Alireza Korangy,
Corey Miller (Eds.)

TRENDS
IN IRANIAN
AND PERSIAN
LINGUISTICS

Alireza Korangy, Corey Miller (Eds.)
Trends in Iranian and Persian Linguistics

Trends in Linguistics Studies and Monographs

Volume 313

Trends in Iranian and Persian Linguistics

—

Edited by
Alireza Korangy
Corey Miller

DE GRUYTER
MOUTON

Acknowledgments

The editors would like to thank their preternaturally patient contributors who braved through the long and arduous – yet delightful – preparation of this volume. The complexities associated with correctly representing linguistic details in all forms and shepherding computer irregularities stemming from multiple systems and countries are not to be taken lightly. We would especially like to thank Martin Schwartz who personifies the spirit of scholarship with his herculean attention to detail and who was a source of great motivation for the editors by example. Corey Miller would like to thank Lixin Yang for his emotional and technical support. With respect to Corey Miller, the views and opinions expressed in this volume are those of the authors and editors, and not those of The MITRE Corporation.

*Alireza Korangy dedicates this book to his daughter
Iran Ghazal Korangy: "a little linguist, who is my motivation
for all things good and positive".*

Table of contents

Introduction

The editors of this volume first met at a conference entitled "The Wide World of Persian: Connections and Contestations, 1500–Present", held at the University of Maryland and the Library of Congress in 2014. That conference brought together literary and linguistic researchers focusing on chronologically and geographically diverse varieties of Persian. This volume, *Trends in Iranian and Persian Linguistics*, inherits from that conference its dedication to diversity in time and space, as well as a deep rooting in the cultural and historical underpinnings of so many linguistic facts. In fact, we have broadened the spatial aspect to encompass a wide range of Iranian languages in addition to Persian as the title implies. While we are more strictly in the linguistic camp, we span the broader range of linguistic endeavor, as will be outlined below.

There are many ways in which a volume of linguistic articles could be assembled. One, of course, would be by linguistic subfield; for example, a collection of articles on nasal vowels, with contributions discussing the phenomenon in different languages. Our collection is more akin to the kind that emanates from the highly successful biennial International Conference on Iranian Linguistics. In fact, we have had the good fortune to meet most of the authors at that conference.

In this introduction, we will situate the articles herein in the broader context of linguistics and Iranian studies. They span the gamut of linguistic studies, from both the historiography of the field and historical linguistics to studies examining specific morphological, phonetic, syntactic, semantic, and pragmatic questions, many providing a deep sociolinguistic context to the evolution and maintenance of the phenomena under study. While the majority of the articles employ painstaking tools of analysis that have characterized the field since time immemorial, the article by Marszalek-Kowalewska illustrates the evolution of those tools into the modern computational realm.

While Hoenigswald (1999) has said that for linguists, "It is their glory to do their work without looking right or left, letting themselves be guided by their subject matter, language, and nothing else", Van Hal's article on "The alleged Persian-Germanic connection: A remarkable chapter in the study of Persian from the sixteenth through the nineteenth centuries" offers an important introspective look at the history of our field. An historiographic study such as this propels us to consider the contemporary societal motivations and influences on our work.

Coupled with the complexities of historical views of our field are the inevitable complexities offered by exocentric perspectives. Ido's paper on "*Huihuiguan zazi*:

A New Persian glossary compiled in Ming China", provides us with an unusual Eastern perspective on Persian. The phonetic and semantic perceptions of Persian from Ming China examined in Ido's lexicographic study provide us with a more detailed canvas upon which to consider the evolution of the language, and enrich our understanding of subsequent developments.

Continuing the plumbing of the lexicographic vein is Rossi's masterful study of Balochi vocabulary, "Glimpses of Balochi lexicography: Some iconyms for the landscape and their motivation". Through an examination of the integral connection between language and space, Rossi shows us how language cannot be studied in a theoretical vacuum that ignores the unique spatio-temporal evolution of a people's means of communication.

Schwartz's "On some Iranian secret vocabularies, as evidenced by a fourteenth-century Persian manuscript" is also a critical lexicographic study, this time underlining how Persian is the heritage of a rich melting pot of peoples and cultures, each of them fashioning the language to some extent in their own image, but contributing to the shared vocabulary and ultimately the legacy of the entire *Sprachbund*.

Shortly following the time period of the manuscript described by Schwartz is a morphosyntactic discussion by Lenepveu-Hotz, "Specialization of an ancient object marker in the New Persian of the fifteenth century". Lenepveu-Hotz's detailed examination of the variability of expression of direct and indirect objects foreshadows the set of articles probing the form and function of a wide set of phenomena in contemporary Iranian Persian discussed in this volume.

Following these diachronically focused articles, we are fortunate to have several contributions navigating the geographical extent of Persian and other Iranian languages, among these, two articles detailing phenomena characterizing its Eastern periphery. Rzehak's study, "Fillers, emphasizers, and other adjuncts in spoken Dari and Pashto", while deepening our understanding of his socio-pragmatic subject matter, also buttresses the thesis expressed in Miller et al. (2013) that Dari and Pashto in Afghanistan have, through their long history in the same space, developed much in common with each other and distinctive from other varieties of the same languages spoken in other countries. Ioannesyan's "The historically unmotivated *majhul* vowel as a significant areal dialectological feature" emphasizes the value of examining contemporary varieties of Tajik and Dari in order to gain perspective on the development of phonetic and phonological features that would remain shrouded in mystery if one were to focus solely on contemporary Iranian Persian.

That is, of course, not to say that the study of contemporary Iranian Persian does not yield enormous treasures of its own, as evinced by a wide spectrum of

articles exploring a range of grammatical topics in that variety. Two of our contributions deal with variation in forms of address due to sociolinguistic factors. Eslami, Abdolhosseini and Dini's "Variability in Persian forms of address as represented in the works of Iranian playwrights" explores the sociopragmatic options available to Persian speakers and what each tells us about the interlocutors' relationships. Saeli and Miller's "Some linguistic indicators of sociocultural formality in Persian" focuses on differences between same and differing gender interactions, while offering a novel accounting for the role played by variation between standard and colloquial morphology and lexis in politeness. The standard/colloquial distinction is thoroughly described in Mahmoodi-Bakhtiari's "Spoken vs. written Persian: Is Persian diglossic?" All areas of grammar are described along with the pervasive register distinctions that uniquely characterize the language.

In the morphosyntactic realm, we have two articles dealing with particular phenomena in counting and plurality. Gebhardt's "Accounting for *yek ta* in Persian" explores both the number marker and numeral classifier properties of *ta*. Ghomeshi's "The associative plural and related constructions in Persian" explores constructions with *ina* from different angles, including the associative plural, extenders, and compounding. Rounding out the morphosyntactic analyses of phenomena in contemporary Iranian Persian is Mahootian and Gebhardt's "Revisiting the status of *-eš* in Persian", which provides a thorough investigation of the clitic and agreement marker properties of *-eš*, including an important section in which speaker judgments were systematically elicited.

Our collection benefits from three articles exploring morphological, syntactic, and semantic phenomena, which extend our geographic scope to the north and west and the south and east. Vydrin's "'Difficult' and 'easy' in Ossetic" extends our bulwark of theoretical linguistic analyses into the Ossetian sphere. By examining the relationship of the construction under study in both the passive and modal constructions of possibility, Vydrin sheds light on a phenomenon that has not received sufficient attention in grammars of the language. Yusupova's "Possessive construction in Kurdish" explores differences in the expression of this construction in northern and southern dialects on the basis of both poetry and folklore texts. Along similar lines, Jahani's "To bring the distant near: On deixis in Iranian oral literature", probes the expression of deixis and its interaction with tense in two varieties of Balochi as well as Vafsi and Gorani.

Finally, Marszalek-Kowalewska's "Extracting semantic similarity from Persian texts" thrusts the linguistics of Iranian languages into the computer era. Pursuing lexicographic themes discussed in the earlier articles, Marszalek shows how computational semantics offers new ways of exploring synonymy and the effects of normative efforts on the way the language is actually used.

It is our fervent hope that this volume will be useful both to those seeking an introduction to some of the themes prevalent in studies of the linguistics of Iranian languages, as well as to those already deeply immersed in the field. Even though, as Sa'di tells us:

بنی آدم اعضای یک پیکرند

[The children of Adam are limbs of one body]

Rumi teaches us that we may still find enrichment through our linguistic diversity and our attempts in reaching a common ground through linguistic understanding:

هین سخن تازه بگو تا دو جهان تازه شود

[Go speak a new language so to replenish both worlds anew!]

The Editors

References

Hoenigswald, Henry M. 1999. Review of Mark Durie & Malcolm Ross, *The comparative method reviewed: Regularity and irregularity in language change. Diachronica* 16 (1): 165–176.

Miller, Corey, Evan Jones, Rachel Strong & Mark Vinson. 2013. Reflections on Dari linguistic identity through toponyms. In Rudolf Muhr, Carla Amorós Negre, Carmen Fernández Juncal, Klaus Zimmermann, Emilio Prieto & Natividad Hernández (eds.), *Exploring linguistic standards in non-dominant varieties of pluricentric languages*, 319–330. Frankfurt: Peter Lang.

Toon Van Hal

1 The alleged Persian-Germanic connection: A remarkable chapter in the study of Persian from the sixteenth through the nineteenth centuries[1]

Abstract: The idea that Modern Persian (Farsi) and the Germanic language group (especially Dutch and German) were connected was formulated around the end of the sixteenth century and remained influential, also after the (re)discovery of Sanskrit at the end of the eighteenth century and the foundation of comparative linguistics in the first half of the nineteenth century. This contribution aims at outlining the history of this remarkably persistent idea and will discuss the linguistic arguments used by Western scholars to substantiate these claims. It will show how Western authors compiled lexical parallels between both language groups and how they explored morphological similarities. Rather than casting new light on the history of the Iranian languages and dialects, the article will reveal how the Persian language was studied, received, and even appropriated in Early Modern Europe.

Keywords: Early Modern language comparison, basic vocabulary, appropriation

1 Introduction

One year before his untimely death, the Leiden physician and Orientalist Johannes Elichmann (1601/1602–1639) expressed his admiration for the renowned female scholar Anna Maria van Schurman (1607–1678) by sending her a quatrain on parchment, calligraphically written in both Persian and Dutch (see Figure 1).[2] Despite the wide attention Van Schurman has attracted in the past few decades

---

1 Sections 2 and 3 of this article consist of abridged and rewritten parts of Van Hal (2007 and 2011). All translations are mine. I am indebted to an anonymous reviewer and to the editors for their comments on earlier drafts of this article.

2 Preserved in The Hague, Royal Library of the Netherlands (Ms 133 B 8, N 75). See Van der Stighelen (1987: 239).

Toon Van Hal, University of Leuven
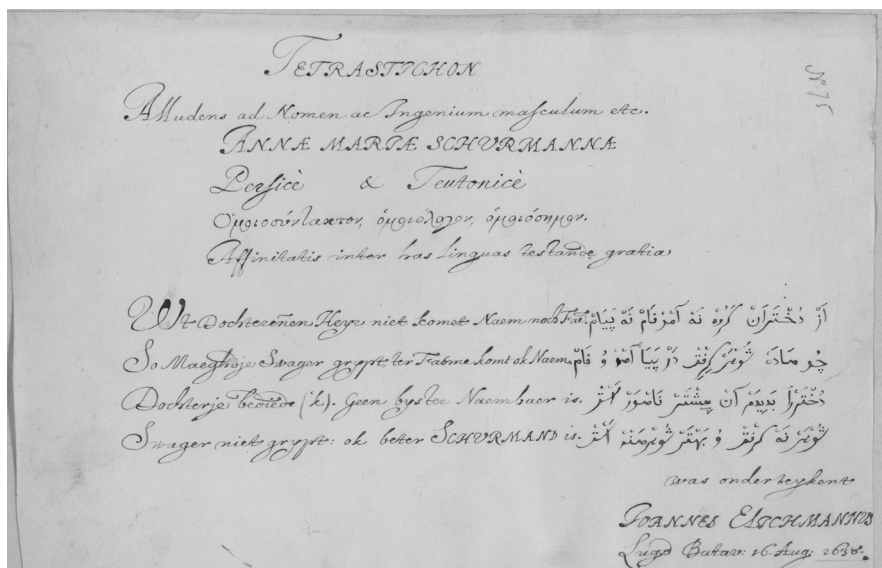
Figure 1: Quatrain by Johannes Elichmann destined for Anna Maria van Schurman (Royal Library of the Netherlands, The Hague: Ms 133 B 8, N 75)

as one of the first female students admitted to a university, both the Dutch and the Persian lines have remained untranslated until recently.[3] This most likely has to do with the fact that Elichmann, an expert in Persian, wanted this poem to demonstrate the affinity between these languages (*affinitatis inter has linguas testandae gratia*). He therefore selected words that were, in both languages, similar in meaning (ὁμοιόσημον) as well as in sound (ὁμοιόλογον); on top of all this, he ordered the words in the same sequence (ὁμοιοσύντακτον). Needless to say, this experiment resulted in a somewhat artificial and hard-to-understand text.

This article aims to present a remarkable chapter in the early history of Western scholars studying Iranian languages, against the background of which Elichmann's astonishing endeavor can be understood. After a period of about 1,000 years during which few scholars from Western Europe studied foreign languages other than Latin, the year 1492 marks the breakthrough of multilingualism on several fronts (see, e.g., Aarsleff 1982: 281). The discovery of America

---

3 See the forthcoming edition by Larsen and Maiullo (2017). I would like to thank Anne Larsen for drawing my attention to this curious epigram (see Larsen [2016] for recent work on Anna Maria van Schurman).

confronted Europe with numerous Amerindian languages, which had to be mastered by Spanish missionaries in order to conduct evangelization. In the wake of the 1492 conquest of Granada, a large number of Jews took refuge in northern cities of Europe, which catalyzed the study of Hebrew. It was also in 1492 that Antonio de Nebrija (1441–1522) published a grammar of Spanish (*Gramática de la lengua castellana*), thus beginning the trend of writing grammars of vernacular languages. It would take another one hundred years for Early Modern Europe, in the wake of the Early Modern globalization, to become familiar with, and conduct investigations into, the Persian language. Europe-based scholars were surprised to notice that the Farsi language bore a surprising number of parallels with the languages spoken in Europe. In particular, the alleged link between German – which included at that time not only High German, but also Low German, Dutch, and at times even English – and Persian was highlighted throughout the entire premodern period and even well into the nineteenth century. This contribution seeks to outline the history of this remarkably persistent idea and to discuss the linguistic arguments Western scholars used to substantiate these claims. It will show against what background Western authors compiled lexical parallels between both language groups and how they explored morphological similarities. Rather than casting new light on the history of the Iranian languages and dialects, the article will reveal how the Persian language was studied, received, and even appropriated in Early Modern Europe.

2 Raphelengius, Scaliger, and Lipsius: The launch of the theory in the last quarter of the sixteenth century

In the first book of his *Historiae*, the Greek historian Herodotus (fifth century BC) refers to a Persian tribe called the *Germanioi* (1.125). As both earlier and more recent scholars have pointed out,[4] it is likely that Herodotus was hinting at the inhabitants of the satrapy *Karmania*.[5] For many humanist scholars, however, it was tempting to suppose that Herodotus was referring to the *Germani*.[6] Only a very small number of humanists voiced some occasional protest against this

---

4 See, e.g., Schmitt (1996) and Cluverius (1631: 30).
5 See Wiesehöfer (2006). Today, Iran has a province and a city that are still called Kerman.
6 This was also defended by modern scholars such as Hallenberg (1816: 129) and Neckel (1929).

tendency to read so much into a mere similarity between two ethnonyms.[7] Indeed, inferring genealogical relationships between two tribes on the basis of similar names was a widespread historical device in the Early Modern period.[8] Pointing out the similarities between ethnonyms of peoples was, in the opinion of many a scholar, sufficient to establish the historical affinity or even identity of these peoples.

In view of all this, it is very possible that Herodotus's brief remark underlies a very noteworthy and long-lasting theory that emerged in the Early Modern period. Indeed, some sixteenth-century scholars alluded to the alleged special link between Germans and Persians, although without addressing the issue of language. For Johannes Goropius Becanus (1519–1573), Herodotus's testimony was sufficient to draw the far-reaching conclusion that the Persians originally spoke the same language as the Dutch.[9] Although extremely well-versed in the classical languages and Hebrew, Goropius – who owes his enduring fame, or infamy, to his view that all languages, including Hebrew, stemmed from the Dutch (or "Cimbric") language – lived during a period in which Persian was still hardly known in the West. This situation changed from the end of the sixteenth century onward once Shah ʿAbbas I (r. 1588–1629) came to power and opened Persia to the outside world.[10] There are two main reasons why the first students of Persian would be inclined to think that Persian would be one of the so-called *linguae orientales* (~ Semitic languages, cf. Gruntfest 1995). First, Persian has many Arabic loanwords in its vocabulary. Second, the language was written mostly in Hebrew or Arabic characters. It is, however, interesting to observe how some of the greatest scholars of the sixteenth century developed an interest in Persian and in its parallels with German and Dutch.

To the best of my knowledge, the first scholar hinting at the parallels between the Persian and German languages in a published piece of scholarship was Joseph Justus Scaliger (1540–1609), a French Huguenot scholar who would

---

7 "De cetero, eorum hic maxime notanda est parum felix coniectura, qui a Persarum gente Germanos ortos, ex Herodoto se probare posse arbitrantur. scilicet quia huic, in libr. 1, sub Persarum imperio populi recensentur Γερμάνιοι. quibus equidem duplex huius sententiae ratio: primum, quia multa Persarum, cum in sermone, tum in moribus, ac vivendi ratione, putant Germanis convenire; deinde, quia nomen idem" (Cluverius 1631: 30).

8 Some humanists had been eager to prove the common roots of the *Armenians* and *Arameans*, while others sought to show that the *Dani* and *Daci* were basically the same tribe. Much ink has been spilled over the question of whether the *Gothi* and the *Getae* were basically the same tribe or not.

9 See Goropius Becanus (1580 [*Hermathena*]: 226). For another example, see, e.g., Neander (1581: 31v).

10 See de Bruijn (1987, 1990) and Matthee (2009).

become world-famous for the breadth of his studies and the scope of classical languages he mastered. In 1579, he remarked in a side note in his edition of the works of the Roman poet Marcus Manilius (fl. first century AD) that Goropius Becanus should have studied Persian, which would have allowed him to discover that words such as *fader*, *muder*, *brader*, *tuchther*, and *band* surfaced both in Persian and in his own Dutch language (Scaliger 1579: 244). Scaliger was a fierce opponent of Goropius Becanus's ideas, and it is therefore difficult to assess what his ultimate point was. It is, however, Franciscus Raphelengius Sr. (1539–1597) who is in earlier publications credited with having launched the Persian-Germanic theory. This specialist in Oriental languages (see e.g., Hamilton 1989) embraced the idea much more warmly than Scaliger did. Before running the new Leiden branch office of the publisher's house of his father-in-law Christopher Plantin (ca. 1520–1589), Raphelengius was active as a proof-reader at Plantin's Antwerp office. There he contributed to the *Biblia Polyglotta* project undertaken by his father-in-law and the Spanish Orientalist Benedictus Arias Montanus (1527–1598). By studying a Persian translation of the Bible, Raphelengius noticed some striking parallels between the Persian and Dutch vocabulary. In 1584, he sent a letter to the professor Justus Lipsius (1547–1606), then at Leiden, in which he included a new list of lexical parallels – it was not the first one, but any previous lists communicated are, unfortunately, no longer extant. In this letter he also suggested that he regarded Goropius Becanus's "Cimbric theory" as a plausible background against which the Persian-Dutch correspondences could be understood. He also emphasized that parallels could also be drawn with Latin, Greek, and Oriental languages, which implies that he did not regard the Persian-Germanic connection as an exclusive one.[11]

Although Raphelengius, who eventually became the first professor of Hebrew at the new university of Leiden, failed to publish his discovery, he appears to have been the most prolific proponent of the idea. It is plausible that he communicated it to the jurist Hugo Grotius (1583–1645), as one of Grotius's texts written in his younger years encompasses a discussion of an extensive number of words shared by Germans and Persians, particularly words belonging to what is now some-times styled "basic vocabulary", such as kinship terms.[12] Raphelengius's Leiden colleague Bonaventura Vulcanius (1538–1614), a professor of Greek, published in 1597 a work containing numerous language specimens (see Van Hal 2010c). This treatise also surveys the similarities between Persian and German, for which Vulcanius acknowledges a debt to Raphelengius (Vulcanius 1597: 87). Hence,

---

11 The letter was published by Nauwelaerts and Sué (1983: 84 09 23 R).

12 This *Parallelon rerumpublicarum liber tertius* was likely written about 1602, but published only in 1801–1803. See Grotius (1801–1803: 62–63).

Vulcanius can be credited with having published the first list of Persian-Germanic parallels. Meanwhile, Lipsius had failed to meet his promise to Raphelengius to discuss the Persian-Germanic connections in the second volume of a miscellaneous book project entitled *Electa* (1585). In 1602, however, he addressed the issue in a very influential treatise, in which he firmly criticized Goropius Becanus's arguments supporting the primacy of Dutch. The letter included a list of Persian-Dutch word comparisons (Lipsius 1602: 56), probably the second one ever published (see Deneire and Van Hal [2006] for further background). Finally, there are strong indications that Raphelengius also inspired Philips van Marnix van Sint-Aldegonde (1540–1598), an influential Protestant politician, to further explore the Persian-Germanic correspondences (see Schulte [1879: 331] and Van Hal [2011: 154–155]).

Although Scaliger and Lipsius were both key figures in launching the theory, both scholars had raised large questions over the general validity of the exclusive connection. While not casting doubt on the similarities between Dutch and Persian, Lipsius remarked that for many Persian words, the Latin equivalents seemed to be closer than the Dutch ones. We will see, indeed, how a considerable number of later authors did not exclusively focus on the correspondences between Persian and German, but investigated Persian's parallels with languages such as Greek and Latin as well.[13] Lipsius's general tenet was that the evidential value of languages, given the continuous changes they underwent throughout time, was much too slippery to be of any significance in historical enquiries: "whoever is looking for solidity in a topic that is fundamentally unstable, viz. language, makes a mistake".[14] In other words, he thought that this kind of scholarship was doomed to failure. In addition, Scaliger emphasized that it was very hazardous to assume linguistic kinship from a narrow empirical foundation. Unlike Lipsius, however, Scaliger was very interested in exploring sound bases to compare languages (see Van Hal 2010b). When it comes to assessing the similarities between Persian and Germanic, Scaliger – who was actually the first to have the idea – seems to have been puzzled by the nature of this closeness. Toward the end of his life, he stated that "nothing differs more from German than Persian".[15] However, there are good reasons to assume that this judgment should not be taken at face value. Judging by his "table talks" (published posthumously without his agreement and therefore not

---

13 Jan van den Driessche's (Joannes Drusius, 1550–1616) plan to publish a book on Greek borrowings from Persian did not materialize (Drusius 1622: 925).
14 "Errat enim qui in re instabili maxime, id est lingua, quaerit firmitatem" (Lipsius 1602: 55).
15 "Nihil tam dissimile alii rei, quam Teutonismus linguae Persicae" (Scaliger in the introduction to Pontanus [1606]).

Figure 2: List of Persian words that are similar to words in other languages, in Scaliger's hand (Collection Leiden University Libraries: Cod. Scal. 57, f. 51)

entirely reliable), he had stressed the lexical similarities between the two languages (Scaliger 1740: 111). In addition, the Leiden University Library holds a draft of a Persian lexicon initiated by Raphelengius and continued by Scaliger (Cod. Scal. 57, f. 25–50). This draft is followed by another draft (f. 51), unmistakably in Scaliger's hand, encompassing a list of Persian words that are similar to words in other languages, among which many comparisons between Dutch and Persian appear (see Figure 2). He offers examples such as שַׁ֤ךְ / شاخ *shākh* 'branch', Latin *ramus* 'branch', Dutch *tak* 'branch' and عود *'wd* 'wood', Latin *lignum*, Dutch *hout* 'wood'. Relying on all sources known and available, it is difficult to draw another conclusion than that Scaliger remained hesitant regarding this thorny issue.

3 The various explanations for the Persian-Germanic connection

It is safe to posit that the critical attention Lipsius and Scaliger had paid to the Persian-Germanic connection in the early days after its discovery contributed considerably to its swift and wide dissemination, since both scholars were generally regarded as among the very best of their generation. This is not to say that there were no scholars who preferred to simply ignore the hypothesis. In

the preface to his *Lexicon heptaglotton* (1669), the English Orientalist Edmundus Castellus (1606?–1685), for instance, classified Persian among the daughter languages of Hebrew without speculating on its possible connection with any European language. Remarkably, he left unexplained why Persian was given a grammatical and lexical description separate from those he provided for Hebrew, Chaldean, Syrian, Ethiopic, Samaritan, and Arabic.[16] There was also a significant group of scholars who virulently contested this alleged connection (Droixhe [1989] surveys the arguments of a number of such adversaries). Nevertheless, it is hard to deny that, in many texts, the Persian-Germanic theory was, in one way or another, mentioned and addressed as though the alleged connection was a well-established fact. A first glance at the sources immediately reveals that champions of the theory cannot be regarded as one cohesive group. Scholars who acknowledged the special connection did not always share the same views on the underlying causes explaining the linguistic similarities. First of all, some scholars limited themselves to observing the parallels without pondering possible explanations.[17] A large number of scholars were convinced that early contacts between German and Persian tribes had given way to lexical convergences. In a book entitled *De lingua Belgica* [On the Dutch language], the Dordrecht pastor Abraham Mylius (1563–1637) took care to show how Dutch colonizers had once subjected an impressive geographical area, even reaching Persia. The similarities between Dutch and several other languages, including Persian, served to underpin this nationalist-like scheme. In contrast, Bernard Furmer (1542–1616) did not trace the Persians back to the Germans, but the other way around.[18]

Instead of explaining the parallels in terms of borrowings, some scholars preferred to see them as remnants of the *lingua Adamica*. As a matter of fact, the book of Genesis recalls how the infelicitous idea of building a tower in Babel that would reach to Heaven urged God to confuse the original language Adam

---

16 As reflected in the full title of his work: *Lexicon heptaglotton, Hebraicum, Chaldaicum, Syriacum, Samaritanum, Aethiopicum, Arabicum conjunctim, et Persicum separatim.*

17 Vulcanius pointed out that Persian and Dutch are "somehow" cognate without clearly defining the nature and the extent of this affinity: "aliquam enim eius esse cum Teutonica affinitatem vel ex eo constat quod multa vocabula utrique linguae inter se sunt communia" (Vulcanius 1597: 87). John Greaves (Gravius, 1602–1652) rounds off his very systematic Persian grammar with a comparative word list and with the observation that many Persian words are similar to English ones, without advancing a clear explanation (Gravius 1649: 89 *eqs*; see also Droixhe [Forthcoming]).

18 "Germanos autem illos ex Persia ortos nomen suum huic orbis parti contulisse, plerique docti et ex lingua et ex moribus utrique nationi communibus acute coniciunt" (Furmerius 1609: 12).

had invented in paradise. Although this language was thus damaged beyond repair, scholars such as the Leiden geographer Philippus Cluverius (1580–1622) theorized that some bits and pieces of it were still detectable across a number of languages of the world. In other words, he regarded not only the similarities between Persian and German, but also those among Latin, Greek, the Indian languages, and Hebrew as merely accidental relics of the primeval language, which was – apart from these scarce vestiges – irrevocably lost.[19]

It was, however, the previously mentioned Elichmann who would lay the foundations for a significant breakthrough (see Van Hal 2010a). An expert in Persian, he was convinced of its special kinship with Germanic, even to the point of constructing poems that could be understood in both Dutch and Persian alike (see above). The Persian-Germanic connection would become the backbone of the Scythian framework Elichmann designed. Apart from German and Persian, languages such as Latin and Greek would have descended from the languages of the "Scythians". Two Leiden professors would contribute considerably to the further elaboration and dissemination of this groundbreaking idea. While in his earlier years, he still explained the parallels as Persian borrowings from Greek (Salmasius 1629: 1130), Claude de Saumaise (1588–1653) was the first to launch Elichmann's idea in print (Salmasius 1643) and even to offer some reconstructions of Scythian numbers. The second Leiden professor who followed in Elichmann's footsteps was Marcus Zuerius van Boxhorn (1612–1653); see, in this context, especially Boxhornius (1720). His enduring fervor for the "Scythian case" stood in contrast to the more ephemeral (yet more influential) interest shown by his rival Saumaise.

4 Evolving views on the nature of the Persian-Germanic connection and on the way of comparing languages

Throughout the seventeenth and eighteenth centuries, the idea of a Persian-Germanic connection never sank into oblivion. Many allusions, often very concise, were made in major encyclopedias,[20] dictionaries,[21] historical, chorographical,

---

19 "Quamobrem nihil mirum, si diversissimae, longeque inter se remotissimae per universum terrae orbem gentes communia habeant etiam nunc nonnulla rerum vocabula" (Cluverius 1631: 30; see also p. 59).

20 For example, in *Zedler's universallexikon*, the most extensive eighteenth-century encyclopedia; see Ludovici (1741: 661) as well as in the third edition of the *Encyclopaedia Britannica* (Anon. 1797: 564).

21 See, e.g., Eberhard (1800: 248).

and antiquarian works,[22] and even in question-and-answer books tailored for a younger readership.[23] Moreover, knowledge of the hypothesis extended beyond the exclusive circle of learned men and women of letters. A number of missionaries and merchants, who can be regarded as linguistic fieldworkers overseas, were obviously familiar with the Persian-Germanic hypothesis when composing Persian dictionaries and grammars. The French Discalced Carmelite Ange de Saint Joseph (1636–1697) converted souls to the Christian faith in Persia and Turkey between 1664 and 1679 (Walsh 1907). His 1684 *Gazophylacium linguae Persarum* is a Persian grammar and dictionary in Latin, Italian, and French, containing a small section devoted to the analogues between Persian and the European languages (Angelus a S. Joseph 1684: 5–7; see Droixhe [Forthcoming]). His Protestant peers in India observed some correspondences between German and Hindustani, which demonstrates that other languages belonging to what is now called the Indo-Iranian family were compared with the Germanic languages, once they became known to Western explorers.[24] Another example is Joan Josua Kettler (Ketelaar; 1659–1718). After committing a number of crimes in his homeland, this Elbing-born German fled to the Netherlands, where he succeeded in making a career in the East India Company (Vogel 1936). About 1700, he authored a Dutch-language *Instruction of the Hindustani and Persian languages*, in which he compared some Persian words with Latin and Dutch examples.[25] It is interesting to see that none of these authors felt urged to ponder the reasons underlying these parallels.

Needless to say, it would be both unfeasible and undesirable to offer a complete chronological survey of all sources in which the Persian-Germanic hypothesis is mentioned, addressed, and elaborated upon. Quite remarkably, however, such extensive overviews can be found in some Early Modern and Modern sources.[26] By concentrating on lesser-known source texts, this section

---

22 See, e.g., Schotanus (1664: 5–6), Sammes (1676: 423), and Rapin de Thoyras (1743: 26).

23 See, e.g., Anon. (1785: 9).

24 "Sonst finden sich in der Hindostanischen Sprache viele Wörter, die ganz mit den Teutschen überein kommen. Als zum Exempel: *Hand* oder *had*, weil das *n* nicht gehöret wird, die Hand; *Mu*, der Mund; *Bocker*, ein Bock; *Kamerband*, ein Gürtel oder Band um die Lenden, *Bandà*, binden; *Binde-bando*, biedet einen Bündel; *Man*, ein Mann; welches letztere im *Composito* viel im Gebrauch ist, als *Beraman*, ein braver oder grosser Mann; *Dutchman*, ein Feind, u.s.w." (Schultze 1747: 713).

25 Ketelaar (1700) – I owe this reference to Anna Pytlowany, who is currently preparing an edition of Ketelaar's manuscript.

26 For example, Eccardus (1711: 209–211), Odhelius and Celsus (1723: 18–19), and especially Dorn (1827: 91–135), presenting first the views of the champions and subsequently the criticisms of the opponents. Droixhe (1978: 81–83) and Helander (2004: 367–369) offer more recent lists.

will therefore explore the general dynamics of the ongoing debate and examine the gradual refinement and sophistication of the linguistic arguments used either to buttress or to undermine the theory.

Many scholars relied solely on the lexical similarities between both languages in order to advocate the mutual kinship. From Vulcanius and Lipsius onward, champions of the Persian-Germanic connection typically offered a small list of about twenty similar words, often adding that many additional words could be supplied without any difficulty. Table 1, which is representative of later lists, reproduces the early yet little-known list provided by the Leiden professor Paullus Merula (1558–1607).

Table 1: List of Persian-Germanic comparisons, offered in Merula and Merula (1627: 544); the English translation column has been added by the author

"Persian"	Dutch	Latin	[English translation]
Choda	God	Deus	'God'
Phedar	Vader	Pater	'father'
Madar	Moeder	Mater	'mother'
Berader	Broeder	Frater	'brother'
Dochtar	Dochter	Filia	'daughter'
Nam	Naem	Nomen	'name'
Dandan	Tanden	Dentes	'teeth'
Lab	Lip	Labium	'lip'
Drog	Bedrogh	Mendacium	'fraud', 'lie'
Star	Ster	Stella	'star'
Mus	Muys	Mus	'mouse'
Casti	Casse	Cista	'box'
Band	Band	Vinculum	'band'
Must	Most	Mustum	'unfermented wine'
Nau	Nieu	Novus	'new'
Du	Du	Tu	'you'
Begrijst	Beschreyt	Lacrymis oppletus	'filled with tears'
Grijft	Grijpt	Tene	'hold'
Murd	Vermoord	Obtruncatus est	'murdered'
Ses	Ses	Sex	'six'
Ta	Daer toe	Usque ad	'until'

This way of comparing languages met with thorough criticism. Reviewing Simon Pelloutier's *Histoire des Celtes*, in which the similarities between German and Persian were discussed in some detail (Pelloutier 1740: 86–89), an anonymous reviewer remarked: "Nevertheless, the presence of a few words that are somehow similar do not prove the identity of two languages. Nor do they even

demonstrate a common origin. They only hint at a natural adoption of words, which easily pass from one language to another".[27]

Such criticisms had also been aired by Richard Verstegan (1548?–1636?), an English antiquary who had been working in Antwerp since 1585 or so. In his *Restitution of decayed intelligence in antiquities concerning the most noble and renovvmned English nation* (1605), we read:

> Surely it is an opinion of a very slender confirmation, for that in deed there is no affinitie at all between those two languages, and albeit there may some half a dozen or half a score woords be found in the Persian, that are broken German woords, as *Choda*, *Phedar*, *Madar*, *Beradar*, *Dochtar*, *Star*, *Band*, for God / Father / Mother / Brother / Daughter / Star / Band / what affinitie makes this, when all the rest is altogether different? [...] By this it may be seen espetially to such as have any knowledge in the Duytsh toung, that between that and this, heer is no neernesse of affinitie at all, but as much farnesse as needeth to be (Verstegan 1605: 26–27, 29).

He added that an investigation by Anthony Sherley (1565–1635), who was ambassador at the Persian court, had clearly revealed that Latin was closer to Persian than Dutch was to Persian. Both critics thus highlighted the poor quantitative foundations on which the claimed Persian-Germanic connections rested. It is precisely against Verstegan that the late seventeenth-century antiquarian Aylett Sammes (1636?–1679?) reacts. Verstegan's argument, this Essex-born scholar states,

> would be true if the words alledged were far fetched, and we were forced to run through a whole Dictionary to find only a few, and those as distant in signification as the Heaven and Earth is from each other, but were so nigh Relations, as *Father*, *Mother*, *Brother* and *Daughter*, which are alwaies in Peoples mouths, are called by the same names in two Languages, it seemeth not to happen by chance. (Sammes 1676: 423)

This remark testifies to the growing awareness of the importance attached to "basic vocabulary" when comparing languages in order to establish genealogical relations (see Van Hal 2015).

This is, however, not to say that the line of reasoning followed by all authors relied solely on the shared lexicon. It is indeed noteworthy to find that, even in the early stages of this argument, both Raphelengius and Lipsius highlighted some grammatical similarities. Nevertheless, it is most likely in the wake of

---

27 "Cependant quelques termes à peu près semblables ne prouvent pas l'identité de deux langues, ni même une commune origine, mais seulement une adoption naturelle de mots, qui passent aisément d'une langue dans une autre" (Anon. 1741: 238). The controversial Jesuit Pierre François Guyot-Desfontaines (1685–1745) might have authored this review, as he was the initiator of the *Observations sur les écrits modernes* series.

Johannes Elichmann that Salmasius observed some similarities in inflection, conjugation, and word composition. One can safely state that the importance attached to structural and grammatical parallels gradually increased over time. The Swedish authors of an academic dissertation dedicated to the Persian-Gothic parallels emphasized the methodological significance of taking into account grammatical structures by relying on the authority of William Wotton (1666–1727), who had highlighted the importance of grammatical correspondences in another context.[28] Other authors still had called attention to similarities in phraseology and syntax.[29] Finally, it is interesting to observe that some scholars argued that the similarities in language went hand in hand with parallels in customs (see, e.g., Wellern 1753: 331–332; Furmerius 1609: 12).

Despite the various explanations given for the Persian-Germanic hypothesis in the first half of the seventeenth century (see above, section 3), it is remarkable to find that a very large number of late seventeenth- and eighteenth-century authors deliberately regarded Boxhorn's Scythian theory as the most plausible framework for explaining the similarities between the languages. As a young man, Hugo Grotius saw no other explanation for the similarities he had observed between Persian and German than by contact, colonization, and borrowing (see above, section 2). By the end of his life, the meanwhile renowned diplomat would also subscribe to the Scythian hypothesis (Grotius 1655: 8). In his sixteenth *Prolegomenon*, discussing the "Persian language and the Persian versions of this Bible", the biblical scholar Brian Walton (1600–1661), editor of a renowned Polyglot Bible, also brings up the Persian-Germanic similarities. After discussing the solutions and answers given by other scholars, he presents Boxhorn's Scythian thesis, "which he will embrace, until a more plausible explanation will be advanced".[30] Along with the above-mentioned Simon Pelloutier (Pelloutier 1740), the Frisian professor Campegius Vitringa the Elder (1658–1722) was a champion of the Scythian framework. He asked whether the contacts and commerce between Persians and Germans had been so intense that they could explain such striking linguistic similarities, as some other Early Modern authors were inclined to think. In this respect, Vitringa himself refers, for instance, to Mylius (1612), but Thomas Hyde (1636–1703) still adhered to the contact theory in 1700 (see Hyde 1700). Vitringa, in turn, was convinced "that both languages,

---

28 Odhelius and Celsius (1723: 14–15, 24), referring to Wottonius (1715) (cf. Droixhe [1994] on this dissertation). See also the substantial grammar-based motivation offered by Henselius (1741: 437–460).

29 See, e.g., Levinus Warner (1619–1665) and August Pfeiffer (1640–1698) in Warnerus (1644: 14) and Pfeifferius (1704: 690).

30 "Probabilis tamen mihi videtur Boxhornii sententia, quam amplectendam sentio, donec aliquid probabilius adferatur" (Walton 1673: 419–420).

Persian as well as Germanic, were born from Scythian".[31] The great polyhistor Gottfried Wilhelm Leibniz (1646–1716), whose interest in the world's languages was ancillary to his fascination with prehistory and early migrations, also mentioned the Scythian theory as a possible explanation of the Persian-Germanic parallels (see, e.g., Babin and van den Heuvel 2004: 365, 843, 877), although he remained somewhat doubtful of the significance of the kinship (see now especially Droixhe [Forthcoming]). A final example is the Dutch Orientalist Albert Schultens (1686–1750), who gained his fame by offering a systematic and grammar-based account of the Semitic – or in his terminology, "Oriental" – languages. Schultens lucidly underlined that the very structure of Persian indicated that this language could not belong to this Oriental language group. Before massively adopting Arabic words, Schultens claimed, Persian was unmistakably of a *Scythian-European* origin (Schultens 1761: 194). It is, however, important to emphasize that the Scythian language was not always regarded as a lost parent language of German and Persian. On a number of occasions, scholars equated Scythian with their own mother tongues, so as to grant their language a time-honored status.

Gradually, more details on the Persian language became known. It was probably due to a lack of Persian data that the Dordrecht preacher Abraham Mylius could not deal with this language in depth, despite the prominent presence of Persian in the subtitle of his 1612 book (see Mylius 1612). Most scholars entirely focused on Modern Persian. One of the sole seventeenth-century exceptions was Boxhorn, who sought to rely on old Persian words mentioned by classical authors – an exercise continued later by the Utrecht scholar Hadrianus Relandus (1676–1718) (Relandus 1707). From the second half of the eighteenth century onward, the European Republic of Letters witnessed a new and substantial extension of the study of Asian linguistics. The publication of new dictionaries describing Iranian languages other than Farsi enabled scholars to further substantiate the links.[32] It was also in this period that a hitherto unknown, extinct Iranian language, so-called Zend-Avestan, came to the surface. A first edition of the text corpus was published by Abraham Hyacinthe Anquetil-Duperron (1731–1805) (see especially App 2010). It was, however, the (re)discovery of Sanskrit that would mesmerize the majority of scholars from the 1780s onward, thus somewhat eclipsing the focus on Persian.[33]

---

31 "Ego igitur multo faciliorem puto dari posse hujus convenientiae rationem ex illa hypothesi, quod utraque illa Lingua tam Persica quam Germanica nata sit ex Scythica" (Vitringa 1712: 100). See also Huet (1722: 102), discussed by Droixhe (1980).

32 See, e.g., Garzoni's lexicon and grammar of the language of the Kurds (Garzoni 1787), which was commented upon by Kinderling (1795: 95).

33 See especially Benes (2008: 72–73), who also offers some examples of scholars who kept believing in the primacy of Persian.

At the turn of the nineteenth century, we see that many scholars are inclined to explain the new data offered by Sanskrit in the context of the age-old Persian-Germanic (Scythian) theory. In other words, it seems safe to state that the (re)discovery of Sanskrit did not immediately give way to a fundamental rethinking of contemporary ideas on linguistic kinship. In a work published as late as 1819, the scholar Joseph Cherade de Montbron (1768–1854) still explained the similarities between Persian and German by referring to the invading *Tartars* or Scythians (Ch[erade] de Montbron 1819: 588–589), although he also knew about new developments in Indian philology.[34] The theory's ongoing impact still echoes in Franz Bopp's (1791–1861) groundbreaking work *Über das Conjugationssystem* (Bopp 1816), whose fifth chapter is devoted to the "conjugation der persischen Sprache und der alten germanischen Mundarten" (see in this respect Hiersche 1985: 157).[35] Without providing any references, Windfuhr (1979: 155) even asserts that the idea of a Persian-German affinity is still somehow present in the folk wisdom of some regions in both Iran and Germany.

It was the general aim of this article to show that the study of Persian greatly stimulated European ideas on linguistic kinship from the end of the sixteenth century onward. Droixhe (1984) has rightly stressed that the idea of the Persian-Germanic connection is thus a remarkable instance of *continuity* in the history of (pre)comparative linguistics, which was in other respects very often characterized by discontinuity.

References

Aarsleff, Hans. 1982. *From Locke to Saussure*. Minneapolis: University of Minnesota Press.

Angelus a S. Joseph. 1684. *Gazophylacium Linguae Persarum, Triplici linguarum Clavi Italicae, Latinae, Gallicae, nec non specialibus praeceptis ejusdem linguae reseratum*. Amstelodami: Ex officina Jansonio-Waesbergiana.

Anon. 1741. *Observations sur les écrits modernes*. Paris: Chaubert.

Anon. 1785. *Epitome historica scientiarum et artium. Ad usum studiosae iuventutis*. Dresdae: Ex officina Waltheria.

Anon. 1797. Philology. *Encyclopædia britannica, or, A dictionary of arts, sciences, and miscellaneous literature*, vol. 14: 485–569.

App, Urs. 2010. *The birth of orientalism*. Philadelphia: University of Pennsylvania Press.

Babin, Malte-Ludolf & Gerd van den Heuvel (eds.). 2004. *[Gottfried Wilhelm Leibniz]: Schriften und Briefe zur Geschichte*. Hannover: Verlag Hahnsche Buchhandlung.

---

34 See, e.g., Kolbe (1819: 504–507), Hallenberg (1816), and Van Hal (2012) for other examples.

35 For similar nineteenth-century references to the Persian-German connection, see also Vater (1815: 186) and especially Frank (1808: 11).

Benes, Tuska. 2008. *In Babel's shadow: Language, philology, and the nation in nineteenth-century Germany*. Detroit: Wayne State University Press.

Bopp, Franz. 1816. *Über das Conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprache [...]*. Frankfurt am Main: in der Andreäischen Buchhandlung.

Boxhornius, Marcus Zuerius. 1720. *Epistola de Persicis Curtio memoratis vocabulis, eorumque cum Germanicis cognatione notis instructa*. Edited by Jo. Henr. Von Seelen. Lubeccae: apud Petrum Boeckmannum.

Bruijn, Johannes T. P. de. 1987. Iranian studies in the Netherlands. *Iranian Studies* 20 (2/4). 161–177.

Bruijn, Johannes T. P. de. 1990. *De ontdekking van het Perzisch. Rede uitgesproken bij de aanvaarding van het ambt van bijzonder hoogleraar in de cultuurgeschiedenis van Iran sedert de opkomst van de Islam aan de Rijksuniversiteit te Leiden op 9 maart 1990*. Leiden: Rijksuniversiteit Leiden.

Ch[erade] de Montbron, J[oseph]. 1819. *Essais sur la littérature des Hébreux*. Paris: Louis Janet.

Cluverius, Philippus. 1631. [1616]. *Germaniae antiquae libri tres [...]*. Lugduni Batavorum: Ex officina Elzeveriana.

Deneire, Tom & Toon Van Hal. 2006. *Lipsius tegen Becanus. Over het Nederlands als oertaal. Editie, vertaling en interpretatie van zijn brief aan Hendrik Schotti (19 december 1598)*. Amersfoort: Florivallis.

Dorn, Boris Andreevich. 1827. *Über die Verwandtschaft des persischen, germanischen und griechisch-lateinischen Sprachstammes*. Hamburg: J. A. Meissner.

Droixhe, Daniel. 1978. *La linguistique et l'appel de l'histoire (1600–1800). Rationalisme et révolutions positivistes*. Genève: Droz.

Droixhe, Daniel. 1980. Le prototype défiguré. L'idée scythique et la France gauloise (XVIIe–XVIIIe siècles). In E. F. K. Koerner (ed.), *Progress in linguistic historiography. Papers from the International Conference on the History of the Language Sciences (Ottawa, 28–31 August 1978)*, 123–137. Amsterdam: Benjamins.

Droixhe, Daniel. 1984. Avant-propos. *Genèse du comparatisme indo-européen = Histoire, Épistémologie, Langage* 6 (2). 5–15.

Droixhe, Daniel. 1989. Boxhorn's bad reputation. A chapter in academic linguistics. In Klaus D. Dutz (ed.), *Kurzbeiträge der IV. Internationalen Konferenz zur Geschichte der Sprachwissenschaften (ICHoLS IV). Trier, 24–27. August 1987*, 359–384. Münster: Nodus.

Droixhe, Daniel. 1994. En attendant Bopp. Une dissertation sur la convenance du perse et du gothique de 1723. In Reinhard Sternemann (ed.), *Bopp-Symposium 1992 der Humboldt-Universität zu Berlin. Akten der Konferenz vom 24.3.–26.3.1992 aus Anlaß von Franz Bopps zweihundertjährigem Geburtstag am 14.9.1991*, 53–71. Heidelberg: Winter.

Droixhe, Daniel. Forthcoming. *The Failure of the Germano-Persian Kinship*.

Drusius, Johannes. 1622. *Veterum interpretum Graecorum in totum vetus Testamentum fragmenta*. Arnhemiae: apud Iohannem Ianssonium.

Eberhard, Johann August. 1800. *Versuch einer allgemeinen deutschen Synonymik in einem kritisch-philosophischen Wörterbuche der sinnverwandten Wörter der hochdeutschen Mundart*. Halle & Leipzig: J. G. Ruff.

Eccardus, Johannes Georgius. 1711. *Historia studii etymologici linguae Germanicae hactenus impensi; ubi scriptores plerique recensentur et diiudicantur, qui in origines et antiquitates linguae Teutonica, Saxonicae, Belgicae, Danicae, Suecicae, Norwegicae et Islandicae veteris item Celticae, Gothicae, Francicae atque Anglo-Saxonicae inquisiverunt, aut libros*

studium nostrae linguae criticum promoventes alios ediderunt. Accedunt et quaedam de lingua Venedorum in Germania habitantium, tandemque proprium de lexico linguae gG aperitur. Hannoverae: apud Nicolaum Foersterum.

Frank, Othmar. 1808. *Das Licht vom Orient*. Nürnberg & Leipzig: Lechner & Besson.

Furmerius, Bernardus. 1609. *Annalium Phrisicorum libri tres*. Franecarae: excudebat Aegidius Radaeus.

Garzoni, Maurizio. 1787. *Grammatica e vocabolario della lingua Kurda*. Roma: Sacra Congregazione di Propaganda Fide.

Goropius Becanus, Johannes. 1580. *Opera hactenus in lucem non edita, nempe Hermathena, Hieroglyphica, Vertumnus, Gallica, Francica, Hispanica*. Antverpiae: excudebat Christophorus Plantinus.

Gravius, Johannes. 1649. *Elementa linguae Persicae*. London: Flesher.

Grotius, Hugo. 1655. *Historia Gotthorum Vandalorum et Langobardorum; praemissa sunt eiusdem Prolegomena ubi regum Gotthorum ordo et chronologia, cum elogiis; accedunt nomina appellativa et verba Gotthica, Vandalica, Langobardica cum explicatione*. Amstelodami: apud Ludovicum Elzevirium.

Grotius, Hugo. 1801–1803. *Batavi, parallelon rerumpublicarum liber tertius: de moribus ingenioque populorum Atheniensium, Romanorum, Batavorum. Vergelijking der Gemeenebesten. Derde boek: over de zeden en den inborst der Athenienseren, Romeinen en Hollanderen*. Edited by Johan Meerman. Haarlem: Loosjes.

Gruntfest, Yaakov. 1995. On the history of the classification of Semitic languages. In Kurt R. Jankowsky (ed.), *History of linguistics 1993. Papers from the Sixth International Conference on the History of the Language Sciences (ICHoIS VI), Washington D.C., 9–14 August 1993*, 147–156. Amsterdam & Philadelphia: Benjamins.

Hallenberg, Jonas. 1816. *Disquisitio de nominibus in lingua Suiogothica, lucis et visus cultusque solaris in eadem lingua vestigiis. Additae hinc inde sunt generaliores de linguarum origine observationes*. Stockholmiae: Apud Direct. Henr. A. Nordström.

Hamilton, Alastair. 1989. Nam tirones sumus. Franciscus Raphelengius' *Lexicon Arabico-Latinum* (Leiden 1613). In Marcus de Schepper and Francine De Nave (eds.), *Ex officina Plantiniana. Studia in memoriam Christophori Plantini (ca. 1520–1589)*, 523–556. Antwerpen: Vereeniging der Antwerpsche Bibliophielen.

Helander, Hans. 2004. *Neo-Latin literature in Sweden in the period 1620–1720. Stylistics, vocabulary and characteristic ideas*. Uppsala: University Library.

Henselius, Godofredus. 1741. *Synopsis universae philologiae, in qua miranda unitas et harmonia linguarum totius orbis terrarum occulta e literarum, syllabarum vocumque natura et recessibus eruitur*. Norimbergae: In commissis apud heredes Homannianos.

Hiersche, Rolf. 1985. Zu Etymologie und Sprachvergleichung vor Bopp. In Hermann M. Ölberg & Gernot Schmidt (eds.), *Sprachwissenschaftliche Forschungen. Festschrift für Johann Knobloch*, 157–169. Innsbruck: Institut für Sprachwissenschaft.

Huet, Pierre-Daniel. 1722. *Huetiana, ou, Pensées diverses de M. Huet [. . .]*. Paris: Chez Jacques Estienne.

Hyde, Thomas. 1700. *Historia religionis veterum Persarum, eorumque magorum [. . .]*. Oxonii: Sheldon.

Ketelaar, Johan Josua. 1700. *Instructie of onderwijsinghe der Hindoustanse en Persiaanse taalen*. http://objects.library.uu.nl/reader/index.php?obj=1874-44896&lan=nl&_ga=1.31740217. 641540806.1461772719#page//31/92/68/3192688800940483454676002176042800784. jpg/mode/1up (accessed 31 December 2016).

Kinderling, Johann Friedrich August. 1795. *Über die Reinigkeit der Deutschen Sprache und die Beförderungsmittel derselben [. . .]*. Berlin: Bey Friedrich Maurer.

Kolbe, Karl Wilhelm. 1819. *Über den Wortreichtum der deutschen und französischen Sprache*. Zweite, ganz umgearbeitete Ausgabe. Berlin: In der Realschulbuchhandlung.

Larsen, Anne R. 2016. *Anna Maria van Schurman, "the star of Utrecht": The educational vision and reception of a savante*. Farnham & Burlington: Ashgate.

Larsen, Anne R. & Steve Maiullo (eds.). 2017. *Anna Maria van Schurman: Letters and poems to her mentor and other members of her circle*. Toronto: The Toronto Center for Renaissance and Reformation Studies.

Lipsius, Justus. 1602. *Epistolarum selectarum centuria prima [-tertia] ad Belgas*. Antverpiae: ex officina Plantiniana, apud Ioannem Moretum.

Ludovici, Carl Günther (ed.). 1741. *Grosses vollständiges Universal-Lexicon aller Wissenschafften und Künste [. . .]*. Leipzig & Halle: J. H. Zedler.

Matthee, Rudi. 2009. The Safavids under western eyes: Seventeenth-century European travelers to Iran. *Journal of Early Modern History* 13 (2) . 137–171.

Merula, Paullus & Guilielmus Merula. 1627. *Tiid-thresoor: Ofte kort ende bondich verhael van den standt der kercken ende de vvereltlicke regieringe: Vervatende een beschrijvinge van alle de gedencvvaerdichste geschiedenissen over den gantschen aertbodem [. . .] : Alles beginnende vande geboorte Iesu Christi, totten jare 1627*. Tot Leyden: Jan Claesz. van Dorp.

Mylius, Abraham. 1612. *Lingua Belgica sive de linguae illius communitate tum cum plerisque aliis, tum praesertim cum Latina, Graeca, Persica, deque communitatis illius causis, tum de linguae illius origine et latissima per nationes quamplurimas diffusione, ut et de ejus praestantia*. Lugduni Batavorum: Ulricus Cornelii et G. Abrahami.

Nauwelaerts, Marcel A. M. & Sylvette Sué (eds.). 1983. *Iusti Lipsi Epistolae. 2: 1584–1587*. Brussels: Koninklijke Academie voor Wetenschappen, Letteren en Schone Kunsten van België.

Neander, Michael. 1581. *Bedencken Michaelis Neandri An einenn Guten Herrn und Freund, Wie ein Knabe zu leiten und zu unterweisen [. . .]*. s.n.: s.n.

Neckel, Gustav. 1929. *Germanen und Kelten: historisch-linguistisch-rassenkundliche Forschungen und Gedanken zur Geisteskrisis*. Heidelberg: Winter.

Odhelius, Olaus & Olof Celsius. 1723. *Dissertatio philologico-historica de convenientia linguae Persicae cum Gothica [. . .]*. Upsaliae: Literis Wernerianis.

Pelloutier, Simon. 1740. *Histoire des Celtes et particulièrement des Gaulois et des Germains depuis les temps fabuleux jusqu'à la prise de Rome par les Gaulois*. La Haye: chez Isaac Beauregard.

Pfeifferius, Augustus. 1704. *Opera omnia quae extant philologica*. Ultrajecti: ex officina Guilielmi Broedelet.

Pontanus, Johannes Isacius. 1606. *Itinerarium Galliae Narbonensis, cum duplici appendice id est universae fere Galliae descriptione Philologica ac Politica. Cui accedit glossarium Prisco-Gallicum seu de lingua Gallorum veteri Dissertatio*. Lugduni Batavorum: ex officina Thomae Basson.

Rapin de Thoyras, Paul. 1743. *The history of England*, 3rd edn. London: John and Paul Knapton.

Relandus, Hadrianus. 1707. De reliquiis veteris linguae Persicae. In *Dissertationum Miscellanearum partes tres*, vol. 2, 97–266. Trajecti ad Rhenum: Guilielmus Broedelet.

Salmasius, Claudius. 1629. *Plinianae exercitationes in Caii Julii Solini Polyhistora. Item Caii Julii Solini Polyhistor ex veteribus libris emendatus*. Parisiis: apud Hieronymum Drouart.

Salmasius, Claudius. 1643. *De Hellenistica commentarius, controversiam de lingua Hellenistica decidens et plenissime pertractans originem et dialectos Graecae linguae*. Lugd[uni] Batavor[um]: ex officina Elseviriorum.

Sammes, Aylett. 1676. *Britannia Antiqua Illustrata: Or, the antiquities of ancient Britain, derived from the Phœnicians*. London: Tho. Roycroft.

Scaliger, Josephus Justus. 1579. *In Manilii libros astronomicon commentarius et castigationes*. Lutetiae: apud Mamertum Patissonium.

Scaliger, Josephus Justus. 1740. *Scaligerana, Thuana, Perroniana, Pithoeana, et Colomesiana. Ou remarques historiques, critiques, morales, et littéraires de Jos. Scaliger, J. Aug. de Thou, le Cardinal du Perron, Fr. Pithou, et P. Colomie's. Avec les notes de plusieurs savans*. Amsterdam: Covens & Mortier.

Schmitt, Rudiger. 1996. Zu den "Germánioi" bei Herodot. *Historische Sprachforschung* 109 (1). 45–52.

[Schotanus], [Christianus]. 1664. *Beschryvinge van de heerlyckheydt van Frieslandt tusschen 't Flie end de Lauwers, met nieuwe caerten van t landschap in't algemeen soo oud als nieuw*. [Franeker]; [Amsterdam]: excudit Johannes Wellens; Jacob van Meurs.

Schulte, Joseph Wilhelm. 1879. (2) Gothica Minora. *Zeitschrift für deutsches Altertum und deutsche Literatur* 23–24. 318–336.

Schultens, Albert. 1761. *Origines Hebraeae sive Hebraeae linguae antiquissima natura et indoles ex Arabiae penetralibus revocata*, 2nd edn. Lugduni Batavorum: Apud Samuelem et Joannem Luchtmans, et Joannem le Mair.

Schultze, Benjamin. 1747. [52. Continuation] Herrn Missionarii Schultzens zu Madras Diarium vom Jahr 1739. In Gotthilf August Francke (ed.), *Der Königl. Dänischen Missionarien aus Ost-Indien eingesandter Ausführlichen Berichten, Von dem Werck ihres Amts unter den Heyden*, 710–742. Halle: Waisenhaus.

Van der Stighelen, Katlijne. 1987. *Anna Maria van Schurman of "Hoe hooge dat een maeght kan in de konsten stijgen"*. Leuven: Universitaire Pers Leuven.

Van Hal, Toon. 2007. Joseph Scaliger, puzzled by the similarities of Persian and Dutch? *Omslag. Bulletin van de Universiteitsbibliotheek Leiden en het Scaliger Instituut*. 1–3.

Van Hal, Toon. 2010a. On "the Scythian Theory". Reconstructing the outlines of Johannes Elichmann's (1601/1602–1639) planned Archaeologia harmonica. *Language & History* 53 (2). 70–80.

Van Hal, Toon. 2010b. "Quam enim periculosa sit ea via...". Josephus Justus Scaliger's views on linguistic kinship. *Beiträge zur Geschichte der Sprachwissenschaft* 20 (1). 111–140.

Van Hal, Toon. 2010c. Vulcanius and his network of language lovers. *De literis et lingua Getarum sive Gothorum* (1597). In Hélène Cazes (ed.), *Bonaventura Vulcanius, Works and Networks* (Bruges 1538 – *Leiden 1614*), 387–401. Leiden & Boston: Brill.

Van Hal, Toon. 2011. The earliest stages of Persian-German language comparison. In Gerda Hassler (ed.), *History of linguistics 2008. Selected papers from the Eleventh International Conference on the History of the Language Sciences (ICHoLS XI), Potsdam, 28th August–2nd September 2008*, 147–165. Amsterdam & Philadelphia: Benjamins.

Van Hal, Toon. 2012. Linguistics *ante litteram*. Compiling and transmitting views on language diversity and relatedness before the nineteenth century. In Rens Bod, Jaap Maat, and Thijs Weststeijn (eds.), *The making of the humanities. From early modern to modern disciplines*, 37–53. Amsterdam: Amsterdam University Press.

Van Hal, Toon. 2015. Friedrich Gedike on why and how to compare the world's languages: A stepping stone between Gottfried Wilhelm Leibniz and Wilhelm von Humboldt? *Beiträge zur Geschichte der Sprachwissenschaft* 25 (1). 53–76.

Vater, Johann Severin. 1815. *Linguarum totius orbis index alphabeticus: quarum grammaticae, lexica, collectiones vocabulorum recensetur, patria significatur, historia adumbatur*. Berolini: In officina libraria Fr. Nicolai.

Verstegan, Richard. 1605. *A restitution of decayed intelligence in antiquities concerning the most noble and renovvmed English nation*. Antwerpen: by Robert Bruney.

Vitringa, Campegius. 1712. *Observationum sacrarum libri sex, in quibus de rebus varii argumenti, & utilissimae investigationis, critice ac theologice, disseritur; Sacrorum imprimis librorum loca multa obscuriora nova vel clariore luce perfunduntur*. Franequerae: ex officina Wibii Bleck.

Vogel, Jean Philippe. 1936. Joan Josua Ketelaar of Elbing, author of the first Hindūstānī grammar. *Bulletin of the School of Oriental and African Studies* 8 (2–3). 817–822.

Vulcanius, Bonaventura. 1597. *De literis et lingua Getarum sive Gothorum. Item de notis Lombardicis. Quibus accesserunt specimina variarum linguarum, quarum indicem pagina quae praefationem sequitur ostendit*. Lugduni Batavorum: ex officina Plantiniana, apud Franciscum Raphelengium.

Walsh, Thomas. 1907. Ange de Saint Joseph. *The Catholic encyclopedia* 1. http://www.newadvent.org/cathen/01476b.htm (accessed 31 December 2016).

Walton, Brian. 1673. *Biblicus apparatus chronologico-topographico-philologicus. [. . .]*. Tiguri: Ex Typographeo Bodmeriano.

Warnerus, Levinus. 1644. *Proverbiorum et sententiarum Persicarum centuria collecta*. Lugduni Batavorum: Joannis Maire.

Wellern, M. J. G. 1753. Gesammelte Spuren von dem Namen Germanus außer Deutschland. In *Das Neueste aus der anmuthigen Gelehrsamkeit*, 325–345. Leipzig: Breitkopf.

Wiesehöfer, Josef. 2006. Karmania. In Hubert Cancik and Helmuth Schneider (eds.), *Der Neue Pauly*. Brill Online. http://referenceworks.brillonline.com/entries/der-neue-pauly/karmania-e609260 (accessed 31 December 2016).

Windfuhr, Gernot L. 1979. *Persian grammar. History and state of its study*. Den Haag, Paris & New York: Mouton.

Wottonius, Guilielmus. 1715. Oratio dominica in diversas omnium fere gentium linguas versa et propriis cujusque linguae characteribus expressa. Una cum dissertationibus nonnullis de linguarum origine variisque ipsarum permutationibus. In Johannes Chamberlaynius (ed.), *Dissertatio de confusione linguarum Babylonica*, 37–75. Amstelaedami: Typis Guilielmi & Davidis Goerei.

Shinji Ido

2 *Huihuiguan zazi*: A New Persian glossary compiled in Ming China

Abstract: *Huihuiguan zazi*, a New Persian glossary compiled in China during the Ming period (1368–1644), has been largely neglected in the linguistic study of Persian despite its obvious importance as a source of data on the historical development of New Persian. In this article, all entries in one particular manuscript of *huihuiguan zazi* are tabulated and supplemented with translations and transcriptions, thus rendering the linguistic information contained in the glossary easily accessible to linguists.

Keywords: *huihuiguan zazi*, Chinese, Persian

Huihuiguan zazi (lit. '*Huihuiguan*[1] literacy primer'), a New Persian[2] glossary compiled in China during the Ming period (1368–1644), has been largely neglected in the linguistic study of Persian despite its obvious importance as a source of data on the historical development of New Persian.[3] In this article, all entries in one particular manuscript of *huihuiguan zazi* (hereafter abbreviated as *zazi*) are tabulated and supplemented with translations and transcriptions, thus rendering the linguistic information contained in the glossary easily accessible to linguists. The entries are reproduced in typescript in Table 1.

    Zazi comprises hundreds of New Persian lexical items and their equivalents in Chinese. It thus offers a unique insight into the historical lexicology of New Persian as well as that of Chinese. Besides its value as a source of lexicological data, *zazi* has a different kind of value that derives from the script in which it presents New Persian lexical items; *zazi* presents New Persian lexical items in Chinese script, which, unlike Arabic script, does not (or rather cannot) dispense

---

1 *Huihuiguan* 'office for/of Islamic states' was the division that was in charge of Persian translation within the Ming dynasty *Siyiguan* 'Four Barbarians' Office'.

2 In the present paper, the term "New Persian" is used somewhat loosely in general reference to the Persian language after the Islamic conquest (see Paul 2013 and Utas 2006: 244–245) for issues of periodization of New Persian) while the present-day varieties of New Persian such as Tajik and Afghan Dari are referred to by their respective names.

3 The reader is referred to Honda (1963) and Liu (2008) for more detailed philological explanations of the glossary.

Shinji Ido, Nagoya University

with the representation of vowels. *Zazi* hence serves as a rare source of data on the historical vowel phonology of New Persian.

Surviving manuscripts and printed copies of *zazi* are not uniform in their contents and formats (see Honda 1963: 2, 56–58). There are basically two types of *zazi*; type 1 *zazi* lists New Persian lexical items in both their original forms (i.e., in Arabic script) and in their transcribed forms (i.e., in Chinese script), whereas type 2 *zazi* dispenses with the former and is written entirely in Chinese script. The two types also differ in the entries they comprise (see Honda 1963: 2–3). The manuscript whose entries are reproduced in this article is the Berlin Library Manuscript (hereafter Berlin Manuscript), which is a type 1 *zazi* manuscript. The manuscript contains a total of 773 entries, all of which are reproduced in Table 1. Table 1 is supplemented with four entries (531st to 534th entries in Table 1), which are absent in the Berlin Manuscript but are present in other copies of type 1 *zazi*.[4]

This article follows Ido (2015: 102–104) in identifying the variety of New Persian documented in type 1 *zazi* as an early fifteenth-century variety of New Persian that had currency in the Timurid court in Samarkand. Accordingly, the New Persian variety whose words are recorded in the Berlin Manuscript will be referred to simply as Timurid Persian in the remainder of this article. This article also follows Ido (2015: 107) in assuming that the Chinese-script transcription of Timurid Persian words in type 1 *zazi* is based on the Ming-period Chinese dialect of Beijing, where *Siyiguan* (lit. 'Four barbarians' office'), the bureau of translation responsible for the compilation of type 1 *zazi*, was situated.

The Berlin Manuscript divides, as do other copies of type 1 *zazi*, its entries into eighteen sections, whose headings may be translated into English as (the sections for) "astrology", "geography", "season/time", "person", "human", "body", "palace", "birds and beasts", "flowers and trees", "utensils", "clothing", "eating and drinking", "treasure", "voice and countenance", "literature and history", "four quarters", "amount/number", and "currency", respectively.[5]

Each entry in the Berlin Manuscript consists of a Timurid Persian lexical item written in Arabic script, its equivalent in Chinese, and the Chinese-script

---

4 The Berlin Manuscript is unique among various copies of type 1 *zazi* in comprising an appendix that contains 223 entries (Honda 1963: 2). The appendix is not reproduced in this paper. The Berlin Manuscript is available online on the website of the Staatsbibliothek zu Berlin (Siyiguan 2013 [1579]). This article concerns the section titled "6. Persisch (Suppl. Vol. 12; Texte Vol. 23)". The appendix, on the other hand, is indexed as "12. Persisch (Suppl. zu Vol. 6)".

5 The section headings are 天文門, 地理門, 時令門, 人物門, 人事門, 身體門, 宮室門, 鳥獸門, 花木門, 器用門, 衣服門, 飲食門, 珍寶門, 聲色門, 文史門, 方隅門, 數目門, and 通用門, respectively.

transcription of the Timurid Persian lexical item. See, for example, the entry for مغول (91st entry in Table 1) reproduced below in typescript.

مغول	Timurid Persian
饝靼	Ming-period Beijing Chinese
卯斡勒	Chinese-script transcription

In this entry, the Timurid Persian word مغول is followed by its Ming-period Beijing Chinese equivalent, namely 饝靼, which in turn is followed by 卯斡勒, a string of Chinese glyphs used as syllabic phonograms to jointly represent the Timurid Persian pronunciation of مغول.

In Table 1, each entry in the Berlin Manuscript is assigned a unique number and is supplemented with the following: an English equivalent of the Chinese word, a reconstructed Ming-period Beijing pronunciation of the Chinese-script transcription, the present-day Standard Chinese pronunciation of the Chinese-script transcription, and the present-day Tajik word that appears to be cognate with (or, in the case of an entry word that is a loanword, the Tajik form of) the Timurid Persian word. The order in which these items appear in Table 1 is as follows.

1. 91   Entry number
2. مغول   Timurid Persian word
3. 饝靼   Ming-period Beijing Chinese equivalent of 2
4. 'post-imperial Mongol(ian)'   English equivalent of 3
5. 卯斡勒   Chinese-script transcription of 2
6. /muʋɒ.•.lɛ/   Reconstructed Ming-period Beijing Chinese pronunciation of 5
7. /mɑu.wo.lɤ/   Present-day Standard Chinese pronunciation of 5
8. муғул   Present-day Tajik word that corresponds to, or is cognate with, 2

In Table 1, these items are arranged horizontally, thus:

Column	1.	2.	3.	4.	5.	6.	7.	8.
Entry	91	مغول	饝靼	post-imperial Mongol(ian)	卯斡勒	muʋɒ.•.lɛ	mɑu.wo.lɤ	муғул

Notes on each of columns 1 to 8 follow.

1. Entry numbers in Table 1 correspond with those used by Honda (1963), and hence also with those by Liu (2008), who adopts Honda's numbering.

2. In preparing Table 1, attention was paid not to "correct" the handwritten Timurid Persian words to the current New Persian orthography. In other words, I tried to replicate in Table 1 (as faithfully as the availability of letters and symbols in the font allowed) the handwritten Timurid Persian words as they appear in the manuscript.[6] As a result, Table 1 preserves "deviant" spellings (which may reflect some phonological characteristics of Timurid Persian), misspellings, and genuine slips of the brush found in the manuscript. For instance, the 309th entry word in the Berlin Manuscript, namely اندان, is presented in Table 1 as اندان, though it is apparently cognate with present-day New Persian دندان *dandān*[7] 'tooth'. Similarly, the 732nd entry word in the Berlin Manuscript, namely كدشتن, is reproduced "as is" in Table 1, despite the spelling of present-day New Persian گذشتن *gozaštan*[8] 'to pass'.[9]

   For the simple reason that we know little about Timurid Persian phonology, no phonological representations of Timurid Persian words are provided in Table 1.

3. Where discrepancies between the assumed meaning of a Timurid Persian lexical item and that of its Chinese equivalent are large, such as in the 396th entry where Timurid Persian جانور, an apparent cognate of contemporary New Persian جانور *jān(e)var*[10] 'animal', and Chinese 鶯 'oriole' are shown to be each other's equivalents, they are mentioned in notes.[11]

---

6 The reader is recommended to consult the original manuscript (Siyiguan 2013 [1579]) to examine their forms in handwriting.

7 For simplicity, only the Iranian Persian romanization of دندان is provided here. The same word is romanized as *dandaan* in Sayd's (2009: 153) Dari romanization. The Tajik equivalent of the word is *дандон*.

8 For simplicity, only the Iranian Persian romanization of گذشتن is provided here. The same word is romanized as *gozashtan* in Sayd's (2009: 247) Dari romanization. The Tajik equivalent of the word is *гузаштан*.

9 These examples testify to orthographical differences between Timurid Persian and the New Persian of today, with one salient difference being the absence of the letter گ in the former. This may not be entirely surprising; Kuroyanagi (1984: 98) observes that the invention of گ postdates that of چ, پ, and ژ (which, according to Kuroyanagi, already existed in the late thirteenth century) by several centuries. While the Berlin Manuscript contains no instance of گ, it does contain one instance of the letter گ in the 330th entry word of the manuscript, namely گردن. Incidentally, the use of گ is less restricted in the two Beijing Library manuscripts reproduced in Beijing tushuguan guji chuban bianji zu (1987–[1994]: 465–572), in which there are multiple instances of the letter.

10 For simplicity, only the Iranian Persian romanization of جانور is provided here. The same word is romanized as *jaan(a)war* in Sayd's (2009: 94) Dari romanization. The Tajik equivalent of the word is *ҷон(а)вар*.

11 Such discrepancies do not have to result from mistranslation. At least theoretically, they can be ascribed to "false friendship" in Timurid Persian and contemporary New Persian or that in Ming-period Beijing Chinese and present-day Standard Chinese, as there is no guarantee that جانور meant 'animal' among Timurids in Samarkand in the early fifteenth century.

4. For translating Chinese words into English, I consulted Tōdō and Kanō (2005) and Kleeman and Yu (2010). An effort was made to present the meanings that the Chinese words had in the early Ming period, during which period type 1 *zazi* was likely first compiled (Ido 2015: 100–104). For example, the word 韃靼 in the example shown above, which is currently used primarily in reference to 'Tatar' in Standard Chinese, meant 'post-imperial Mongol(ian)' in Ming China, hence the English translation.[12]

5. Some Chinese-script transcriptions in the Berlin Manuscript contain obvious slips of the brush. Such scribal errors are found in entries 166, 175, 415, 522, 678, and 755, and are pointed out in notes. There is also one illegible glyph in the 270th entry, which is also noted in a note.

6. The reconstructed Ming-period Chinese pronunciations presented in this column are based on Lu's (1988) reconstruction of the readings of the Chinese glyphs contained in *dengyun tujing*, a set of rhyme tables compiled in Ming China.[13] However, a number of Chinese glyphs occur in *zazi* for which their Ming-period readings are not retrievable from *dengyun tujing*. The symbol • in this column serves as a placeholder for such glyphs. Thus, the Ming-period Beijing Chinese pronunciation of 卯斡勒 is shown to be /muʊʋ.•.lɛ/ in this column because the reading of the glyph 斡 is not retrievable from *dengyun tujing*.

   The glyph 阿, which appears in column 5, is shown to have the reading of "ɔ/ɑ" in this column. This is because the glyph likely had multiple readings in Ming-period Beijing (as in fact it still does in Standard Chinese today) and no good reason seems to exist for limiting its reading in *zazi* to one of /ɔ/ and /ɑ/.[14]

   Note that the reconstructed Ming-period Beijing Chinese pronunciations presented in this article are not beyond dispute, nor do they certainly represent the pronunciations that the compiler(s) of type 1 *zazi* used. For instance,

---

12 Many of the English translations supplied in this column abstract away from the distinction between different parts of speech because many Chinese words are not readily amenable to that distinction in the absence of context (see Norman 1998: 157). For example, the glyph 差 in the 227th entry of type 1 *zazi*, for which only 'send; differ' is shown in Table 1, may mean any of 'send', 'differ', 'different', 'difference', and 'somewhat', depending on the context in which it occurs.

13 Lu's reconstruction is used here with slight modifications made in accordance with Tōdō (1957: 104–108), Satoh (1981), and Ye (2001: 140–153). See Nagashima (1941: 23–24) for a more detailed philological explanation of *dengyung tujing*, which can be literally translated as 'illustrated book of rhyme classification'.

14 An analysis of *dengyun tujing* reveals multiple readings of some glyphs that occur in column 5, but, in this column, with the exception of 阿 (indicated in this section as ɔ/ɑ), only one reading is assigned to each of the glyphs. The reader is referred to Ido (2015: 117–118) for an explanation of how particular readings are selected for the glyphs.

法, 府, 番, and 夫, which typically appear in the Chinese-script transcriptions of Timurid Persian words comprising the letter ف, are respectively pronounced /fa/, /fu/, /fan/, and /fu/ (i.e., with the voiceless labio-dental fricative) in today's Standard Chinese, while the glyphs' reconstructed Ming-period Beijing Chinese pronunciations presented in this column are /puɑ/, /piu/, /puiɛn/, and /piu/, respectively. This may suggest that the reconstruction needs revision, and/or that considerable variation exists within what I simply refer to as Ming-period Beijing Chinese in this article.[15]

7. The present-day Standard Chinese pronunciations of the Chinese-script transcriptions in the Berlin Manuscript are supplied in Table 1 primarily to allow an inference as to how the syllables shown with • in column 6 (see above) may have been pronounced in Ming-period Beijing.[16] A number of glyphs have multiple readings in Standard Chinese, but in this column only one reading is assigned to each glyph. For example, I identify the reading of the glyph 都 not as *dōu* /tou/ but as *dū* /tu/ and that of 血 not as *xuè* /ɕɥe/ but as *xiě* /ɕje/ because the latter seem more in line with the reconstructed Ming-period Beijing Chinese readings of the glyphs.[17]

8. The Tajik data in this column are provided in the hope that they are useful in studying the phonology of Timurid Persian because Tajik is apparently more closely related to Timurid Persian than are other major varieties of contemporary New Persian such as Iranian Persian and Afghan Dari (in terms of the vowel system at any rate; see Ido [2015]). The symbol × used in this column indicates that I could not find a Tajik word that is in apparent correspondence with (or that is an apparent cognate of) the Timurid Persian entry in the published Tajik-language sources available to me.[18] English translations are provided in notes for Tajik words whose meanings differ markedly from those presented in column 4.

---

15 Alternatively, it could suggest variation within Timurid Persian or that Timurid Persian ف in *zazi* represents a bilabial plosive.

16 The IPA transcription in this section is based on Lin's (2007: 123–135, 283–292) phonetic description of the *pīnyīn* romanization system.

17 Similarly, 恁 is read not *nèn* /nən/ but *nín* /nin/ while 塞 is read not *sāi* /sai/ but *sè* /sɤ/ in this section.

18 The Tajik spellings of many archaic words are retrieved from Šukurov et al. (1969a, 1969b).

Table 1: *Huihuiguan zazi* (the Berlin Library Manuscript)

	天文門		The "astrology" section				
1	آسمان	天	sky	阿思媽恩	ɔ/ɑ.sɹ.muɑ.ən	a.sɹ.ma.ən	осмон
2	آفتاب	日	sun	阿夫他卜	ɔ/ɑ.piu.tʰɑ.•	a.fu.tʰa.pu	офтоб
3	ماه	月	moon	媽黑	muɑ.xɛ	ma.xei	моҳ
4	ستاره	星	star	洗他勒	si.tʰɑ.lɛ	ɕi.tʰa.lɤ	ситора
5	ابر	雲	cloud	阿卜兒	ɔ/ɑ.•.•	a.pu.əɹ	абр
6	باد	風	wind	巴得	puɑ.tɛ	pa.tɤ	бод
7	باران	雨	rain	把剌恩	puɑ.•.ən	pa.la.ən	борон
8	شبنم	露	dew	舍卜南	•.•.nan	ʂɤ.pu.nan	шабнам
9	پشک	霜	frost	僕石克	•.•.•	pʰu.ʂɹ.kʰɤ	пашк
10	برف	雪	snow	百兒夫	puɛ.•.piu	pai.əɹ.fu	барф
11	رعد	雷	thunder	勒阿得	lɛ.ɔ/ɑ.tɛ	lɤ.a.tɤ	раъд
12	برق	電	lightning	百兒革	puɛ.•.kɛ	pai.əɹ.kɤ	барқ
13	قوس قزخ	虹	rainbow	高思古則黑	kɒʊ.sɹ.ku.tsɛ.xɛ	kɑu.sɹ.ku.tsɤ.xei	қавси қузаҳ
14	بنات النعش	斗	Plough (Big Dipper)	百納塔納阿石	puɛ.nɑ.tʰun.nɑ.ɔ/ɑ.•	pai.na.tʰwən.na.a.ʂɹ	Банотуннаъш
15	بخار	煙	smoke, vapour	卜哈兒	•.xɑ.•	pu.xa.əɹ	бухор
16	غبار	霧	fog, mist	五巴兒	•.puɑ.•	wu.pa.əɹ	ғубор
17	يخ	氷	ice	夜黑	iɛ.xɛ	je.xei	ях
18	يخچه	雹	hail	夜黑徹	iɛ.xɛ.•	je.xei.tʂʰɤ	яхча
19	صاعقه	霆	thunderbolt	撒額革	sɑ.•.kɛ	sa.ɤ.kɤ	соиқа
20	آتش	火	fire	阿弐石	ɔ/ɑ.tʰɛ.•	a.tʰɤ.ʂɹ	оташ
21	نور	光	light	奴兒	nu.•	nu.əɹ	нур
22	سايه	影	shadow	撒夜	sɑ.iɛ	sa.je	соя
23	روشن	明	clear, bright	羅山	luɔ.ʂan	lwo.ʂan	равшан
24	تاریک	暗	dark	他列克	tʰɑ.liɛ.•	tʰa.lje.kʰɤ	торик
25	باد صبا	東風	east wind; spring wind	巴得塞巴	puɑ.tɛ.sɛ.puɑ	pa.tɤ.sɤ.pa	боди сабо
26	باد سموم	薰風	south wind; early summer breeze	巴得塞木恩	puɑ.tɛ.sɛ.•.ən	pa.tɤ.sɤ.mu.ən	боди самум[19]
27	باد دبور	金風	autumn breeze	巴得得卜兒	puɑ.tɛ.tɛ.•.•	pa.tɤ.tɤ.pu.əɹ	боди дабур[20]
28	باد صايم	朔風	north wind	巴得撒因	puɑ.tɛ.sɑ.in	pa.tɤ.sa.jin	х[21]
29	داره	日運[22]	solar halo	打勒	tɑ.lɛ	ta.lɤ	х
30	هاله	月運[23]	lunar halo	哈勒	xɑ.lɛ	xa.lɤ	ҳола
31	بدر	圓月	full moon	百得兒	puɛ.tɛ.•	pai.tɤ.əɹ	бадр
32	محاق	殘月	morning moon	母哈革	mu.xɑ.kɛ	mu.xa.kɤ	муҳоқ / маҳоқ / миҳоқ
33	کسوف	日蝕	solar eclipse	苦蘇夫	kʰu.su.piu	kʰu.su.fu	кусуф
34	خسوف	月蝕	lunar eclipse	虎蘇夫	xu.su.piu	xu.su.fu	хусуф
35	ژاله	霖雨	long-continued rain	惹勒	ʐɛ.lɛ	ɹɤ.lɤ	жола

19 Modern Tajik *боди самум* 'hot, harmful desert wind'.
20 Modern Tajik *боди дабур* 'west wind'.
21 Timurid Persian باد صایم would be written in Tajik as *боди соим* lit. 'faster's wind', which, however, seems to lack direct semantic correspondence with 朔風 'north breeze'.
22 This is actually the meaning of 日暈 *rìyùn*, with which 日運 is homophonous (in present-day Standard Chinese).
23 This is actually the meaning of 月暈 *yuèyùn*, with which 月運 is homophonous (in present-day Standard Chinese).

36	ثابتات	雜星	miscellaneous stars	撒必他忒	sɑ.pi.tʰa.tʰɛ	sa.pi.tʰa.tʰɤ	собитот
37	سيارات	七政	sun, moon, Mercury, Mars, Venus, Jupiter, and Saturn	塞呀剌忒	sɛ.•.•.tʰɛ	sɤ.ja.la.tʰɤ	саёрот[24]
38	صبح صادق	天曉	dawn	速卜黑撒的革	•.•.xɛ.sɑ.ti.kɛ	su.pu.xei.sa.ti.kɤ	субҳи содиқ
39	هوا	天氣	weather	黑洼	xɛ.•	xei.wa	ҳаво
40	انجلا	復圓	fourth contact (astronomy)	尹知剌	•.•.•	jin.tʂ.la	инчило[25]

	地理門		The "geography" section				
41	كوه	山	mountain	科黑	kʰuɔ.xɛ	kʰɤ.xei	кӯҳ
42	جوی	河	stream; river; Yellow River	卓衣	tʂuɔ.i	tʂwo.ji	чӯй[26]
43	رود	江	(large) river; Yangtze River	魯得	lu.tɛ	lu.tɤ	рӯд[27]
44	دريا	海	sea	得兒呀	tɛ.•.•	tɤ.əɹ.ja	дарё[28]
45	خاك	土	earth	哈克	xɑ.•	xa.kʰɤ	хок
46	زمين	地	ground	則米尹	tsɛ.mi.•	tsɤ.mi.jin	замин
47	آب	水	water	阿卜	ɔ/ɑ.•	a.pu	об
48	چشمه	泉	fountain; spring	扯石黙	tʂʰɛ.•.•	tʂʰɤ.ʂ.mwo	чашма
49	خانبالغ[29]	京	capital (city)	罕巴力額	xan.puɑ.•.•	xan.pa.li.ɤ	×
50	مملكت	國	country; state	滿剌克忒	muan.•.•.tʰɛ	man.la.kʰɤ.tʰɤ	мамлакат
51	شهر	城	city; town	舍黑兒	•.xɛ.•	ʂɤ.xei.əɹ	шаҳр
52	نواحی	境	border; region	納洼黑	nɑ.•.xɛ	na.wa.xei	навоҳӣ
53	روستا	村	village	羅思他	luɔ.sɹ.tʰa	lwo.sɹ.tʰa	русто
54	بيابان	野	wide empty plains; field	比呀巴恩	pi.•.puɑ.ən	pi.ja.pa.ən	биёбон[30]
55	زراعت	田	field; farmland	即剌額忒	•.•.•.tʰɛ	tɕi.la.ɤ.tʰɤ	зироат[31]
56	باغ	園	garden	巴額	puɑ.•	pa.ɤ	боғ
57	كرد	塵	dust, dirt	革兒得	kɛ.•.tɛ	kɤ.əɹ.tɤ	гард
58	ريك	沙	sand	列克	liɛ.•	lje.kʰɤ	рег
59	سنگ	石	stone	桑克	sɑŋ.•	saŋ.kʰɤ	санг
60	راه	路	road	剌黑	•.xɛ	la.xei	роҳ / раҳ
61	بازار	市	market	把咱兒	puɑ.•.•	pa.tsa.əɹ	бозор
62	چاه	井	well	叉黑	tʂʰa.xɛ	tʂʰa.xei	чоҳ / чаҳ
63	تواره	籬	hedge	忒洼勒	tʰɛ.•.lɛ	tʰɤ.wa.lɤ	тавора / тувара
64	ديوار	墻	wall; fence	迭洼兒	•.•.•	tje.wa.əɹ	девор
65	عقبه	嶺	ridge; mountain	阿革百	ɔ/ɑ.kɛ.puɛ	a.kɤ.pai	ақаба
66	غار	洞	cave	阿兒	ɔ/ɑ.•	a.əɹ	ғор
67	خوض	潭	deep water/pool	蒿子	xɒʊ.tsɹ	xɑu.tsɹ	ҳавз

24 Modern Tajik *саёрот* 'planets' appears in Odilov (1974: 82).

25 Modern Tajik *инчило* 'light/bright/manifestation' does not specifically refer to fourth contact.

26 Modern Tajik *чӯй* means 'brook; stream' rather than 'river'.

27 Modern Tajik *рӯд* 'canal; river' is not used in reference to large rivers.

28 Modern Tajik *дарё* 'river'.

29 خانبالغ (modern Beijing) is variously spelt Khan-baliq (Franke 1966: 57), Ḵẖānbaliḳ (Barthold 1987: 898), Khanbaligh (Atwood 2004: 123), etc., in English.

30 Modern Tajik *биёбон* 'desert'.

31 Modern Tajik *зироат* 'agriculture'.

68	چول	川	river; plain	搠勒	•.lɛ	ʂwo.lɣ	чӯл[32]
69	جویچه	溝	ditch	卓衣徹	tʂuɔ.i.•	tʂwo.ji.tʂʰɣ	чӯйча
70	کدرکاه	渡	cross	古得兒噶黑	ku.tɛ.•.•.xɛ	ku.tɣ.əɹ.ka.xei	гузаргоҳ
71	لب جوی	岸	bank; shore	勒比卓衣	lɛ.pi.tʂuɔ.i	lɣ.pi.tʂwo.ji	лаби чӯй
72	دوتاه	徑	path	堵他黑	•.tʰɑ.xɛ	tu.tʰa.xei	дутоҳ[33]
73	مزار	墳	grave	默咱兒	•.•.•	mwo.tsa.əɹ	мазор
74	موج	潮	tide	卯知	mʊɒ.•	mɑu.tʂɹ	мавҷ[34]
75	بیخان	庄	manor, village	擺哈恩	puai.xɑ.ən	pai.xa.ən	×
76	جهان	世	world	者哈恩	tʂɛ.xɑ.ən	tʂɣ.xa.ən	ҷаҳон
77	جنکل	林	forest	展革勒	tʂan.kɛ.lɛ	tʂan.kɣ.lɣ	ҷангал
78	معدن	鑛	ore; mineral deposit	母阿定	mu.ɔ/ɑ.tiŋ	mu.a.tjəŋ	маъдан / маъдин
79	حضیر	街	street	黑雖兒	xɛ.suei.•	xei.swei.əɹ	×
80	کل	泥	mud	吉勒	•.lɛ	tɕi.lɣ	гил
81	تر	濕	wet	忒兒	tʰɛ.•	tʰɣ.əɹ	тар
82	خشک	乾	dry	戶石克	xu.•.•	xu.ʂɻ.kʰɣ	хушк
83	مغاک	深	deep; depth	母阿克	mu.ɔ/ɑ.•	mu.a.kʰɣ	мағок
84	پایاب	淺	shallow	吧呀卜	pʰuɑ.•.•	pʰa.ja.pu	поёб
85	خندق	城壕	city moat	罕得革	xan.tɛ.kɛ	xan.tɣ.kɣ	хандақ
86	میدان	教場	square for training/reviewing troops	買搭恩	muai.tɑ.ən	mai.ta.ən	майдон[35]
87	دروازه	關廂	area outside of a city gate	得兒洼則	tɛ.•.•.tsɛ	tɣ.əɹ.wa.tsɣ	дарвоза
88	دهانه	關口	pass; juncture	得哈納	tɛ.xɑ.nɑ	tɣ.xa.na	даҳона
89	مسلمان	回回	Islam(ic)	母蘇里媽恩	mu.su.•.mʊɑ.ən	mu.su.li.ma.ən	мусулмон
90	ترکی	高昌	Gaochang/Qara-hoja	土兒期	tʰu.•.•	tʰu.əɹ.tɕʰi	туркӣ[36]
91	مغول[37]	韃靼	post-imperial Mongol(ian)	卯幹勒	mʊɒ.•.lɛ	mɑu.wo.lɣ	муғул
92	جورجی	女直	Jurchen	卓兒知	tʂuɔ.•.•	tʂwo.əɹ.tʂɹ	×
93	تبت	西番	(Kham) Tibetans	土百忒	tʰu.puɛ.tʰɛ	tʰu.pai.tʰɣ	Тибет[38]
94	قریانی	雲南	Yunnan	古兒呀你	ku.•.•.ni	ku.əɹ.ja.ni	×
95	کنجانفو[39]	陝西	Shaanxi	欽張夫	kʰin.•.piu	tɕʰin.tʂaŋ.fu	×
96	تنغوت[40]	河西	west of the Yellow River	湯屋忒	tʰɑŋ.•.tʰɛ	tʰɑŋ.wu.tʰɣ	×

---

32 Modern Tajik *чӯл* 'steppe', which is a Mongolian loan; does not mean 'river'.

33 Modern Tajik *дутоҳ* 'bent; curved'.

34 Modern Tajik *мавҷ* 'wave'.

35 Modern Tajik *майдон* 'public square'.

36 Modern Tajik *туркӣ* 'Turkic' does not refer specifically to 'Gaochang/Qara-hoja'.

37 韃靼 *Dádá*, which is currently used in reference to 'Tatar', was used in the Ming to refer to (post-imperial Mongolian) Northern Yuan, hence the entry word and Tajik *муғул* 'Mongol'.

38 Present-day Tajik *Тибет* 'Tibet' is probably a loanword from Russian.

39 کنجانفو has been identified in the literature as reflecting the pronunciation of 京兆府 *Jīngzhàofǔ* or 咸陽府 *Xiányángfǔ* (see Haw 2014: 7).

40 Haw (2014: 22) identifies "the *Hexi* 河西 region" as "[t]he Tangut region, that is, the former Xi Xia 西夏 state".

	時令門		The "season/time" section				
97	سال	年	year	撒勒	sɑ.lɛ	sa.lɣ	сол
98	ماه	月	month	媽黑	muɑ.xɛ	ma.xei	моҳ / маҳ
99	روز	日	day	羅子	luɔ.tsɹ	lwo.tsɹ	рӯз
100	ساعت	時	hour	撒額弍	sɑ.•.tʰɛ	sa.ɣ.tʰɣ	соат
101	بهار	春	spring	卜哈兒	•.xɑ.•	pu.xa.əɹ	баҳор
102	تابستان	夏	summer	他比思他恩	tʰɑ.pi.sɹ.tʰɑ.ən	tʰa.pi.sɹ.tʰa.ən	тобистон
103	تیرماه	秋	autumn	體兒媽黑	tʰi.•.muɑ.xɛ	tʰi.əɹ.ma.xei	тирамоҳ
104	زمستان	冬	winter	即米思他恩	•.mi.sɹ.tʰɑ.ən	tɕi.mi.sɹ.tʰa.ən	зимистон
105	بامداد	早	morning	榜搭得	puaŋ.ta.tɛ	pɑŋ.ta.tɣ	бомдод
106	شبانگاه	晚	evening	舍榜噶黑	•.puaŋ.•.xɛ	ʂɣ.pɑŋ.ka.xei	шабонгоҳ / шабонгаҳ
107	اجتماع	朔	first day of a lunar month	以知體媽額	•.•.tʰi.muɑ.•	ji.tʂɹ.tʰi.ma.ɣ	иҷтимоъ[41]
108	استقبال	望	fifteenth day of a lunar month	以思體革巴勒	•.sɹ.tʰi.kɛ.puɑ.lɛ	ji.sɹ.tʰi.kɣ.pa.lɣ	истиқбол[42]
109	سرما	寒	cold	塞兒媽	sɛ.•.muɑ	sɣ.əɹ.ma	сармо
110	گرما	熱	hot	革兒媽	kɛ.•.muɑ	kɣ.əɹ.ma	гармо
111	تیره	陰	cloudy	體勒	tʰi.lɛ	tʰi.lɣ	тира
112	صاف	晴	clear, fine (weather)	撒夫	sɑ.piu	sa.fu	соф
113	نیم شب	子	11 pm–1 am	恁舍卜	nin.•.•	nin.ʂɣ.pu	нимшаб
114	سحر	丑	1 am–3 am	塞黑兒	sɛ.xɛ.•	sɣ.xei.əɹ	саҳар
115	پگاه	寅	3 am–5 am	僕噶黑	•.•.xɛ	pʰu.ka.xei	пагоҳ / пагаҳ
116	ضخوه[43]	卯	5 am–7 am	祖黑幹	•.xɛ.•	tsu.xei.wo	×
117	اژدر	辰	7 am–9 am	阿日得兒	ɔ/ɑ.zɻ.tɛ.•	a.ɻɹ.tɣ.əɹ	аждар
118	چاشتگاه	巳	9 am–11 am	叉石弍噶黑	tʂʰɑ.•.tʰɛ.•.xɛ	tʂʰa.ʂɹ.tʰɣ.ka.xei	чоштгоҳ
119	استوا	午	11 am–1 pm	以思體洼	•.sɹ.tʰi.•	ji.sɹ.tʰi.wa	истиво
120	بیشین	未	1 pm–3 pm	撒石尹	pʰiɛ.•.•	pʰje.ʂɹ.jin	пешин
121	دیگر	申	3 pm–5 pm	底革兒	ti.kɛ.•	ti.kɣ.əɹ	дигар
122	آفتاب فرورفتن	酉	5 pm–7 pm	阿夫他卜府羅勒夫飡	ɔ/ɑ.piu.tʰɑ.•.piu.luɔ.lɛ.piu.tʰan	a.fu.tʰa.pu.fu.lwo.lɣ.fu.tʰan	офтоб фурӯ рафтан
123	شام	戌	7 pm–9 pm	沙恩	ʂɑ.ən	ʂa.ən	шом
124	خفتن	灾	9 pm–11 pm	虎夫飡	xu.piu.tʰan	xu.fu.tʰan	хуфтан
125	معتدل	温	warm	母阿弍的勒	mu.ɔ/ɑ.tʰɛ.ti.lɛ	mu.a.tʰɣ.ti.lɣ	мӯътадил
126	فسردن	凍	freezing	非思反丹	•.sɹ.•.tan	fei.sɹ.əɹ.tan	фусурдан
127	سال عشرت	稔年	year of bumper harvest	撒勒額石勒弍	sɑ.lɛ.•.•.lɛ.tʰɛ	sa.lɣ.ɣ.ʂɹ.lɣ.tʰɣ	соли ишрат
128	سال قحطی	歉年	lean year	撒勒革黑推	sɑ.lɛ.kɛ.xɛ.•	sa.lɣ.kɣ.xei.tʰwei	соли қаҳтӣ
129	آفت سماوی	水災	flood disaster	阿法梯塞媽弍	ɔ/ɑ.puɑ.tʰi.sɛ.muɑ.•	a.fa.tʰi.sɹ.ma.ɥy	офати самовӣ[44]
130	آفت خشک	旱災	drought	阿法梯戶石克	ɔ/ɑ.puɑ.tʰi.xu.•.•	a.fa.tʰi.xu.ʂɹ.kʰɣ	офати хушк
131	دینه	昨日	yesterday	底納	ti.nɑ	ti.na	дина

---

41 Modern Tajik *иҷтимоъ* 'society; gathering'.

42 Modern Tajik *истиқбол* 'festive welcome; future; prospects'.

43 According to Steingass (2012: 800), ضحوة means "[f]orenoon; luncheon-time".

44 Modern Tajik *офати самовӣ* lit. 'skyey/celestial disaster'.

132	فردا	明日	tomorrow	法兒搭	pua.•.ta	fa.ɹ.ta	фардо
133	هرروز	逐日	day by day; everyday	哈兒羅子	xɑ.•.luɔ.tsɿ	xa.ɹ.lwo.tsɿ	ҳар рӯз
134	چهار فصل	四季	four seasons	叉哈兒法思勒	tʂʰa.xɑ.•.pua.sɿ.lɛ	tʂʰa.xa.ɹ.fa.sɿ.lɤ	ч(ах)ор фасл
135	پنج عناصر	五行[45]	five elements	潘知額納速兒	pʰuan.•.•.nɑ.•.•	pʰan.tʂɿ.ɤ.na.su.əɹ	панч аносир
136	سال رونده	去年	last year	撒勒勒灣得	sa.lɛ.lɛ.•.tɛ	sa.lɤ.lɤ.wan.tɤ	соли раванда
137	سال آینده	来年	next year	撒勒阿言得	sa.lɛ.ɔ/a.iɛn.tɛ	sa.lɤ.a.jɛn.tɤ	соли оянда
	人物門		The "person" section				
138	بادشاه	君	monarch	吧得沙黑	pʰua.tɛ.ʂa.xɛ	pʰa.tɤ.ʂa.xei	подшоҳ
139	وزیر	臣	liege; vassal	我即兒	ɔ.•.•	wo.tɕi.əɹ	вазир
140	بیغامبر	聖	saint; sage; master	迫昂百兒	•.aŋ.puɛ.•	pʰwo.aŋ.pai.əɹ	пайғамбор / пайғомбор
141	حکیم	賢	wise; able person	黑期尹	xɛ.•.•	xei.tɕʰi.jin	ҳаким
142	امیر	官	bureaucrat; official	阿米兒	ɔ/a.mi.•	a.mi.əɹ	амир
143	نویسنده	吏	minor official	你傘得	ni.san.tɛ	ni.san.tɤ	нависанда
144	ایلجی	使	envoy	以里赤	•.•.•	ji.li.tʂʅ	элчй[46]
145	لشکر	軍	army	勒石克兒	lɛ.•.•.•	lɤ.ʂʅ.kʰɤ.əɹ	лашкар
146	جد	祖	ancestor; grandfather	折得	•.tɛ	tʂɤ.tɤ	чадд
147	پدر	父	father	迫得兒	•.tɛ.•	pʰwo.tɤ.əɹ	падар
148	مادر	母	mother	媽得兒	mua.tɛ.•	ma.tɤ.əɹ	модар
149	زن	妻	wife	簪	•	tsan	зан
150	دادر	兄	elder brother	打得兒	ta.tɛ.•	ta.tɤ.əɹ	додар
151	برادر	弟	younger brother	比剌則兒	pi.•.tsɛ.•	pi.la.tsɤ.əɹ	бародар
152	فرزند	子	child; son	法兒簪得	pua.•.•.tɛ	fa.əɹ.tsan.tɤ	фарзанд
153	دختر	女	woman; daughter	朵黑弍兒	tuɔ.xɛ.tʰɛ.•	two.xei.tʰɤ.əɹ	духтар
154	خویش	親	parent; relative	或石	xuɛ.•	xwo.ʂʅ	хеш
155	یار	朋	companion; friend	呀兒	•.•	ja.əɹ	ёр
156	خاوند	主	host; owner	哈灣得	xɑ.•.tɛ	xa.wan.tɤ	хованд
157	مهمان	客	guest	米黑媽恩	mi.xɛ.mua.ən	mi.xei.ma.ən	меҳмон
158	استاد	師	teacher; master	五思他得	•.sɿ.tʰa.tɛ	wu.sɿ.tʰa.tɤ	устод
159	شاکرد	徒	apprentice; pupil	沙吉兒得	ʂa.•.•.tɛ	ʂa.tɕi.əɹ.tɤ	шогирд
160	پیر	老	old	批兒	pʰi.•	pʰi.əɹ	пир
161	جوان	少	young; few	主洼恩	tʂu.•.ən	tʂu.wa.ən	чавон
162	توین[47]	僧	monk	脱因	tʰuɔ.in	tʰwo.jin	×
163	مغ	道	Taoist	木額	•.•	mu.ɤ	муғ[48]
164	کشاورز	農	agriculture; peasant	起沙幹兒子	kʰi.ʂa.•.•.tsɿ	tɕʰi.ʂa.wo.əɹ.tsɿ	кашоварз
165	سوداکر	商	trade; tradesman	嫂搭革兒	•.ta.kɛ.•	sɑu.ta.kɤ.əɹ	савдогар
166	طبیب	醫	doctor	弍比弍[49]	tʰɛ.pi.tʰɛ.	tʰɤ.pi.tʰɤ	табиб

---

45 五行 *wǔxíng* refers to the five elements in an ancient Chinese doctrine. The elements, namely wood, fire, earth, metal, and water, are "allotted" respectively to spring, summer, midsummer, autumn, and winter.

46 Clauson's (1972: 539) dictionary of pre-thirteenth-century Turkic has *élçi:* 'ambassador' as one of its entries.

47 Liu (2008: 93) considers توین to ultimately originate from Chinese 道人 *dàoren* 'Taoist priest'.

48 Modern Tajik *муғ* 'magus'.

49 This probably is a misspelt 弍比卜 (see Honda 1963: 9; Beijing tushuguan guji chuban bianji zu 1987–[1994]: 476, 528), whose reconstructed Ming-period Beijing Chinese pronunciation and Standard Chinese pronunciation would be /tʰɛ.pi.•/ and /tʰɤ.pi.pu/, respectively.

167	فال کوی	卜	fortune-telling	法勒鍋衣	pua.lɛ.kuɔ.i	fa.lɣ.kwo.ji	фолгӯй
168	پری	神	deity; spirit	迫里	•.•	pʰwo.li	пари
169	نیک مرد	仙	hermit; wizard	乜克黙兒得	miɛ.•.•.•.tɛ	mje.kʰɣ.mwo.əɹ.tɣ	некмард
170	بت	佛	Buddha; statue of Buddha	卜忒	•.tʰɛ	pu.tʰɣ	пут
171	دیو	鬼	ghost; devil	刁	•	tjau	дев
172	خرفهور	工	worker; craft(sman)	黑兒法幹兒	xɛ.•.pua.•.•	xei.əɹ.fa.wo.əɹ	хирфавор
173	شوی	夫	husband	朔衣	ʂuɔ.i	ʂwo.ji	шӯй / шӯ
174	آدمی	人	person; human being	阿得密	ɔ/ɑ.tɛ.mi	a.tɣ.mi	одамй
175	رعیت	民	people; subjects	勒勒夜忒⁵⁰	lɛ.lɛ.iɛ.tʰɛ	lɣ.lɣ.je.tʰɣ	раият
176	عمک	叔	third among brothers; uncle	阿黙克	ɔ/ɑ.•.•	a.mwo.kʰɣ	амак
177	دادرزاده	姪	nephew; niece	打得兒咱得	ta.tɛ.•.•.tɛ	ta.tɣ.əɹ.tsa.tɣ	додарзода
178	عمه	姑	mother-in-law; aunt	俺黙	an.•	an.mwo	амма
179	ینکه	嫂	elder brother's wife	眼革	•.kɛ	jɛn.kɣ	янга
180	کنیزک	婢	maidservant	克你則克	•.ni.tsɛ.•	kʰɣ.ni.tsɣ.kʰɣ	канизак
181	جاریه	妾	concubine	鮓兒夜	tʂa.•.iɛ	tʂa.əɹ.je	чория
182	همسایه	隣	neighbour	罕撒夜	xan.sɑ.iɛ	xan.sa.je	ҳамсоя
183	نبیره	孫	grandchild	納必勒	nɑ.pi.lɛ	na.pi.lɣ	набера
184	تو	你	you	禿	tʰu	tʰu	ту
185	من	我	I	鑾	muan	man	ман
186	وی	他	she/he; other; that	歪	uai	wai	вай
187	غلام	僕	servant	五剌恩	•.•.ən	wu.la.ən	ғулом
188	دبیر	秀士	man of knowledge and virtue	得必兒	tɛ.pi.•	tɣ.pi.əɹ	дабир
189	مبارز	将軍	general	母巴力子	mu.pua.•.tsɿ	mu.pa.li.tsɹ	мубориз⁵¹
190	هیزم زننده	樵人	woodcutter	血津則南忒	xiuɛ.tsin.tsɛ.nan.tɛ	ɕje.tɕin.tsɣ.nan.tɣ	ҳезум-зананда
191	ماهی گیر	漁人	fisherman	馬希几兒	mua.xi.•.•	ma.ɕi.tɕi.əɹ	моҳигир
192	طباخ	厨役	cook	忒巴黑	tʰɛ.pua.xɛ	tʰɣ.pa.xei	таббох
193	صیاد	獵人	hunter	塞呀得	sɛ.•.tɛ	sɣ.ja.tɣ	сайёд
194	نقاش	画士	painter	納噶石	nɑ.•.•	na.ka.ʂɹ	наққош
195	مطرب	樂人	musician	母忒力卜	mu.tʰɛ.•.•	mu.tʰɣ.li.pu	мутриб
196	فراست	相士	physiognomist	法剌塞忒	pua.•.sɛ.tʰɛ	fa.la.sɣ.tʰɣ	фиросат⁵²
197	ختایی	漢人⁵³	Han	黑他衣	xɛ.tʰɑ.i	xei.tʰa.ji	хитой
198	شبان	牧羊	shepherd sheep	鼠巴恩	•.pua.ən	ʂu.pa.ən	чӯпон
199	کلهبان	牧馬	herd horses	克勒巴恩	•.lɛ.pua.ən	kʰɣ.lɣ.pa.ən	га(л)лабон
200	کاوبان	牧牛	graze cattle	稿巴恩	kɔʊ.pua.ən	kɑu.pa.ən	говбон
201	فیلبان	牧象	tend elephants	非勒巴恩	•.lɛ.pua.ən	fei.lɣ.pa.ən	филбон
202	دزد	盗賊	robber	杜子得	tu.tsɿ.tɛ	tu.tsɹ.tɣ	дузд

50 This is probably a misspelt 勒額夜忒 (see Honda 1963: 9; Beijing tushuguan guji chuban bianji zu 1987–[1994]: 476, 529), whose reconstructed Ming-period Beijing Chinese pronunciation and Standard Chinese pronunciation would be /lɛ.•.iɛ.tʰɛ/ and /lɣ.ɣ.je.tʰɣ/, respectively.

51 Modern Tajik *мубориз* 'fighter'.

52 Modern Tajik *фиросат* 'comrehension; cleverness'.

53 Haw (2014: 22) writes that 漢人 *Hànrén* "included not just Chinese, but all the peoples who had been subjects of the Jin empire, including Jurchens and Khitans, among others" during the Yuan period. How or whether this is related to the placement of the entry for 漢人 in the "person" section, which consists mostly of occupation names and kinship terms, is unclear. Incidentally, 女直 *Nǚzhēn* 'Jurchen' is an entry (number 92) in the "geography" section.

		人事門	The "human" section				
203	دولت	福	good fortune	倒勒式	•.lɛ.tʰɛ	tɑu.lɤ.tʰɤ	давлат
204	عمر	壽	longevity	兀木兒	•.•.•	wu.mu.əɹ	умр
205	شاد	喜	delight; happy	沙得	ʂɑ.tɛ	ʂa.tɤ	шод
206	نشاط	樂	joyous	你沙式	ni.ʂɑ.tʰɛ	ni.ʂa.tʰɤ	нишот / нашот
207	جد	勤	deligent	只得	•.tɛ	tʂɻ.tɤ	чидд
208	كاهلی	懶	lazy	噶黑里	•.xɛ.•	ka.xei.li	коҳилӣ
209	دوست	愛	be fond of; cherish	多思式	tuɔ.sɻ.tʰɛ	two.sɻ.tʰɤ	дӯст
210	رحم	憐	pity; sympathize	勒罕	lɛ.xan	lɤ.xan	раҳм
211	برآمدن	出	go/come out	百剌黙丹	puɛ.•.•.tan	pai.la.mwo.tan	баромадан
212	درآمدن	入	enter	得剌黙丹	tɛ.•.•.tan	tɤ.la.mwo.tan	даромадан
213	دید	見	see	底丹	ti.tan	ti.tan	дидан
214	دانستن	知	know	打你思貪	tɑ.ni.sɻ.tʰan	ta.ni.sɻ.tʰan	донистан
215	جستن	尋	seek	主思貪	tʂɰ.sɻ.tʰan	tʂu.sɻ.tʰan	чустан
216	اندیشه	想	think; miss	俺迭舍	an.•.•	an.tje.ʂɤ	андеша
217	كار	事	matter	噶兒	•.•	ka.əɹ	кор
218	آموختن	學	learn; knowledge	阿母黑貪	ɔ/ɑ.mu.xɛ.tʰan	a.mu.xei.tʰan	омӯхтан
219	خاص	專	specific	哈思	xɑ.sɻ	xa.sɻ	хос
220	صدق	誠	sincere	遂得革	suei.tɛ.kɛ	swei.tɤ.kɤ	сидқ
221	انعام	恩	favour	尹阿恩	•.ɔ/ɑ.ən	jin.a.ən	инъом
222	تشریف	賞	reward	式石里夫	tʰɛ.•.•.piu	tʰɤ.ʂɻ.li.fu	ташриф
223	عرضه	奏	present a memorial to an emperor; play (music)	阿兒則	ɔ/ɑ.•.tsɛ	a.əɹ.tsɤ	арза
224	خواستن	討	ask for; send punitive expedition	花思貪	xuɑ.sɻ.tʰan	xwa.sɻ.tʰan	хостан
225	نیاز	拜	make obeisance	你呀子	ni.•.tsɻ	ni.ja.tsɻ	ниёз
226	عنایت	望	gaze; hope for	額納夜式	•.nɑ.iɛ.tʰɛ	ɤ.na.je.tʰɤ	иноят[54]
227	فرستادن	差	send; differ	法思你丹[55]	puɑ.•.sɻ.ni.tan	fa.əɹ.sɻ.ni.tan	фиристодан
228	دادن	與	give	打丹	tɑ.tan	ta.tan	додан
229	توانگر	富	abundant	土往革兒	tʰu.uɑŋ.kɛ.•	tʰu.waŋ.kɤ.əɹ	тавонгар
230	مهتر	貴	noble	米黑式兒	mi.xɛ.tʰɛ.•	mi.xei.tʰɤ.əɹ	меҳтар
231	فقیر	貧	poor; poverty; avaricious	法革兒	puɑ.kɛ.•	fa.kɤ.əɹ	фақир
232	كهتر	賤	humble; cheap	起黑式兒	kʰi.xɛ.tʰɛ.•	tɕʰi.xei.tʰɤ.əɹ	кеҳтар
233	انسانیت	仁	humanity; human	因撒你夜式	in.sɑ.ni.iɛ.tʰɛ	jin.sa.ni.je.tʰɤ	инсоният
234	مروت	義	justice; rightous	母魯幹式	mu.lu.•.tʰɛ	mu.lu.wo.tʰɤ	мурувват
235	آدب	禮	propriety	阿得卜	ɔ/ɑ.tɛ.•	a.tɤ.pu	одоб
236	خرد	智	wisdom	黑勒得	xɛ.lɛ.tɛ	xei.lɤ.tɤ	хирад
237	وفا	信	trust	我法	ɔ.puɑ	wo.fa	вафо
238	ظریف	清	clear; clean	則里夫	tsɛ.•.piu	tsɻ.li.fu	зариф
239	عزیزی	濁	turbid	額即即	•.•.•	ɤ.tɕi.tɕi	азизӣ[56]
240	حركات	動	move	黑兒噶式	xɛ.•.•.tʰɛ	xei.əɹ.ka.tʰɤ	ҳаракат

54 Modern Tajik *иноят* 'thoughtfulness; assistance; favour'.

55 This is probably a misspelt 法兒思他丹 (see Honda 1963: 11; Beijing tushuguan guji chuban bianji zu 1987–[1994]: 480, 533), whose reconstructed Ming-period Beijing Chinese pronunciation and Standard Chinese pronunciation would be /puɑ.•.sɻ.tʰɑ.tan/ and /fa.əɹ.sɻ.tʰa.tan/, respectively.

56 Modern Tajik *азизӣ* 'valuableness'.

241	سكنات	静	still; calm; motionless	塞克納式	sɛ.•.na.tʰɛ	sɤ.kʰɤ.na.tʰɤ	сукунат
242	كريستن	哭	cry	己里思貪	•.•.sɹ.tʰan	tɕi.li.sɹ.tʰan	гиристан
243	خنديدن	笑	laugh	罕底丹	xan.ti.tan	xan.ti.tan	хандидан
244	مكر	詐	deceive	黙克兒	•.•.•	mwo.kʰɤ.əɹ	макр
245	عام	愚	stupid	阿恩	ɔ/ɑ.ən	a.ən	ом
246	مزيدن	添	add; append	黙即丹	•.•.tan	mwo.tɕi.tan	мазидан
247	كم	減	reduce; subtract	堪	kʰan	kʰan	кам
248	صفت	誇	to be proud of; exaggerate	髓法式	suei.pua.tʰɛ	swei.fa.tʰɤ	сифат[57]
249	پند	勸	advise	潘得	pʰuan.tɛ	pʰan.tɤ	панд
250	مطيع	順	smooth, in order	母推額	mu.•.•	mu.tʰwei.ɤ	мутеъ
251	ياغى	逆	adverse; contrary	呀額	•.•	ja.ɤ	ёғӣ
252	حكم	斷	judge; cut off	戸坤	xu.kʰun	xu.kʰwən	ҳукм
253	خدك	疑	doubt	黑杜克	xɛ.tu.•	xei.tu.kʰɤ	хазук / хадук /худук[58]
254	خريدن	買	buy	黑里丹	xɛ.•.tan	xei.li.tan	харидан
255	فروختن	賣	sell	府羅黑貪	piu.luɔ.xɛ.tʰan	fu.lwo.xei.tʰan	фурӯхтан
256	بيا	来	come	必呀	pi.•	pi.ja	биё
257	برو	去	go	卜勞	•.lɒʊ	pu.lɑu	бирав / бурав
258	غضب	怒	anger; get angry	額則卜	•.tsɛ.•	ɤ.tsɤ.pu	ғазаб
259	تغريم	罰	punish	式額里尹	tʰɛ.•.•.•	tʰɤ.ɤ.li.jin	×
260	شكايت	怨	resent	石噶夜式	•.•.iɛ.tʰɛ	ʂɹ.ka.je.tʰɤ	шикоят
261	تقاضا	催	urge: hasten	式噶咱	tʰɛ.•.•	tʰɤ.ka.tsa	тақозо
262	مست	醉	drunk	黙思式	•.sɹ.tʰɛ	mwo.sɹ.tʰɤ	маст
263	بيدار	醒	awake; sober	別搭兒	piɛ.ta.•	pje.ta.əɹ	бедор
264	ماندهكى	勞	toil; fatigue	媽恩得几	muɑ.ən.tɛ.•	ma.ən.tɤ.tɕi	мондагӣ
265	عفو	赦	amnesty	阿夫	ɔ/ɑ.piu	a.fu	афв
266	بازداشتن	留	remain; retain	巴子打石貪	puɑ.tsɹ.tɑ.•.tʰan	pa.tsɹ.ta.ʂɹ.tʰan	боздоштан
267	تعليم	訓	teach; train	式阿里尹	tʰɛ.ɔ/ɑ.•.•	tʰɤ.a.li.jin	таълим
268	برخاستن	起	rise; raise	百兒哈思貪	puɛ.•.xɑ.sɹ.tʰan	pai.əɹ.xa.sɹ.tʰan	бархостан
269	باش	居	reside; be; occupy	巴石	puɑ.•	pa.ʂɹ	бош
270	مقبول	受	accept; receive	黙革□[59]勒	•.kɛ.□.lɛ	mwo.kɤ.□.lɤ	мақбул
271	دستكيرى	救	rescue	得思式己里	tɛ.sɹ.tʰɛ.•.•	tɤ.sɹ.tʰɤ.tɕi.li	дастгирӣ
272	سياست	刑	punishment	洗呀塞式	si.•.sɛ.tʰɛ	ɕi.ja.sɤ.tʰɤ	сиёсат
273	كشتن	殺	kill	苦石貪	kʰu.•.tʰan	kʰu.ʂɹ.tʰan	куштан
274	خواب ديدن	夢	dream	花卜底丹	xuɑ.•.ti.tan	xwa.pu.ti.tan	хоб дидан
275	خسبيدن	睡	sleep	虎思比丹	xu.sɹ.pi.tan	xu.sɹ.pi.tan	хусбидан / хуспидан
276	شناختن	認	recognize	石納黑貪	•.nɑ.xɛ.tʰan	ʂɹ.na.xei.tʰan	шинохтан
277	وعده	許	allow	我阿得	ɔ.ɔ/ɑ.tɛ	wo.a.tɤ	ваъда[60]
278	طلبيدن	請	ask for	式勒比丹	tʰɛ.lɛ.pi.tan	tʰɤ.lɤ.pi.tan	талбидан / талабидан
279	خواندن	讀	read	花恩丹	xuɑ.ən.tan	xwa.ən.tan	хондан
280	نظر	觀	watch	納則兒	nɑ.tsɛ.•	na.tsɤ.əɹ	назар
281	آه	嘆	sigh; lament	阿黑	ɔ/ɑ.xɛ	a.xei	оҳ
282	توانا	能	capable	土洼納	tʰu.•.nɑ	tʰu.wa.na	тавоно

57 Modern Tajik *сифат* 'quality; attribute; adjective'.

58 Modern Tajik *хазук* 'distressed'.

59 The white square □ here represents an illegible Chinese glyph. Judging from Honda (1963: 12), the glyph is probably 卜, whose Standard Chinese pronunciation is /pu/.

60 Modern Tajik *ваъда* 'promise'.

283	شادباش	謝	thank	沙得巴石	ʂɑ.tɛ.puɑ.•	ʂa.tɤ.pa.ʂɹ̩	шодбош
284	كناه	罪	guilt	古納黑	ku.nɑ.xɛ	ku.na.xei	гуноҳ / гунаҳ
285	كردن	為	do; carry out	克兒丹	kʰɤ.əɹ.tan	kʰɤ.əɹ.tan	кардан
286	آویختن	掛	hang	阿月黑貪	ɔ/ɑ.iuɛ.xɛ.tʰan	a.ɥe.xei.tʰan	овехтан
287	تجسس	查	investigate	忒折速思	tʰɛ.•.•.sɹ	tʰɤ.tʂɤ.su.sɹ̩	таҷассус
288	پشیمان	悔	regret	迫舍媽恩	•.•.muɑ.ən	pʰwo.ʂɤ.ma.ən	пушаймон
289	بازی	戲	play	巴即	puɑ.•	pa.tɕi	бозӣ
290	سازواری	和睦	concord; amity; harmonious	撒子洼力	sɑ.tsɹ.•.•	sa.tsɹ̩.wa.li	созворӣ / созгорӣ
291	آواز	聲音	sound	阿洼子	ɔ/ɑ.•.tsɹ	a.wa.tsɹ̩	овоз
292	تعلق	管束	keep control over	忒俺路革	tʰɛ.an.•.kɛ	tʰɤ.an.lu.kɤ	тааллуқ
293	تقضیر	怠慢	neglect (of duty)	忒革雖兒	tʰɛ.kɛ.suei.•	tʰɤ.kɤ.swei.əɹ	тақсир
294	دلالت	導引	guide	得剌勒忒	tɛ.•.lɛ.tʰɛ	tɤ.la.lɤ.tʰɤ	далолат
295	زیرک	聰明	clever; bright	即勒克	•.lɛ.•	tɕi.lɤ.kʰɤ	зирак
296	مشورت	商議	confer; discuss	默束勒忒	•.•.lɛ.tʰɛ	mwo.ʂu.lɤ.tʰɤ	машварат
297	زینهار	叮嚀	exhort	儘哈兒	tsin.xɑ.•	tɕin.xa.əɹ	зинхор
298	اختیار	選擇	choose	以黑體呀兒	•.xɛ.tʰi.•.•	ji.xei.tʰi.ja.əɹ	ихтиёр
299	خصومت	爭競	contend; compete	虎蘇默忒	xu.su.•.tʰɛ	xu.su.mwo.tʰɤ	хусумат
300	فسوس	欺凌	insult	府蘇思	piu.su.sɹ	fu.su.sɹ̩	фусӯс / фусус
	身體門		The "body" section				
301	سر	頭	head	塞兒	sɛ.•	sɤ.əɹ	сар
302	روی	面	face	羅衣	luɔ.i	lwo.ji	рӯй
303	ابرو	眉	eyebrow	阿卜羅	ɔ/ɑ.•.luɔ	a.pu.lwo	абрӯ / абру
304	موی	髮	hair	母衣	mu.i	mu.ji	мӯй / муй
305	چشم	眼	eye	徹深	•.•	tʂʰɤ.ʂən	чашм
306	كوش	耳	ear	鍋石	kuɔ.•	kwo.ʂɹ̩	гӯш
307	بینی	鼻	nose	比你	pi.ni	pi.ni	бинӣ
308	دهان	口	mouth	得哈恩	tɛ.xɑ.ən	tɤ.xa.ən	даҳан / даҳон
309	اندان	牙	tooth; posterior tooth	膽搭恩	•.tɑ.ən	tan.ta.ən	дандон
310	زبان	舌	tongue	則巴恩	tsɛ.puɑ.ən	tsɤ.pa.ən	забон
311	دل	心	heart	的勒	ti.lɛ	ti.lɤ	дил
312	شكم	腹	belly	石堪	•.kʰan	ʂɹ̩.kʰan	шикам / ишкам
313	دست	手	hand	得思忒	tɛ.sɹ.tʰɛ	tɤ.sɹ̩.tʰɤ	даст
314	پای	足	leg	吧衣	pʰuɑ.i	pʰa.ji	пой
315	فربه	肥	fat	法兒必黑	puɑ.•.pi.xɛ	fa.əɹ.pi.xei	фарбеҳ (фарбӣ)
316	لاغر	瘦	thin; skinny	剌額兒	•.•.•	la.ɤ.əɹ	лоғар
317	فرج	陰	yin; concealed; negative; femininity (as the opposite of masculinity)	法兒知	puɑ.•.•	fa.əɹ.tʂɹ̩	фарҷ
318	ذكر	陽	yang; open; positive; masculinity (as the opposite of femininity)	則克兒	tsɛ.•.•	tsɤ.kʰɤ.əɹ	закар
319	خوی	性	character; nature	乎衣	•.i	xu.ji	хӯ(й)
320	جان	命	life	扎恩	•.ən	tʂa.ən	ҷон
321	سخن	言	word; speech	塞昏	sɛ.xun	sɤ.xwən	сухан / сахун
322	عمل	行	conduct; behavior	阿默勒	ɔ/ɑ.•.lɛ	a.mwo.lɤ	амал
323	نیت	意	aspiration	你夜忒	ni.iɛ.tʰɛ	ni.je.tʰɤ	ният

324	صورت	形	shape; body	蘇勒弒	su.lɛ.tʰɛ	su.lɤ.tʰɤ	сурат
325	جکر	肝	liver	止革兒	tʃ.kɛ.•	tʂʰ.kɤ.əɹ	чигар
326	زهره	膽	gall bladder	則黑勒	tsɛ.xɛ.lɛ	tsɤ.xei.lɤ	захра
327	استخوان	骨	bone	五思土黑洼恩	•.sɹ.tʰu.xɛ.•.ən	wu.sɹ.tʰu.xei.wa.ən	устухон
328	کوشت	肉	flesh	鍋石弒	kuɔ.•.tʰɛ	kwo.ʂɹ.tʰɤ	гӯшт
329	زنخ	頰	cheek	則納黑	tsɛ.nɑ.xɛ	tsɤ.na.xei	занах
330	گردن	項	nape of the neck	革兒丹	kɛ.•.tan	kɤ.əɹ.tan	гардан
331	شش	肺	lung	束石	•.•	ʂu.ʂɹ	шуш
332	سینه	胸	chest	洗納	si.nɑ	ɕi.na	сина
333	تارکسر	頂	top; the crown (of the head/hat)	他里起塞兒	tʰɑ.•.kʰi.sɛ.•	tʰa.li.tɕʰi.sɤ.əɹ	тораки сар
334	مشت	拳	fist	木石弒	•.•.tʰɛ	mu.ʂɹ.tʰɤ	мушт
335	کتف	肩	shoulder	克替夫	•.•.piu	kʰɤ.tʰi.fu	китф / кифт
336	پشت	背	back	僕石弒	•.•.tʰɛ	pʰu.ʂɹ.tʰɤ	пушт
337	دیده[61]	晴[61]	clear (see note 61)	底得	ti.tɛ	ti.tɤ	дида
338	معده	胃	stomach	米阿得	mi.ɔ/ɑ.tɛ	mi.a.tɤ	меъда
339	روده	腸	bowel	羅得	luɔ.tɛ	lwo.tɤ	рӯда
340	پی	筋	muscle	拍衣	pʰuɛ.i	pʰai.ji	пай
341	زلف	鬢	sidelock	則勒夫	tsɛ.lɛ.piu	tsɤ.lɤ.fu	зулф
342	ریش	鬚	beard	里石	•.•	li.ʂɹ	риш
343	انکشت	指	finger	俺故石弒	an.•.•.tʰɛ	an.ku.ʂɹ.tʰɤ	ангушт
344	بیشانی	額	forehead	撇沙你	pʰiɛ.ʂɑ.ni	pʰje.ʂa.ni	пешонӣ
345	آفت	灾	disaster	阿法弒	ɔ/ɑ.puɑ.tʰɛ	a.fa.tʰɤ	офат[62]
346	آب دیده	淚	tear	阿卜底得	ɔ/ɑ.•.ti.tɛ	a.pu.ti.tɤ	оби дида
347	خون	血	blood	乎恩	•.ən	xu.ən	хун
348	تن	身	body	貪	tʰan	tʰan	тан
349	خلقوم	咽喉	throat	虎魯故恩	xu.lu.•.ən	xu.lu.ku.ən	ҳалқум / хулқум
350	بیمار	疾病	disease; illness	別媽兒	piɛ.mɑ.•	pje.ma.əɹ	бемор

	宮室門		The "palace" section				
351	قلمه	樓	storied building	革里黙	kɛ.•.•	kɤ.li.mwo	×
352	صفه	臺	raised large building; platform	塞法	sɛ.puɑ	sɤ.fa	суффа
353	کوشک	殿	large building with a substantial base; hall	科石克	kʰuɔ.•.•	kʰɤ.ʂɹ.kʰɤ	кӯшк
354	غرفه	閣	pavilion	五兒法	•.•.puɑ	wu.əɹ.fa	ғурфа
355	انبار	倉	granary; magazine	俺巴兒	an.puɑ.•	an.pa.əɹ	анбор
356	خزینه	庫	storehouse	黑即納	xɛ.•.nɑ	xei.tɕi.na	хазина
357	مسجد	寺	temple	黙思只得	•.sɹ.•.tɛ	mwo.sɹ.tʂɹ.tɤ	масҷид
358	خانه	房	house; room	哈納	xɑ.nɑ	xa.na	хона
359	در	門	door; gate	得兒	tɛ.•	tɤ.əɹ	дар
360	دریچه	窗	window	得兒徹	tɛ.•.•	tɤ.əɹ.tʂʰɤ	дарича

61 晴 *qíng* 'clear' here is probably a misspelt 睛 *jīng* 'eye'.

62 Tajik *офат* does mean 'disaster', but seems out of place in the "body" section. This entry word may hence be a misspelt افت, which is current in modern Tajik (as well as in modern Uzbek) in the form of *афт* 'face; appearance; facial expression'. Incidentally, *афт* can be identified as synonymous with modern Tajik *рӯй* 'face' (see entry 302) in certain contexts (see Muhammadiev 1975: 165).

361	سقف	梁	roof beam	塞革夫	sɛ.kɛ.piu	sɤ.kɤ.fu	сақф
362	ستون	柱	pillar; column	速禿恩	•.tʰu.ən	su.tʰu.ən	сутун
363	پغنه	堦	step; stairs	迫額納	•.•.nɑ	pʰwo.ɤ.na	пағна
364	مناره	塔	pagoda	母納勒	mu.na.lɛ	mu.na.lɤ	минора
365	تیم	店	shop	梯尹	tʰi.•	tʰi.jin	тим[63]
366	بام	簷	eave	巴木	puɑ.•	pa.mu	бом
367	پل	橋	bridge	僕勒	•.lɛ	pʰu.lɤ	пул
368	مجور	廊	corridor; porch	米知幹兒	mi.•.•.•	mi.tʂɹ.wo.əɹ	×
369	خشت	磚	brick	黑石弍	xɛ.•.tʰɛ	xei.ʂɹ.tʰɤ	хишт / ғишт
370	سفال	瓦	tile	塞法勒	sɛ.puɑ.lɛ	sɤ.fa.lɤ	сафол
371	سرای	廳	hall	塞刺衣	sɛ.•.i	sɤ.la.ji	сарой
372	وسه	橡	rafter	我塞	ɔ.sɛ	wo.sɤ	васса
373	دارافزین	欄杆	railing	打兒阿夫即尹	tɑ.•.ɔ/ɑ.piu.•.•	ta.əɹ.a.fu.tɕi.jin	дорафзин / дарафзин
374	بارگاه	丹墀	red steps leading to a palace	巴兒噶黑	puɑ.•.•.xɛ	pa.əɹ.ka.xei	боргоҳ
375	یام خانه	館驛	posthouse inn	呀木哈納	•.•.xɑ.nɑ	ja.mu.xa.na	ёмхона

	鳥獸門		The "birds and beasts" section				
376	اژدژ	龍	dragon	阿日得兒	ɔ/ɑ.ʐ̩.tɛ.•	a.ɹɹ.tɤ.əɹ	аждар
377	پلنک	虎	tiger	迫郎克	•.lɑŋ.•	pʰwo.lɑŋ.kʰɤ	паланг
378	شیر	獅[64]	lion	賒兒	•.•	ʂɤ.əɹ	шер
379	فیل	象	elephant	非勒	•.lɛ	fei.lɤ	фил
380	اشتر	駝	camel	五束土兒	•.•.tʰu.•	wu.ʂu.tʰu.əɹ	уштур / шутур
381	اسب	馬	horse	阿思卜	ɔ/ɑ.sɹ.•	a.sɹ.pu	асп / асб
382	کاو	牛	cow	高	kɔʊ	kɑu	гов / гав
383	کوسفند	羊	sheep	果思番得	kuɔ.sɹ.puiɛn.tɛ	kwo.sɹ.fan.tɤ	гӯсфанд / гусфанд
384	مرغ	雞	fowl; chicken	木兒額	•.•.•	mu.əɹ.ɤ	мурғ
385	ماهی	魚	fish	馬希	muɑ.xi	ma.ɕi	моҳӣ
386	قاز	鵝	goose	噶子	•.tsɹ	ka.tsɹ	қоз
387	بط	鴨	duck	百弍	puɛ.tʰɛ	pai.tʰɤ	бат
388	روباه	狐	fox	魯巴黑	lu.puɑ.xɛ	lu.pa.xei	рӯбоҳ / рӯбаҳ / рубоҳ
389	خرکوش	兔	rabbit	黑兒鍋石	xɛ.•.kuɔ.•	xei.əɹ.kwo.ʂɹ	харгӯш / харгуш
390	خوک	猪	pig	乎克	•.•	xu.kʰɤ	хук
391	سک	犬	dog	塞克	sɛ.•	sɤ.kʰɤ	саг
392	فراشترک	燕	swallow	法剌石土路克	puɑ.•.•.tʰu.•.•	fa.la.ʂɹ.tʰu.lu.kʰɤ	фароштурук
393	اسور	鴈	wild goose	阿思幹兒	ɔ/ɑ.sɹ.•.•	a.sɹ.wo.əɹ	×
394	حمدونه	猴	monkey	罕都納	xan.tu.nɑ	xan.tu.na	ҳамдуна / ҳамдӯна
395	یوز	豹	leopard	迋子	•.tsɹ	ɥy.tsɹ	юз
396	جانور	鶯	oriole	扎奴幹兒	•.nu.•.•	tʂa.nu.wo.əɹ	ҷонвар / ҷонавар[65]
397	باشه	鷂	sparrow hawk	巴舍	puɑ.•	pa.ʂɤ	боша
398	مار	蛇	snake	媽兒	muɑ.•	ma.əɹ	мор
399	مشک	麝	musk deer	木石克	•.•.•	mu.ʂɹ.kʰɤ	мушк
400	کربه	猫	cat	古兒百	ku.•.puɛ	ku.əɹ.pai	гурба
401	موش	鼠	mouse; rat	母石	mu.•	mu.ʂɹ	муш

---

63 Modern Tajik *тим* 'large caravanserai; covered bazaar' seems to diverge semantically from 店 'shop'. Honda (1963: 15) identifies Timurid Persian تیم as a loanword from Chinese.

64 According to Tōdō and Kanō (2005: 1132), 獅 *shī* 'lion' is a loanword from *kodai perushago* lit. 'ancient Persian'.

65 Modern Tajik *ҷон(а)вар* 'animal'.

402	مکس	蠅	fly	黙革思	•.kɛ.sɹ	магас
403	پشه	蚊	mosquito	迫舎	•.•	pʰwo.ʂɤ	пашша
404	ژمره	蟬	cicada	日母勒	ʐ].mu.lɛ	ɻɹ.mu.lɤ	×
405	پروانه	蛾	moth	迫兒洼納	•.•.•.nɑ	pʰwo.əɹ.wa.na	парвона
406	کرم	蟲	insect; worm	乞林	•.lin	tɕʰi.lin	кирм
407	مورچه	蟻	ant	抹兒徹	muɔ.•.•	mwo.əɹ.tʂɤ	мӯрча / мурча
408	پریدن	飛	fly	迫里丹	•.•.tan	pʰwo.li.tan	паридан
409	بانک	鳴	cry	邦克	puɑŋ.•	pɑŋ.kʰɤ	бонг
410	پشم	毛	hair; feather	迫深	•.•	pʰwo.ʂən	пашм
411	بال	翅	wing	巴勒	pua.lɛ	pa.lɤ	бол
412	سنب	蹄	hoof	孫卜	sun.•	swən.pu	сунб / сум(м)
413	جنکال	爪	claw; talon	展噶勒	tʂan.•.lɛ	tʂan.ka.lɤ	чангол
414	منقار	嘴	mouth; beak	敏噶兒	min.•.•	min.ka.əɹ	минқор
415	فلوس	鱗⁶⁶	(female) unicorn	府羅思	piu.luɔ.sɹ	fu.lwo.sɹ	фулус⁶⁷
416	ظرافه	麒麟⁶⁸	kylin; unicorn	祖刺法	•.•.pua	tsu.la.fa	заррофа⁶⁹
417	سیمرغ	鳳凰	phoenix	洗木兒額	si.•.•.•	ɕi.mu.əɹ.ɤ	симургʻ
418	طرطی	鸚鵡	parrot	脱推	tʰuɔ.•	tʰwo.tʰwei	тӯтӣ
419	لحام	鴛鴦	mandarin duck	魯哈木	lu.xa.•	lu.xa.mu	×
420	طاوس	孔雀	peacock	他屋思	tʰɑ.•.sɹ	tʰa.wu.sɹ	товус
421	چغز	蝦蟇	frog; toad	徹額子	•.•.tsɹ	tʂɤ.ɤ.tsɹ	чаʻз
422	چرچری	翡翠	halcyon	赤兒赤里	•.•.•.•	tʂʰɻ.əɹ.tʂʰɻ.li	чирчирй
423	ورتیج	鶴鶉	quail	我兒梯知	ɔ.•.tʰi.•	wo.əɹ.tʰi.tʂ	вартиш
424	ملخ	蝗虫	grasshopper; locust	黙勒黒	•.lɛ.xɛ	mwo.lɤ.xei	малах

	花木門		The "flowers and trees" section			
425	درخت	樹	plant; tree	得忒黒忒	tɛ.lɛ.xɛ.tʰɛ	tɤ.lɤ.xei.tʰɤ	дарахт
426	چوب	木	tree; wood	棚卜	•.•	ʂwo.pu	чӯб
427	توت	桑	mulberry	秃忒	tʰu.tʰɛ	tʰu.tʰɤ	тут
428	بید	柳	willow	別得	piɛ.tɛ	pje.tɤ	бед
429	سرو	松	pine tree	速魯	•.lu	su.lu	сарв
430	ابخل	栢	cypress	五卜戸勒	•.•.xu.lɛ	wu.pu.xu.lɤ	×
431	کل	花	flower	故勒	•.lɛ	ku.lɤ	гул
432	علف	草	grass	阿勒夫	ɔ/ɑ.lɛ.piu	a.lɤ.fu	алаф
433	بادرنک	瓜	melon	巴得郎克	pua.tɛ.lɑŋ.•	pa.tɤ.lɑŋ.kʰɤ	бодиринг / бодранг⁷⁰
434	میوه	果	fruit	滅幹	miɛ.•	mje.wo	мева
435	مرود	梨	pear	母魯得	mu.lu.tɛ	mu.lu.tɤ	муруд / амруд
436	چیغان	棗	jujube	赤卜阿恩	•.•.ɔ/ɑ.ən	tʂʰɻ.pu.a.ən	×
437	آلو	李	plum	阿魯	ɔ/ɑ.lu	a.lu	олу
438	زردارو	杏	apricot	則兒打魯	tsɛ.•.ta.lu	tsɤ.əɹ.ta.lu	зардолу
439	شفتالو	桃	peach	舎夫他魯	•.piu.tʰa.lu	ʂɤ.fu.tʰa.lu	шафтолу
440	ابار	榴	pomegranate	阿納兒	ɔ/ɑ.nɑ.•	a.na.ɹ	анор

66 This 麟 *lín* is probably a mistakenly written 鱗 *lín* 'fish scale' (see Honda 1963: 17).

67 Modern Tajik *фулус* 'fish scales'.

68 The Japanese word 麒麟 *kirin* (*qílín* in Chinese pīnyīn) is used in reference to 'giraffe'.

69 Modern Tajik *заррофа* 'giraffe'.

70 Modern Tajik *бодиринг* means not 'melon' but 'cucumber'. Cucumber is called 黄瓜 *huángguā* lit. 'yellow melon' or 胡瓜 *húguā* lit. 'melon introduced from the northern/western ethnic groups' in modern Standard Chinese. (Note that both 黄瓜 and 胡瓜 contain the glyph 瓜 *guā* 'melon'.)

441	ختمی	葵	geraniums, holly-hocks, mallows, etc.	黑忒密	xɛ.tʰɛ.mi	xei.tʰɤ.mi	хатмӣ
442	اقخوان	桂	fragrant olive	五革黑洼恩	•.kɛ.xɛ.•.ən	wu.kɤ.xei.wa.ən	уқҳувон / ақҳавон
443	بابونج	菊	chrysanthemum	巴卜納知	puɑ.•.nɑ.•	pa.pu.na.tʂɹ	×[71]
444	غنچه	蕋	stamen	穏徹	un.•	wən.tʂʰɤ	ғунча
445	نی	竹	bamboo	奈	nai	nai	най
446	گل نیلوفر	蓮	lotus	故勒你魯法兒	•.lɛ.ni.lu.puɑ.•	ku.lɤ.ni.lu.fa.ɹ	гули нилуфар
447	بادنجان	茄	aubergine	把廷扎恩	puɑ.•.•.ən	pa.tʰjəŋ.tʂa.ən	бодинчон / бодимчон
448	پیاز	葱	onion	痞呀子	pʰi.•.tsɹ	pʰi.ja.tsɹ	пиёз
449	زنجبیل	薑	ginger	簪知必勒	•.•.pi.lɛ	tsan.tʂɹ.pi.lɤ	занчабил
450	سیر	蒜	garlic	西兒	•.•	ɕi.əɹ	сир
451	شالی	稻	rice	沙里	ʂɑ.•	ʂa.li	шолӣ
452	کندم	麥	wheat, barley, oats, etc.	敢敦	kan.tun	kan.twən	гандум
453	طراوت	新鮮	fresh	忒剌幹忒	tʰɛ.•.•.tʰɛ	tʰɤ.la.wo.tʰɤ	тароват
454	ماش	豆	bean	媽石	muɑ.•	ma.ʂɹ	мош
455	قلقاس[72]	芋	taro; potatoes	古勒嘎思	ku.lɛ.•.sɹ	ku.lɤ.ka.sɹ	×
456	کنب	麻	hemp, flax, jute, etc.	克納卜	•.nɑ.•	kʰɤ.na.pu	канаб
457	کندنا	韭	leek; chive	敢得納	kan.tɛ.nɑ	kan.tɤ.na	гандано
458	شاخ	枝	branch	沙黑	ʂɑ.xɛ	ʂa.xei	шох
459	برک	葉	leaf	百兒克	puɛ.•.•	pai.əɹ.kʰɤ	барг
460	نیش نی	笋	bamboo shoot	你石奈	ni.•.nai	ni.ʂɹ.nai	неши най
461	بیخ	根	root	別耶	piɛ.xɛ	pje.xei	бех
462	سلیجه[73]	牡丹	peony	塞里黑	sɛ.•.xɛ	sɤ.li.xei	×
463	سبست[74]	苜蓿	lucerne	洗卜細忒	si.•.•.tʰɛ	ɕi.pu.ɕi.tʰɤ	×
464	انکور	蒲萄	grape	俺姑兒	an.•.•	an.ku.əɹ	ангур
465	فخرب	浮萍	duckweed	法黑路卜	puɑ.xɛ.•.•	fa.xei.lu.pu	×
466	کلاب	薔薇	rose	古剌卜	ku.•.•	ku.la.pu	гулоб

	器用門		The "utensils" section				
467	کمان	弓	bow	克媽恩	•.muɑ.ən	kʰɤ.ma.ən	камон
468	تیر	箭	arrow	梯兒	tʰi.•	tʰi.əɹ	тир
469	خود	盔	helmet	乎得	•.tɛ	xu.tɤ	хӯд
470	جوشن	甲	armour	卓山	tʂuɔ.ʂan	tʂwo.ʂan	чавшан
471	نیزه	鎗	spear	乃則	nai.tsɛ	nai.tsɤ	найза
472	کارد	刀	knife; sword	嘎兒得	•.•.tɛ	ka.əɹ.tɤ	корд
473	رکاب	鐙	stirrup	里嘎卜	•.•.•	li.ka.pu	рикоб
474	جناق	鞊	saddleflap	主納革	tʂʉ.nɑ.kɛ	tʂu.na.kɤ	чаноғ
475	طبق	盤	plate; dish	忒百革	tʰɛ.puɛ.kɛ	tʰɤ.pai.kɤ	табақ
476	کاسه	碗	bowl; cup	嘎塞	•.sɛ	ka.sɤ	коса
477	صراحی	壺	pot; bottle	速剌黑	•.•.xɛ	su.la.xei	сурохӣ

---

71 Modern Tajik *бобуна* means 'camomile'.

72 According to Steingass (2012: 985), قلقاس means "[t]he root of a plant which is edible when cooked".

73 According to Steingass (2012: 695), سلیخة means "[a] certain perfume; benzoin or balsam of the *bān*-tree before it is prepared".

74 According to Steingass (2012: 652), سپست has "[t]refoil, clover" as one of its meanings.

478	چوکی[75]	筯	chopsticks	搊几	•.•	ʂwo.tɕi	×
479	تمغا	印	stamp; seal	貪阿	tʰan.ɔ/ɑ	tʰan.a	тамға
480	آینه	鏡	mirror	阿衣納	ɔ/ɑ.i.nɑ	a.ji.na	оина
481	غژک	琴	zither	額日克	•.ʐ̩.•	ɣ.ɻ.kʰɣ	ғижжак
482	شطرنج	棋	chess	舍忒藍知	•.tʰɛ.lan.•	ʂɣ.tʰɣ.lan.tʂ̩	шатранч / сатранч
483	یوی	香	perfume	鉢衣	•.i	pwo.ji	бӯй / бӯ / бу
484	نقشین	畫	paint; drawing	納革石尹	nɑ.kɛ.•.•	na.kɣ.ʂ̩.jin	нақшин
485	علم	旗	flag	阿藍	ɔ/ɑ.lan	a.lan	алам
486	چتر	傘	umbrella; parasol	徹忒兒	•.tʰɛ.•	tʂʰɣ.tʰɣ.əɻ	чатр
487	کوزه	瓶	bottle	科則	kʰuɔ.tsɛ	kʰɣ.tsɣ	кӯза
488	کشتی	船	boat	起石梯	kʰi.•.tʰi	tɕʰi.ʂ̩.tʰi	киштӣ
489	کردون	車	vehicle; wheeled instrument/ machine	革兒都恩	kɛ.•.tu.ən	kɣ.əɻ.tu.ən	гардун
490	بوریا	席	seat	鉢兒呀	•.•.ja	pwo.ɻɛ.ja	бӯрё / бурё
491	نای	笛	flute	納衣	nɑ.i	na.ji	най / ной
492	دهل	鼓	drum	堵戶勒	•.xu.lɛ	tu.xu.lɣ	духул / дӯл
493	قدح	鍾	bell; goblet	革得黑	kɛ.tɛ.xɛ	kɣ.tɣ.xei	қадаҳ
494	لکام	嗇	bridle	魯噶木	lu.•.•	lu.ka.mu	лигом / лагом
495	اکز	鈎	hook	阿革子	ɔ/ɑ.kɛ.tsɹ̩	a.kɣ.tsɹ̩	окач
496	پرده	簾	curtain	迫兒得	•.•.tɛ	pʰwo.əɻ.tɣ	парда
497	درفش	錐	awl	堵路夫石	•.•.piu.•	tu.lu.fu.ʂɹ̩	дарафш / дирафш
498	آسیا	磨	mill	阿洗呀	ɔ/ɑ.si.•	a.ɕi.ja	осиёб / осиё
499	کوبه	杵	pestle	科百	kʰuɔ.puɛ	kʰɣ.pai	кӯба[76]
500	جواز	臼	mortar	主洼子	tʂɥ.•.tsɹ̩	tʂu.wa.tsɹ̩	чувоз
501	چراغ	燈	lamp	赤刺額	•.•.•	tʂʰɻ̩.la.ɣ	чироғ / чароғ
502	خم	甕	urn	昏	xun	xwən	хум
503	بادبان	篷	awning; sail	巴得巴恩	puɑ.tɛ.puɑ.ən	pa.tɣ.pa.ən	бодбон
504	تازیانه	鞭	whip	他子呀納	tʰa.tsɹ̩.•.nɑ	tʰa.tsɹ̩.ja.na	тозиёна
505	کوی	毬	ball	鍋衣	kuɔ.i	kwo.ji	гӯ(й)
506	مسقار	笙	reed pipe	母洗噶兒	mu.si.•.•	mu.ɕi.ka.əɻ	мусиқор
507	شیره[77]	卓	table	史勒	ʂɹ̩.lɛ	ʂɹ̩.lɣ	×
508	سبندلی	橙	stool	散得里	san.tɛ.•	san.tɣ.li	сандалӣ
509	لطن[78]	盆	basin	勒團	lɛ.•	lɣ.tʰwan	лаган
510	بادویزن	扇	fan	巴得月簪	puɑ.tɛ.iuɛ.•	pa.tɣ.ɥe.tsan	бодбезан / бодбизан
511	دیک	鍋	pot; cauldron	迭克	•.•	tje.kʰɣ	дег
512	سفط	箱	box; chest	塞法式	sɛ.puɑ.tʰɛ	sɣ.fa.tʰɣ	сабад
513	فانوس	燈籠	lantern	法奴思	puɑ.nu.sɹ̩	fa.nu.sɹ̩	фонус
514	شمع	蠟燭	candle	舍黙額	•.•.•	ʂɣ.mwo.ɣ	шамъ
515	محفه	轎子	sedan chair	黙黑法	•.xɛ.puɑ	mwo.xei.fa	миҳаффа
516	مجمر	香爐	incense burner	米知黙兒	mi.•.•.•	mi.tʂɹ̩.mwo.əɻ	миҷмар

---

75 According to Jarring (1964: 76), *čökɛ* 'chop-sticks' exists in Turkic dialects spoken in the southern part of Xinjiang.

76 Modern Tajik *кӯба* 'mallet; hammer'.

77 The meanings of شیره listed in Steingass (2012: 774) include "a tray with a leg to stand upon".

78 لطن may be a misspelt لکن, which, according to Steingass (2012: 1128), has the meaning of "a basin, bowl".

	衣服門		The "clothing" section				
517	جامه	衣	clothes; garment	扎黙	•.•	tʂa.mwo	ҷома
518	تاج	冠	hat	他知	tʰɑ.•	tʰa.tʂɻ	тоҷ
519	موزه	靴	boot	抹則	muɔ.tsɛ	mwo.tsɤ	мӯза
520	کمر	帶	belt	克黙兒	•.•.•	kʰɤ.mwo.əɻ	камар
521	کتی	錦	brocade	克梯	•.tʰi	kʰɤ.tʰi	×
522	توار⁷⁹	叚⁸⁰	satin⁸¹	忒洼兒	tʰɛ.•.•	tʰɤ.wa.əɻ	×
523	حریر	綾	thin silk	黒里兒	xɛ.•.•	xei.li.əɻ	ҳарир
524	لای	羅	net; silk gauze	剌衣	•.i	la.ji	лой
525	کرباس	布	cloth	克兒巴思	•.•.puɑ.sɿ	kʰɤ.əɻ.pa.sɿ	карбос
526	تورقو⁸²	絹	plain silk	土兒孤	tʰu.•.ku	tʰu.əɻ.ku	×
527	ابریشیم	絲	raw silk; silk thread	阿卜列石尹	ɔ/ɑ.•.liɛ.•.•	a.pu.lje.ʂɻ.jin	абрешим
528	ریشته	線	thread; string; line	里石忒	•.•.tʰɛ	li.ʂɻ.tʰɤ	ришта
529	ابره	表	surface	阿卜勒	ɔ/ɑ.•.lɛ	a.pu.lɤ	абра
530	استر	裏	lining; inside	阿思忒兒	ɔ/ɑ.sɿ.tʰɛ.•	a.sɿ.tʰɤ.əɻ	астар
531	⁸³	襟	front of a garment	打鸞	tɑ.muan	ta.man	доман
532	⁸⁴	袖	sleeve	阿思梯尹	ɔ/ɑ.sɿ.tʰi.•	a.sɿ.tʰi.jin	остин
533	⁸⁵	綿	cotton	敏搭禿	min.tɑ.tʰu	min.ta.tʰu	×
534	⁸⁶	帽	hat; cap	苦剌黑	kʰu.•.xɛ	kʰu.la.xei	кулоҳ
535	کریبان	領	collar; neck	己里巴思	•.•.puɑ.ən	tɕi.li.pa.ən	гиребон
536	نمد	氊	felt	納黙得	nɑ.•.tɛ	na.mwo.tɤ	намад
537	قضاغند	被	quilt; blanket	革咱安得	kɛ.•.an.tɛ	kɤ.tsa.an.tɤ	қазоган(д) / қазоған(д)

---

79 According to Steingass (2012: 332), توار means "[a] rope for tying on a load".

80 Honda (1963: 20) identifies the glyph 叚 *jiǎ* 'false; borrow', which has 叚 as one of its alternative forms, as 段 *duàn* 'step'.

81 This is not the meaning of 叚 *jiǎ* 'false; borrow', but that of 緞 *duàn* 'satin'. I tentatively assume that 叚 here is misspelt for, or is meant to represent, 緞, because this particular entry is in the 'clothing' section, and also because 緞 appears in the place of 叚 in another copy of *huihuiguan zazi* (Beijing tushuguan guji chuban bianji zu 1987–[1994]: 553).

82 Clauson's dictionary of pre-thirteenth-century Turkic (1972: 539) has *torku:* 'silk fabric' as one of its entries, but Clauson suspects that it may be a loanword.

83 The entry numbered 531 in Honda (1963: 20) is absent in the Berlin Manuscript, hence the blank. The entry is دامن in Beijing tushuguan guji chuban bianji zu (1987–[1994]: 499, 554), from which the Chinese translation 襟 and Chinese-script transcription 打鸞 in this row are retrieved.

84 The entry numbered 532 in Honda (1963: 20) is absent in the Berlin Manuscript, hence the blank. The entry appears as آستیی and آستن in different pages of Beijing tushuguan guji chuban bianji zu (1987–[1994]: 499, 554), from which the Chinese translation 袖 and Chinese-script transcription 阿思梯尹 in this row are retrieved.

85 The entry numbered 533 in Honda (1963: 20) is absent in the Berlin Manuscript, hence the blank. The entry is منداتو in Beijing tushuguan guji chuban bianji zu (1987–[1994]: 500, 554), from which the Chinese translation 綿 and Chinese-script transcription 敏搭禿 in this row are retrieved.

86 The entry numbered 534 in Honda (1963: 20) is absent in the Berlin Manuscript, hence the blank. The entry appears as کداه and کلاه in different pages of Beijing tushuguan guji chuban bianji zu (1987–[1994]: 500, 554), from which the Chinese translation 帽 and Chinese-script transcription 苦剌黑 in this row are retrieved.

538	پستر	褥	mattress	痞思忒兒	pʰi.sɻ.tʰɛ.•	pʰi.sɻ.tʰɤ.əɹ	бистар
539	بالش	枕	pillow	把力石	puɑ.•.•	pa.li.ʂɻ	болиш / болишт
540	جوال	袋	bag; pouch; sack	主洼勒	tʂʮ.•.lɛ	tʂu.wa.lɤ	чувол
541	فوطه	手巾	towel	府忒	piu.tʰɛ	fu.tʰɤ	фута / фӯта
542	نيمتنه	短衫	short upper garment	恁忒納	nin.tʰɛ.nɑ	nin.tʰɤ.na	нимтана

| | 飲食門 | | The "eating and drinking" section | | | | |

543	توگی[87]	米	husked rice	土几	tʰu.•	tʰu.tɕi	×
544	كوشت	肉	meat; flesh	鍋石忒	kuɔ.•.tʰɛ	kwo.ʂɻ.tʰɤ	гӯшт
545	آش	飯	cooked rice; meal	阿石	ɔ/ɑ.•	a.ʂɻ	ош
546	آرد	麪	flour; noodle	阿兒得	ɔ/ɑ.•.tɛ	a.əɹ.tɤ	орд
547	روغن	油	oil	羅安	luɔ.an	lwo.an	равған
548	نمك	鹽	salt	納黙克	nɑ.•.•	na.mwo.kʰɤ	намак
549	شراب	酒	alcoholic drink	捨剌卜	ʂɛ.•.•	ʂɤ.la.pu	шароб
550	سركه	醋	vinegar	洗兒克	si.•.•	ɕi.əɹ.kʰɤ	сирка / сирко
551	شكر	糖	sugar	捨克兒	ʂɛ.•.•	ʂɤ.kʰɤ.əɹ	шакар
552	عسل	蜜	honey	阿塞勒	ɔ/ɑ.sɛ.lɛ	a.sɤ.lɤ	асал
553	دارو	藥	medicine	打魯	tɑ.lu	ta.lu	дору
554	چا	茶	tea	本音[88]	pun.•	pən.jin	чой
555	گرسنه	饑	hungry	古兒思納	ku.•.sɻ.nɑ	ku.əɹ.sɻ.na	гурусна
556	سير	飽	full	㘝兒	siɛ.•	ɕje.əɹ	сер
557	تشنه	渇	thirsty	忒石納	tʰɛ.•.nɑ	tʰɤ.ʂɻ.na	ташна
558	آشاميدن	飲	drink	阿沙米丹	ɔ/ɑ.ʂɑ.mi.tan	a.ʂa.mi.tan	ошомидан
559	خوردن	喫	eat; consume	火兒丹	xuɔ.•.tan	xwo.əɹ.tan	хӯрдан
560	مزه	味	taste	黙則	•.tsɛ	mwo.tsɤ	маз(з)а
561	شيرين	甜	sweet	史里尹	ʂɻ.•.•	ʂɻ.li.jin	ширин
562	تلخ	苦	bitter	忒勒黒	tʰɛ.lɛ.xɛ	tʰɤ.lɤ.xei	талх
563	ترش	酸	sour	土路石	tʰu.•.•	tʰu.lu.ʂɻ	турш
564	تره	辣	peppery; pungent	忒兒勒	tʰɛ.•.lɛ	tʰɤ.əɹ.lɤ	тарра
565	شور	鹹	salty	朔兒	ʂuɔ.•	ʂwo.əɹ	шӯр
566	سوختن	燒	bake	鎖黒貪	suɔ.xɛ.tʰan	swo.xei.tʰan	сӯхтан
567	جوشيدن	煮	boil	卓石丹	tʂuɔ.•.tan	tʂwo.ʂɻ.tan	чӯшидан
568	خام	生	raw	哈恩	xɑ.ən	xa.ən	хом
569	پخته	熟	ripe	僕黒忒	•.xɛ.tʰɛ	pʰu.xei.tʰɤ	пухта
570	شوربا	湯	soup	朔兒巴	ʂuɔ.•.puɑ	ʂwo.əɹ.pa	шӯрбо
571	نان	餅	round flat cake	納恩	nɑ.ən	na.ən	нон
572	روغن كاو	酥	butter	羅安高	luɔ.an.kɒʊ	lwo.an.kɑu	равғани гов
573	جغرات	酪	junket; curd	主額剌忒	tʂʮ.•.•.tʰɛ	tʂu.ɤ.la.tʰɤ	чурғот / чуғрот
574	كومه	醬	(bean) sauce made by fermenting	科黙	kʰuɔ.•	kʰɤ.mwo	×
575	آميختن	調和	mix; blend	阿滅黒貪	ɔ/ɑ.miɛ.xɛ.tʰan	a.mje.xei.tʰan	омехтан

---

87 Clauson's (1972: 478) pre-thirteenth-century Turkic dictionary has *tögi:* 'crushed or cleaned cereal' as one of its entries, while Doerfer (1965: 629–630) lists توگی 'millet' as a loanword in New Persian.

88 This is arguably not a transcription but a note to the reader as 本音 *běn yīn* means 'this sound' in Chinese. This is to say that the Timurid Persian word for "tea" had a similar pronunciation to that of 茶 *chá* 'tea' in Ming-period Beijing Chinese. 本音 also appears in the appendix section of the Berlin Manuscript (entry number 806 in Honda 1963: 29).

		珍寶門		The "treasure" section			
576	زر	金	gold	則兒	tsɛ.•	tsɤ.əɹ	зар
577	نقره	銀	silver	奴革勒	nu.kɛ.lɛ	nu.kɤ.lɤ	нукра
578	مرواريد	珠	pearl	黙兒洼里得	•.•.•.•.tɛ	mwo.əɹ.wa.li.tɤ	марворид
579	يشم	玉	jade	夜深	iɛ.•	je.ʂən	яшм
580	مس	銅	copper	密思	mi.sɿ	mi.sɿ	мис
581	آهن	鐵	iron	阿罕	ɔ/ɑ.xan	a.xan	оҳан
582	قاش[89]	錢	coin	噶石	•.•	ka.ʂɿ	×
583	جيزى	物	thing	赤則	•.tsɛ	tʂʰɿ.tsɤ	чизе
584	سرب	鉛	lead	速兒卜	•.•.•	su.əɹ.pu	сурб
585	ارزيز	錫	tin	阿兒即子	ɔ/ɑ.•.•.tsɿ	a.əɹ.tɕi.tsɿ	арзиз
586	سبيده	粉	powder	洗撒得	si.pʰiɛ.tɛ	ɕi.pʰje.tɤ	сапеда
587	رخت	貨	goods; property	勒黒忒	lɛ.xɛ.tʰɛ	lɤ.xei.tʰɤ	рахт
588	جزع	瑪瑙	agate	止則額	tʂʅ.tsɛ.•	tʂʰɿ.tsɤ.ɤ	чазъ
589	مرجان	珊瑚	coral	黙兒扎恩	•.•.•.ən	mwo.əɹ.tʂa.ən	марчон
590	بلور	水晶	crystal	卜魯兒	•.lu.•	pu.lu.əɹ	булӯр
591	كهربا	琥珀	amber	克黒兒巴衣	•.xɛ.•.puɑ.i	kʰɤ.xei.əɹ.pa.ji	каҳрабо / каҳрабой
592	كوهر	寶貝	treasure	稿黒兒	kʊ.xɛ.•	kɑu.xei.əɹ	гавҳар
593	حلبى[90]	玻璨	glass	黒勒必	xɛ.lɛ.pi	xei.lɤ.pi	×

		聲色門		The "voice and countenance" section			
594	كبود	青	blue-green	克卜得	•.•.tɛ	kʰɤ.pu.tɤ	кабуд
595	سرخ	紅	red	速兒黒	•.•.xɛ	su.əɹ.xei	сурх
596	زرد	黄	yellow	則兒得	tsɛ.•.tɛ	tsɤ.əɹ.tɤ	зард
597	سفيد	白	white	洗法得	si.puɑ.tɛ	ɕi.fa.tɤ	сафед
598	سياه	黑	black	洗呀黒	si.•.xɛ	ɕi.ja.xei	сиёҳ
599	نوک	紫	purple	那克	•.•	na.kʰɤ	×[91]
600	رنكارى	藍	blue	昝噶力	tsan.•.•	tsan.ka.li	зангорӣ
601	سبز	綠	green	塞卜子	sɛ.•.tsɿ	sɤ.pu.tsɿ	сабз
602	رنكين	濃	deep; thick	郎几尹	laŋ.•.•	laŋ.tɕi.jin	рангин
603	بى رنك	淡	pale; light; thin	別郎克	pie.laŋ.•	pje.laŋ.kʰɤ	беранг
604	رنك كردن	染	dye	郎克克兒丹	laŋ.•.•.•.tan	laŋ.kʰɤ.kʰɤ.əɹ.tan	ранг кардан
605	رنك	色	color	郎克	laŋ.•	laŋ.kʰɤ	ранг
606	جوزى	茶褐	dark brown	爪即	•.•	tʂɑu.tɕi	чавзӣ
607	آل	大紅	bright red	阿勒	ɔ/ɑ.lɛ	a.lɤ	ол
608	سبز روشن	明綠	bright green	塞卜子羅山	sɛ.•.tsɿ.luɔ.ʂan	sɤ.pu.tsɿ.lwo.ʂan	сабзи равшан
609	سبز تلخ	黑綠	dark green	塞卜子忒勒黒	sɛ.•.tsɿ.tʰɛ.lɛ.xɛ	sɤ.pu.tsɿ.tʰɤ.lɤ.xei	сабзи талх
610	فستقى	柳青	yellowish green	非思忒革	•.sɿ.tʰɛ.kɛ	fei.sɿ.tʰɤ.kɤ	пистақ(ҟ)й / пистоқӣ / пистагӣ

89 قاش may be related to Turkic *ka:ş* 'jade' (Clauson 1972: 669).

90 According to Steingass (2012: 428), حلبى means "[b]elonging to a milch cow; native of Aleppo; white iron, tin-plate (modern colloquialism)".

91 Modern Tajik *нок* 'unclean; impure (musk, ambergris, etc.)' may be related to this entry, though *нок* would be spelt ناک in Arabic script (Šukurov et al. 1969a: 865).

	文史門		The "literature and history" section				
611	شعر	詩	poetry; poem	舍額兒	•.•.•	ʂɤ.ɣ.əɹ	шеър
612	دفتر	書	script; writing	得夫忒兒	tɛ.piu.tʰɛ.•	tɤ.fu.tʰɤ.əɹ	дафтар
613	عبارت	文	literary composition; literary language	額巴勒忒	•.puɑ.lɛ.tʰɛ	ɣ.pa.lɤ.tʰɤ	иборат / ибора⁹²
614	خط	字	letter; character	黑忒	xɛ.tʰɛ	xei.tʰɤ	хатт
615	كاغذ	紙	paper	噶額子	•.•.tsɿ	ka.ɣ.tsɿ	коғаз
616	بكه⁹³	墨	ink	百克	puɛ.•	pai.kʰɤ	×
617	قلم	筆	writing brush	革藍	kɛ.lan	kɤ.lan	қалам
618	دوات	硯	inkstone	得洼忒	tɛ.•.tʰɛ	tɤ.wa.tʰɤ	давот
619	قرآن	經	scripture	古剌恩	ku.•.ən	ku.la.ən	Қуръон
620	تواریج	史	history	土洼列黑	tʰu.•.liɛ.xɛ	tʰu.wa.lje.xei	таърих
621	سوره	篇	chapter; passus	蘇勒	su.lɛ	su.lɤ	сура
622	دیباچه	序	preface	底巴徹	ti.puɑ.•	ti.pa.tʂʰɤ	дебоча
623	خط تحقیق	真字	regular script	黑忒忒黑革革	xɛ.tʰɛ.tʰɛ.xɛ.kɛ.kɛ	xei.tʰɤ.tʰɤ.xei.kɤ.kɤ	хатти тахқиқ
624	خط مسوده	草字	cursive script	黑忒母嫂幹得	xɛ.tʰɛ.mu.•.•.tɛ	xei.tʰɤ.mu.sɑu.wo.tɤ	хатти мусаввада
625	مخلوج	行書	semicursive script	默黑魯知	•.xɛ.lu.•	mwo.xei.lu.tʂɿ	×⁹⁴
626	خط کوفی	篆字	seal script	黑忒科法	xɛ.tʰɛ.kʰuɔ.puɑ	xei.tʰɤ.kʰɤ.fa	хатти кӯфӣ
627	بیت	詞曲	*ci* and *qu* forms of poetry	擺忒	puai.tʰɛ	pai.tʰɤ	байт

	方隅門		The "four quarters" section				
628	مشرق	東	east	母石力革	mu.•.•.kɛ	mu.ʂɹ.li.kɤ	машрик
629	مغرب	西	west	默額力卜	•.•.•.•	mwo.ɣ.li.pu	мағриб
630	جنوب	南	south	者奴卜	tʂɛ.nu.•	tʂɤ.nu.pu	ҷануб
631	شمال	北	north	石媽勒	•.muɑ.lɛ	ʂɹ.ma.lɤ	шимол
632	زبر	上	up; above	則百兒	tsɛ.puɛ.•	tsɤ.pai.əɹ	забар
633	زیر	下	below; down	節兒	•.•	tɕje.əɹ	зер
634	چپ	左	left	徹卜	•.•	tʂʰɤ.pu	чап
635	راست	右	right	剌思忒	•.sɿ.tʰɛ	la.sɿ.tʰɤ	рост
636	پیش	前	front	撇石	pʰiɛ.•	pʰje.ʂɹ	пеш
637	پس	後	back	迫思	•.sɿ	pʰwo.sɹ	пас
638	اندرون	內	inside	俺得魯恩	an.tɛ.lu.ən	an.tɤ.lu.ən	андарун
639	بیررون	外	outside	別魯恩	piɛ.lu.ən	pje.lu.ən	берун / бурун
640	میان	中	middle; centre	米呀恩	mi.•.ən	mi.ja.ən	миён
641	کرانه	邊	edge; margin	克剌納	•.•.nɑ	kʰɤ.la.na	карона
642	مربغ	方	square	母藍百額	mu.lan.puɛ.•	mu.lan.pai.ɣ	мураббаъ
643	مدور	圓	circle; round	母倒幹兒	mu.•.•.•	mu.tɑu.wo.əɹ	мудаввар
644	فراخ	寬	wide; tolerant	法剌黑	puɑ.•.xɛ	fa.la.xei	фарох
645	تنک	窄	narrow	湯克	tʰɑŋ.•	tʰɑŋ.kʰɤ	танг

92 Modern Tajik *ибора(m)* 'expression; statement'.

93 Honda (1963: 23) identifies this entry (بكه) as a loanword from Chinese. 墨 was /muɛ/ in the Ming-period Beijing Chinese according to Lu (1998: 69), and was, according to Tōdō and Kanō (2005: 381), /muək/ in Zhou, Qin, and Han, /m(b)uak/ in Sui and Tang, and /mo/ in Song, Yuan, and Ming. Pullyblank (1991: 218) reconstructs the pronunciation of the same glyph in Late Middle Chinese and Early Middle Chinese as /muɜk/ and /mək/, respectively.

94 The Arabic loanword *махлӯъ* 'deposed; cashiered' (مخلوع in Arabic script) exists in Tajik.

646	کوشه	角	corner; horn	科舍	kʰuɔ.•	kʰɤ.ʂɤ	гӯша
647	تک	底	bottom	忒克	tʰɛ.•	tʰɤ.kʰɤ	таг
648	حرم	家	home	黑藍	xɛ.lan	xei.lan	ҳарам[95]
649	یک در	間	space between; room	夜克得兒	iɛ.•.tɛ.•	je.kʰɤ.tɤ.əɹ	якдар
650	اینجا	這里	here	因扎	in.•	jin.tʂa	инчо
651	انجا	那里	there	昂扎	ɑŋ.•	ɑŋ.tʂa	ончо / унчо

	數目門		The "amount/number" section				
652	یک	一	one	夜克	iɛ.•	je.kʰɤ	як
653	دو	二	two	都	tu	tu	ду
654	سه	三	three	岁	siɛ	ɕje	се
655	جهار	四	four	又哈兒	tʂʰɑ.xɑ.•	tʂʰa.xa.əɹ	чаҳор / чор
656	بنج	五	five	潘知	pʰuan.•	pʰan.tʂɹ	панҷ
657	شش	六	six	舍石	•.•	ʂɤ.ʂɹ	шаш / шиш
658	هفت	七	seven	哈夫忒	xɑ.piu.tʰɛ	xa.fu.tʰɤ	ҳафт
659	هشت	八	eight	哈石忒	xɑ.•.tʰɛ	xa.ʂɹ.tʰɤ	ҳашт
660	نه	九	nine	奴黑	nu.xɛ	nu.xei	нӯх / нух
661	ده	十	ten	得黑	tɛ.xɛ	tɤ.xei	даҳ
662	صد	百	hundred	塞得	sɛ.tɛ	sɤ.tɤ	сад
663	هزار	千	thousand	哈咱兒	xɑ.•.•	xa.tsa.əɹ	ҳазор
664	طاق	單	single	他革	tʰɑ.kɛ	tʰa.kɤ	тоқ
665	جفت	雙	two; double	住夫忒	tʂɥ.piu.tʰɛ	tʂu.fu.tʰɤ	чуфт
666	تمن	萬	ten thousand	土蠻	tʰu.muan	tʰu.man	тӯмон / тумон
667	سیر	兩	two; both	岁兒	siɛ.•	ɕje.əɹ	сер[96]
668	عدد	數	number	阿得得	ɔ/ɑ.tɛ.tɛ	a.tɤ.tɤ	адад
669	ذره	毫釐	a minute amount; the least bit	則兒勒	tsɛ.•.lɛ	tsɤ.əɹ.lɤ	зарра

	通用門		The "currency" section				
670	هست	有	exist	哈思忒	xɑ.sɹ.tʰɛ	xa.sɹ.tʰɤ	ҳаст
671	نیست	無	nothing; nonexistent	乜思忒	miɛ.sɹ.tʰɛ	mje.sɹ.tʰɤ	нест
672	برابر	同	same	百剌百兒	puɛ.•.puɛ.•	pai.la.pai.əɹ	баробар
673	تفاوت	異	different	忒法兀忒	tʰɛ.puɑ.•.tʰɛ	tʰɤ.fa.wu.tʰɤ	тафовут
674	بلی	是	be; right; yes	百列	puɛ.liɛ	pai.lje	бале
675	فتنه	非	be not; wrong; non-	非忒納	•.tʰɛ.nɑ	fei.tʰɤ.na	фитна[97]
676	کاواک	虛	false; empty	噶洼克	•.•.•	ka.wa.kʰɤ	ковок
677	خقیقت	實	fact; full	黑革革忒	xɛ.kɛ.kɛ.tʰɛ	xei.kɤ.kɤ.tʰɤ	ҳақиқат
678	آهسته	緩	slow	阿思思忒[98]	ɔ/ɑ.sɹ.sɹ.tʰɛ	a.sɹ.sɹ.tʰɤ	оҳиста
679	تیز	急	rapid	貼子	tʰiɛ.tsɹ	tʰje.tsɹ	тез
680	دشوار	難	difficult	堵石洼兒	•.•.•.•	tu.ʂɹ.wa.əɹ	душвор

95 Modern Tajik *ҳарам* 'forbidden (place); harem'.

96 Modern Tajik *сер* 'full'. See entry number 556.

97 Modern Tajik *фитна* 'discord'.

98 This is probably a misspelt 阿黑思忒 (see Honda 1963: 25; Beijing tushuguan guji chuban bianji zu 1987–[1994]: 510, 566), whose reconstructed Ming-period Beijing Chinese pronunciation and Standard Chinese pronunciation are /ɔ.xɛ.sɹ.tʰɛ/-/ɑ.xɛ.sɹ.tʰɛ/ and /a.xei.sɹ.tʰɤ/, respectively.

681	آسان	易	easy	阿撒恩	ɔ/ɑ.sɑ.ən	a.sa.ən	осон
682	دور	遠	far	都兒	tu.•	tu.əɹ	дур
683	نزدیک	近	near	納子底克	nɑ.tsɹ.ti.•	na.tsɹ.ti.kʰɤ	наздик
684	کشادن	開	open	苦沙丹	kʰu.ʂɑ.tan	kʰu.ʂa.tan	кушодан
685	بستن	閉	close	百思貪	puɛ.sɹ.tʰan	pai.sɹ.tʰan	бастан
686	درست	精	refined; energy; essence	堵路思忒	•.•.sɹ.tʰɛ	tu.lu.sɹ.tʰɤ	дуруст
687	درشت	粗	coarse	堵路石忒	•.•.•.tʰɛ	tu.lu.ʂɹ.tʰɤ	дурушт
688	دراز	長	long	得剌子	tɛ.•.tsɹ	tɤ.la.tsɹ	дароз
689	کوتاه	短	short	科他黑	kʰuɔ.tʰɑ.xɛ	kʰɤ.tʰa.xei	кӯтоҳ
690	کلان	大	big	克剌恩	•.•.ən	kʰɤ.la.ən	калон
691	خرد	小	small	戶兒得	xu.•.tɛ	xu.əɹ.tɤ	хурд
692	بسیار	多	much; many	比洗呀恩	pi.si.•.•	pi.ɕi.ja.əɹ	бисёр
693	کمتر	寡	few	堪忒兒	kʰan.tʰɛ.•	kʰan.tʰɤ.ɹɛ	камтар
694	چون	如	be like; as	初恩	tʂʰu.ən	tʂʰu.ən	чун
695	اکنون	今	now	阿克奴恩	ɔ/ɑ.•.nu.ən	a.kʰɤ.nu.ən	акнун
696	اکر	若	if; be like	阿革兒	ɔ/ɑ.kɛ.•	a.kɤ.ɹ	агар
697	کهنه	舊	former; old	科黑納	kʰuɔ.xɛ.nɑ	kʰɤ.xei.na	кӯҳна / куҳан
698	تمامت	完	complete	忒媽默忒	tʰɛ.muɑ.•.tʰɛ	tʰɤ.ma.mwo.tʰɤ	тамомат
699	کمال	全	entire	克媽勒	•.muɑ.lɛ	kʰɤ.ma.lɤ	камол
700	اقبال	興	interest	以革跛勒	•.kɛ.puɑ.lɛ	ji.kɤ.pa.lɤ	иқбол
701	ادبار	敗	defeat; fail	以得跛兒	•.tɛ.puɑ.•	ji.tɤ.pa.əɹ	идбор
702	خوش	好	good	或石	xuɛ.•	xwo.ʂɹ	хуш
703	بد	歹	bad	百得	puɛ.tɛ	pai.tɤ	бад
704	یافتن	得	get	呀夫貪	•.piu.tʰan	ja.fu.tʰan	ёфтан
705	نایافتن	失	lose	納呀夫貪	nɑ.•.piu.tʰan	na.ja.fu.tʰan	наёфтан
706	کران	重	heavy	革剌恩	kɛ.•.ən	kɤ.la.ən	гарон / гирон
707	سبک	輕	light	塞卜克	sɛ.•.•	sɤ.pu.kʰɤ	сабук
708	فراغت	閑	idle; vacant	法剌額忒	puɑ.•.•.tʰɛ	fa.la.ɤ.tʰɤ	фароғат
709	شتاب	忙	busy	石他卜	•.tʰɑ.•	ʂɹ.tʰa.pu	шитоб
710	جیر	善	virtuous	灾兒	xai.•	xai.əɹ	хайр
711	شر	惡	evil	舍兒	•.•	ʂɤ.əɹ	шар / шарр
712	رونق	榮	glory	勞納革	loʊ.nɑ.kɛ	lau.na.kɤ	равнақ
713	شرمنده	辱	disgrace	捨兒滿得	ʂɛ.•.muan.tɛ	ʂɤ.əɹ.man.tɤ	шарманда
714	بالا رفتن	升	ascend; go up	把剌勒夫貪	puɑ.•.lɛ.piu.tʰan	pa.la.lɤ.fu.tʰan	боло рафтан
715	فرو رفتن	沉	sink; fall	府羅勒夫貪	piu.luɔ.lɛ.piu.tʰan	fu.lwo.lɤ.fu.tʰan	фурӯ рафтан
716	زشت	醜	ugly	即石忒	•.•.tʰɛ	tɕi.ʂɹ.tʰɤ	зишт
717	خوب	俊	talented; handsome	乎卜	•.•	xu.pu	хуб
718	قوی	強	strong	改迁	•.•	kai.ɥy	қавӣ
719	ضعیف	弱	weak	則額夫	tsɛ.•.piu	tsɤ.ɤ.fu	заиф
720	بها	價	price; value	百哈	puɛ.xɑ	pai.xa	баҳо
721	پر	滿	full	僕兒	•.•	pʰu.əɹ	пур
722	بایستن	用	use	把衣思貪	puɑ.i.sɹ.tʰan	pa.ji.sɹ.tʰan	боистан
723	کفایت	成	accomplish; become; achievement	起法夜忒	kʰi.puɑ.iɛ.tʰɛ	tɕʰi.fa.je.tʰɤ	кифоят[99]
724	بلند	高	high	百藍得	puɛ.lan.tɛ	pai.lan.tɤ	баланд

---

99 Modern Tajik *кифоят* 'sufficiency; contentment'.

725	پستی	低	low	迫思梯	•.sɹ.tʰi	pʰwo.sɹ.tʰi	пастӣ
726	راستی	真	true; genuine	刺思梯	•.sɹ.tʰi	la.sɹ.tʰi	ростӣ
727	دورغ	假	false; fake	朶羅額	tuɔ.luɔ.•	two.lwo.ɤ	дурӯғ
728	فریق	分	divide; division	法里革	puɑ.li.kɛ	fa.li.kɤ	фариқ
729	جماعت	聚	assemble	者媽額弍	tʂɛ.muɑ.•.tʰɛ	tʂɤ.ma.ɤ.tʰɤ	ҷамоат
730	میل	私	private	買勒	muai.lɛ	mai.lɤ	майл[100]
731	عدل	公	public; fair	阿得勒	ɔ/ɑ.tɛ.lɛ	a.tɤ.lɤ	адл[101]
732	کدشتن	過	pass	古得石貪	ku.tɛ.•.tʰan	ku.tɤ.ʂɹ.tʰan	гузаштан
733	رنج	傷	wound; injury	藍知	lan.•	lan.tʂɹ	ранҷ
734	ورزیدن	積	accumulate	我兒即丹	ɔ.•.•.tan	wo.əɹ.tɕi.tan	варзидан
735	مشفق	孝	filial (piety); mourning	母石費革	mu.•.•.kɛ	mu.ʂɹ.fei.kɤ	мушфиқ
736	سطبر	厚	thick	洗弍卜兒	si.tʰɛ.•.•	ɕi.tʰɤ.pu.əɹ	ситабр
737	تنک	薄	thin	土奴克	tʰu.nu.•	tʰu.nu.kʰɤ	тунук
738	انتظار	等	wait	尹體咱兒	•.tʰi.•.•	jin.tʰi.tsa.əɹ	интизор
739	سبب	因	cause	塞百卜	sɛ.puɛ.•	sɤ.pai.pu	сабаб
740	کدام	誰	who	革搭恩	kɛ.tɑ.ən	kɤ.ta.ən	кадом
741	صدق[102]	施	bestow; grant; carry out	塞得革	sɛ.tɛ.kɛ	sɤ.tɤ.kɤ	садақа
742	منع	阻	obstruct	黙納額	•.nɑ.•	mwo.na.ɤ	манъ
743	رها	放	let go; discharge	勒哈	lɛ.xɑ	lɤ.xa	рахо
744	رسیدن	至	reach; arrive	勒洗丹	lɛ.si.tan	lɤ.ɕi.tan	расидан
745	بدین	此	this	百底尹	puɛ.ti.•	pai.ti.jin	бадин[103]
746	فهم	省	conscious; examine oneself	法罕	puɑ.xan	fa.xan	фахм
747	نو	新	new	腦	•	nɑu	нав
748	مجموع	總	put together; total	黙知母額	•.•.mu.•	mwo.tʂɹ.mu.ɤ	маҷмӯъ
749	ولیکن	然	but; so	我列欽	ɔ.liɛ.kʰin	wo.lje.tɕʰin	валекин
750	سلامت	安	safe; calm	塞刺黙弍	sɛ.•.•.tʰɛ	sɤ.la.mwo.tʰɤ	саломат
751	کردانیدن	轉	turn	革兒打你丹	kɛ.•.tɑ.ni.tan	kɤ.əɹ.ta.ni.tan	гардонидан
752	مقصود	緣故	reason	黙革蘇得	•.kɛ.su.tɛ	mwo.kɤ.su.tɤ	мақсуд
753	جکونه	怎生	how	初科納	tʂʰu.kʰuɔ.nɑ	tʂʰu.kʰɤ.na	чӣ гуна
754	غربده	喧嘩	uproar; hubbub	額兒百得	•.•.puɛ.tɛ	ɤ.əɹ.pai.tɤ	арбада
755	سیاخت	遊翫	play; stroll	洗呀塞弍[104]	si.•.sɛ.tʰɛ	ɕi.ja.sɤ.tʰɤ	сиёсат / саёхат
756	آراستن	齊整	neat; in good order	阿刺思貪	ɔ/ɑ.•.sɹ.tʰan	a.la.sɹ.tʰan	оростан
757	پاک	乾浄	clean	叭克	pʰuɑ.•	pʰa.kʰɤ	пок

---

100 Modern Tajik *майл* 'inclination; liking'.

101 Modern Tajik *адл* 'justice; truthfulness'.

102 This might be a misspelt صدقات whose meaning is "alms" according to Steingass (2012: 784).

103 Modern Tajik *бадин* 'to, in, or with this'.

104 Judging from Beijing tushuguan guji chuban bianji zu (1987–[1994]: 514, 571), this 洗呀塞弍 may be a misspelt 洗呀黒弍, whose reconstructed Ming-period Beijing Chinese pronunciation and Standard Chinese pronunciation are /si.•.xɛ.tʰɛ/ and /ɕi.ja.xei.tʰɤ/, respectively. Both *сиёсат* 'politics' and *саёхат* 'travel' are presented in the far-right cell of this row because the pronunciation of the former resembles the Chinese pronunciation of 洗呀塞弍 while the latter appears to be congruent with the reading of 洗呀黒弍 as well as with the spelling of the entry word.

15

758	پدرفتار	舉保	?[105]	迫得勒夫他兒	•.tɛ.lɛ.piu.tʰɑ.•	pʰwo.tɤ.lɤ.fu.tʰa.əɹ	пазируфтор[106]
759	بيكار	無用	useless	別噶兒	piɛ.•.•	pje.ka.əɹ	бекор
760	خودمراد	自由	free	或得母剌得	xuɛ.tɛ.mu.•.tɛ	xwo.tɤ.mu.la.tɤ	×
761	كوچ	遷更	?[107]	科知	kʰʊɔ.•	kʰɤ.tʂɻ	кӯч
762	باك نيست	無妨	would do no harm	巴克乜思忒	puɑ.•.miɛ.sɹ.tʰɛ	pa.kʰɤ.mje.sɹ.tʰɤ	бок нест
763	پاسكونه	顛倒	turn upside down	吧思科納	pʰuɑ.sɹ.kʰʊɔ.nɑ	pʰa.sɹ.kʰɤ.na	божгуна / бозгуна / бошгуна[108]
764	ترتيب	次序	order; sequence	忒兒梯卜	tʰɛ.•.tʰi.•	tʰɤ.əɹ.tʰi.pu	тартиб
765	عادت	慣曾	custom	阿得忒	ɔ/ɑ.tɛ.tʰɛ	a.tɤ.tʰɤ	одат
766	بنهان	遮蔵	hide	潘哈恩	pʰuan.xɑ.ən	pʰan.xa.ən	пинхон
767	بهانه	推辭	decline	百哈納	puɛ.xɑ.nɑ	pai.xa.na	баҳона
768	قناعت	守分	know one's place	革納額忒	kɛ.nɑ.•.tʰɛ	kɤ.na.ɤ.tʰɤ	қаноат
769	آراميدن	穩重	steady	阿剌米丹	ɔ/ɑ.•.mi.tan	a.la.mi.tan	оромидан / орамидан
770	كرچه	雖是	although	革兒赤	kɛ.•.•	kɤ.əɹ.tʂʰɹ	гарчи
771	غيرت	發志	stimulate aspiration	矮勒忒	ai.lɛ.tʰɛ	ai.lɤ.tʰɤ	ғайрат
772	هنر	本事	capability	虎納兒	xu.nɑ.•	xu.na.əɹ	ҳунар
773	دل كمادى	用心	intention; take care	的勒古媽里	ti.lɛ.ku.muɑ.•	ti.lɤ.ku.ma.li	дилгуморӣ
774	خدمت	侍奉	tend to; look after	黑得黙忒	xɛ.tɛ.•.tʰɛ	xei.tɤ.mwo.tʰɤ	хидмат / хизмат
775	تقديم	進貢	pay tribute to	忒革得尹	tʰɛ.kɛ.ti.•	tʰɤ.kɤ.ti.jin	тақдим
776	ابرالدهر	永遠	forever	阿人鈍得黑兒	ɔ/ɑ.•.•.tɛ.xɛ.•	a.pu.twən.tɤ.xei.əɹ	абадуддаҳр
777	آمان	太平	peaceful and tranquil	阿媽恩	ɔ/ɑ.muɑ.ən	a.ma.ən	амон

As is evident from Table 1, New Persian written in Chinese script provides different (more detailed) information about New Persian vowels than is obtainable from Arabic-script sources. Thanks to this information, we can see, for example, that Timurid Persian, unlike present-day Iranian Persian (see Miller 2012: 167), contrasts the two vowels corresponding with present-day Tajik /i/ and /e/, whose orthographic representations are ‹и›/‹ӣ› and ‹е›, respectively (Table 2).[109]

105 舉保 may mean 'recommend', judging from the meanings of the two glyphs that it comprises.

106 Modern Tajik *пазируфтор* means 'accepter; surety'.

107 遷更 may mean 'migrate' judging from the meaning of the two glyphs that it comprises and the meaning of Modern Tajik *кӯч* 'migration', which, incidentally, is a Turkic loanword.

108 These would be spelled بازگونه, باژگونه, and باشگونه in the Perso-Arabic writing system.

109 Note that this correspondence is not absolute. For instance, نبيره (entry number 183) whose present-day Tajik cognate *набера* is transcribed as 納必勒 /nɑ.pi.lɛ/ in the Berlin Manuscript.

Table 2: A vowel contrast shared by Timurid Persian in Chinese-script transcription (/i/ with /iɛ/) and Tajik (/i/ with /e/)

	Entry word in *zazi*	Timurid Persian in Chinese-script transcription		Tajik	Iranian Persian
4	ستاره	**洗**他勒	**si**.tʰɑ.lɛ	**си**тора	**se**tāre
332	سینه	**洗**納	**si**.nɑ	**си**на	**si**ne
46	زمین	則**米**尹	tsɛ.**mi**.•	за**ми**н	za**mi**n
104	زمستان	即**米**思他恩	•.**mi**.sɿ.tʰɑ.ən	зи**ми**стон	ze**me**stān
102	تابستان	他**比**思他恩	tʰɑ.**pi**.sɿ.tʰɑ.ən	тo**би**стон	tā**be**stān
275	خسبیدن	虎思**比**丹	xu.sɿ.**pi**.tan	хус**би**дан	xos**bi**dan
278	طلبیدن	忒勒**比**丹	tʰɛ.lɛ.**pi**.tan	тал**би**дан	tala**bi**dan
307	بینی	**比**你	**pi**.ni	**би**нй	**bi**ni
692	بسیار	**比**洗呀兒	**pi**.si.•.•	би**cё**р	be**sy**ār
556	سیر	**夥**兒	**siɛ**.•	**ce**р	**si**r
654	سه	**夥**	**siɛ**	**ce**	**se**
434	میوه	**滅**斡	**miɛ**.•	**ме**ва	**mi**ve
575	آمیختن	阿**滅**黑貪	ɔ/ɑ.**miɛ**.xɛ.tʰan	о**ме**хтан	ā**mi**xtan
350	بیمار	**別**媽兒	**piɛ**.muɑ.•	**бе**мор	**bi**mār
428	بید	**別**得	**piɛ**.tɛ	**бе**д	**bi**d
461	بیخ	**別**黑	**piɛ**.xɛ	**бе**х	**bi**x
639	بیررون	**別**魯恩	**piɛ**.lu.ən	**бе**рун	**bi**run
759	بیکار	**別**噶兒	**piɛ**.•.•	**бе**кор	**bi**kār

Table 1 also provides data on the nominal morphology of Timurid Persian. For instance, it reveals an apparent absence of (or vowel reduction in) the *ezāfe* marker in a number of entry words. Notice how خط 'script writing' is transcribed identically in entries 614 and 623 (Table 3).

Table 3: Presence and absence of the *ezāfe* marker in the Chinese-script transcription of *zazi*

	Entry word in *zazi*	Timurid Persian in Chinese-script transcription		Tajik
345	آفت	阿法**忒**	ɔ/ɑ.puɑ.tʰɛ	офат
130	آفت خشک	阿法**梯**戸石克	ɔ/ɑ.puɑ.tʰi.xu.•.•	офати хушк
614	خط	黑**忒**	xɛ.tʰɛ	хатт
623	خط تخقیق	黑**忒**忒黑革革	xɛ.tʰɛ.tʰɛ.xɛ.kɛ.kɛ	хатти тахқиқ

In addition, the Chinese-script transcritption of *zazi* allows an estimation of the positions of Timurid Persian vowels relative to one another. This has been attempted in a recently published paper (Ido 2015: 127–128), from which Figure 1 is reproduced with little modification.

In conclusion, Table 1, in which entries in *zazi* are augmented by reconstructed Ming-period Beijing Chinese readings, provides a wealth of linguistic information on the variety of New Persian that was current in the Timurid court in Samarkand about 600 years ago, and hence is particularly useful in the study of the historical development of New Persian.
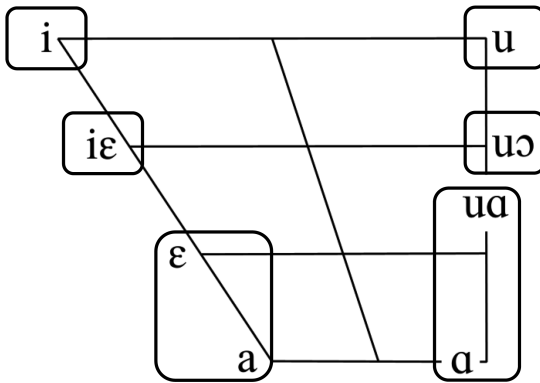
Figure 1: Timurid Persian vowels as they are reflected in the Chinese-script transcription system used in type 1 *zazi*

References

Atwood, Christopher P. 2004. *Encyclopedia of Mongolia and the Mongol empire*. New York: Facts on File, Inc.

Barthold, W. 1987. K̲h̲ānbali̊k̲. In Martijn Theodoor Houtsma, Thomas Walker Arnold, René Basset & Richard Hartmann (eds.), *E. J. Brill's first encyclopaedia of Islam, 1913–1936* (photomechanical reprint edition), 898–899. Leiden: E. J. Brill.

Beijing tushuguan guji chuban bianji zu. (ed.). 1987–[1994]. *Beijing tushuguan guji zhenben congkan 6: huayi yiyu* [Beijing Library collection of ancient books and rare editions 6: *huayi yiyu*]. Beijing: Shumu Wenxian Chubanshe.

Clauson, Gerald. 1972. *An etymological dictionary of pre-thirteenth-century Turkish*. Oxford: Clarendon Press.

Doerfer, Gerhard 1965. *Türkische und mongolische Elemente im Neupersischen*, vol. II, *Türkische Elemente im Neupersischen*. Wiesbaden: Steiner.

Franke, Herbert. 1966. Sino-Western contacts under the Mongol empire. *Journal of the Hong Kong Branch of the Royal Asiatic Society* 6. 49–72.

Haw, Stephen G. 2014. The Persian language in Yuan-dynasty China: A reappraisal. *East Asian history* 39. 5–32.

Honda, Minobu. 1963. ‹Kaikaikan yakugo› ni tsuite [On *huihuiguan yiyu*]. *Hokkaidō daigaku bungakubu kiyō* 11. 1–73.

Ido, Shinji. 2015. New Persian vowels transcribed in Ming China. In Matteo De Chiara & Evelin Grassi (eds.), *Iranian languages and literatures of Central Asia: From the 18th century to the present*, 99–136. Paris: Association pour l'Avancement des Études Iraniennes.

Jarring, Gunnar. 1964. *An Eastern Turki-English dialect dictionary*. Lund: C. W. K. Gleerup.

Kleeman, Julie & Harry Yu (eds.). 2010. *Oxford Chinese dictionary*. Oxford: Oxford University Press.

Kuroyanagi, Tetsuo. 1984. *Perushiago no hanashi* [An account of the Persian language]. Tokyo: Daigaku Shorin.

Lin, Yen-Hwei. 2007. *The sounds of Chinese*. Cambridge: Cambridge University Press.

Liu Yingsheng. 2008. *«Huihuiguan zazi» yu «huihuiguan yiyu» yanjiu* [A study on two medieval Sino-Persian clossaries (*sic*)]. Beijing: Renmin University of China Press.

Lu, Zhiwei. 1988. Ji Xu Xiao «zhongding sima wengong dengyun tujing» [A note on Xu Xiao's *zhongding sima wengong dengyun tujing*]. In *Lu Zhiwei jindai hanyu yinyun lunji*, 54–84. Beijing: Commercial Press.

Miller, Corey. 2012. Variation in Persian vowel systems. *Orientalia Suecana* 61. 156–169.

Muhammadiev, Mardon. 1975. *Luġati muxtasari sinonimhoi zaboni tojikī* [A concise dictionary of Tajik synonyms]. Dushanbe: Maorif.

Nagashima Eiichirō. 1941. Kinsei shinago toku ni hoppōgo keitō ni okeru on'inshi kenkyū shiryō ni tsuite (zoku) [A review of the sources for the study of the early modern Chinese phonological history with a focus on northern varieties of Chinese (part 2)]. *Gengo kenkyū* 9. 17–79.

Norman, Jerry. 1998. *Chinese*. Cambridge: Cambridge University Press.

Odilov, Nodir. 1974. *Mirovozzrenie Džalaliddina Rumi* [The world-view of Jalaluddin Rumi]. Dushanbe: Izdatel'stvo «Irfon».

Paul, Ludwig. 2013. Early New Persian. Encyclopædia Iranica. http://www.iranicaonline.org/articles/persian-language-1-early-new-persian (accessed 30 March 2015).

Pullyblank, Edwin G. 1991. *Lexicon of reconstructed pronunciation in early Middle Chinese, late Middle Chinese, and early Mandarin*. Vancouver: UBC Press.

Satoh, Akira. 1981. Chūko tōkōsetsu nisshōji to pekingo kōgo'on [The literary and colloquial pronunciations in the Peking dialect: With focus on the *ru-sheng* words of the Ancient Chinese *zeng* and *geng* rime-groups]. *Yokohama kokuritsu daigaku jinbun kiyō dai 2 rui gogaku bungaku* 28. 43–64.

Sayd, Mustafa Ajan. 2009. *Dari-English dictionary*. Hyattsville: Dunwoody Press.

Siyiguan. 2013 [1579]. *Hua-i-yi-yü*. Digitalisierte Sammlungen der Staatsbibliothek zu Berlin. http://resolver.staatsbibliothek-berlin.de/SBB000103AF00000000 (accessed 3 August 2015).

Steingass, Francis Joseph. 2012. *A comprehensive Persian-English dictionary (Nataraj edition)*. Springfield: Nataraj Books.

Šukurov, Muhammad Šarifovič, Vladimir Aleksandrovič Kapranov, Rahim Hošim & Nosirjon Asadovič Ma''sumī. 1969a. *Farhangi zaboni tojikī (az asri X to ibtidoi asri XX) I* [A dictionary of the Tajik language (from the tenth century to the beginning of the twentieth century) I]. Moscow: Našriëti «Sovetskaja Ènciklopedija».

Šukurov, Muhammad Šarifovič, Vladimir Aleksandrovič Kapranov, Rahim Hošim & Nosirjon Asadovič Ma''sumī. 1969b. *Farhangi zaboni tojikī (az asri X to ibtidoi asri XX) II* [A dictionary of the Tajik language (from the tenth century to the beginning of the twentieth century) II]. Moscow: Našriëti «Sovetskaja Ènciklopedija».

Tōdō, Akiyasu. 1957. *Chūgokugo on'inron* [Chinese phonology]. Tokyo: Kōnan Shoin.

Tōdō, Akiyasu & Yoshimitsu Kanō (eds.). 2005. *Gakken shin kanwa daijiten* [Gakken great dictionary of Sino-Japanese]. Tokyo: Gakushū Kenkyūsha.

Utas, Bo. 2006. A multiethnic origin of New Persian? In Lars Johanson & Christiane Bulut (eds.), *Turkic-Iranian contact areas: Historical and linguistic aspects*, 241–251. Wiesbaden: Harrassowitz Verlag.

Ye, Baokui. 2001. *Mingqing guanhua yinxi* [Ming-Qing Mandarin phonology]. Xiamen: Xiamen University Press.

Adriano V. Rossi

3 Glimpses of Balochi lexicography: Some iconyms for the landscape and their motivation

Abstract: The speakers of any language, even if at a small extent, concur to change the lexicon, which they have inherited as a whole. They are driven to do that by the necessity of naming something new or optimizing the onomasiological salience of already existing words, with a continuous changing in the way they express concepts. In order to avoid an overloading of the memory system, they are encouraged to recycle what is already existent in the lexicon. Through a small set of associative strategies, people relate a concept which has already been verbalized, with another one which has to be verbalized, producing *lexical changes*. Over time, however, the conceptual motivation which originated a particular designation becomes obscure to speakers. Large scale lexical surveys aid us in discovering recurrent schemas of designating a concept and recovering the relevant motivation for each designation, i.e. its 'iconym' (the Engl. term *iconym* has been currently utilized, e.g., by Joachim Grzega in his contributions to *Onomasiology Online*).

In the general framework of cognitive onomasiology, I have been carrying out since the 1990s (at L'Orientale University, Naples) a project aimed at singling out the different 'pathways' through which natural physical concepts have been designated in the Iranian languages, in order to get insight into the way Iranian speaking peoples have perceived and conceptualized the physical environment which they concurred to change with their millenary activities.

There are several types of associative relations on which lexical innovation relies on; one of these is similarity. The best known process based on similarity is that of metaphor, a process through which we speak of a concept in terms of another, and whose main lines are similarity of shape, similarity of spatial configuration, functional similarity, etc. Since human beings perceive their bodies as an interface between themselves and the surrounding world, the body part lexicon overlaps in many points with those of other conceptual domains; first of all, with the lexicon used to describe the environment.

 Adriano V. Rossi, University L'Orientale

Metaphorical mappings involving human (or animal) body parts as a source, and elements of the landscape as a target, are commonly found in most Iranian languages.

Object of this paper will be a selection of Balochi terms for parts of the human body, variously related to terms used to describe the landscape, studying them from an etymological and areal perspective.

Keywords: linguistics, lexicography, body part terms, Iranian linguistics, Balochi

1 Generalities

The speakers of any language can, at any time, concur to make changes (however minor) in the lexicon they have inherited. They are driven to do that by the necessity of naming something new or optimizing the onomasiological salience of existing words. In order to avoid overloading the memory system, they are encouraged to recycle existing words in the lexicon. Through small associative strategies, people relate a concept that has already been verbalized with another one that has yet to be verbalized, producing *lexical changes*. Over time, however, the conceptual motivation that originated a particular designation becomes obscure to speakers. Large-scale lexical surveys aid us in discovering recurrent schemas of designating a concept and recovering the relevant motivation for each designation, i.e., its 'iconym'.[1]

In the general framework of cognitive onomasiology, we have outlined with our research team at L'Orientale University[2] a project aimed at singling out the

---

1 Cf. Filippone (2006, 365). The English term *iconym*, first introduced by Alinei (1997), has been currently utilized in the subsequent years, particularly by Joachim Grzega in his contributions to *Onomasiology Online*. Alinei's original definition is as follows:

> [B]ecause of the importance of the role of motivation in the genesis of words, I have recently proposed calling it *iconym* (from *icon* + 'name', with the derivations "iconymy", "iconymic", and "iconomastic"), in order to avoid using the much too ambiguous and generic term "motivation" (Alinei 1997c). Only a few linguists, unfortunately, have recently discussed some theoretical aspects of iconymy (e.g., Lakoff and Johnson (1980), Lakoff (1987) and Sanga (1997)). [...] This is exactly what iconyms do, by "representing", as it were, whole concepts. Any new concept that in the process of social communication has become standardized, can thus be collapsed, by means of iconyms, into a new word, allowing us to enrich our knowledge, without changing our abstract, mental categories (Alinei 2003: 108–109).

2 This research is carried out within the frame of the Ethnolinguistics of the Iranian area Project (no. 9710425417), also drawing on lexical material from the Archives of the Comparative Etymological Balochi Dictionary Project (no. MM10422399, hereinafter referred to as *Archive*), both directed by myself and funded by the Italian Ministry of Education at *L'Orientale* University, Naples.

different 'pathways' through which natural physical concepts have been designated in the Iranian languages, in order to get insight into the way Iranian-speaking peoples – and particularly Balochi-speaking peoples – have perceived and conceptualized the physical environment that they concurred to change with their millenary activities. This research has been carried out since the 1990s within the frame of the Ethnolinguistics of the Iranian area project and the Comparative etymological Balochi dictionary project, both of which I direct and which are funded by the Italian Ministry of Education at L'Orientale University, Naples.[3]

To accomplish this work, many years ago we began gathering the relevant lexicon in the Iranian languages, using as sources mostly dictionaries and glossaries and, for a few languages, mostly Western Iranian (including Balochi), information provided by native speakers. The corpus produced so far contains several thousand words of a remarkable interest, many hundreds of which refer to different dialects of Balochi.

1.1 Metaphorical mappings involving human/animal body parts as a source in Balochi

There are several types of associative relations on which lexical innovation relies on; one of these is similarity. The best-known process based on similarity is that of metaphor, a process through which we speak of a concept in terms of another. Since human beings perceive their bodies as an interface between themselves and the surrounding world, the body-part lexicon overlaps in many points with those of other conceptual domains; first of all, with the lexicon used to describe the environment.

Metaphorical mappings involving human (or animal) body parts as a *source*, and elements of the landscape as a *target*, are commonly found in most Iranian languages.[4]

---

3 Two previous studies conducted within this framework are Filippone (2006, 2010), to which the reader is referred. Since the introduction of this methodology in the Iranian studies originates from joint research of Prof. Filippone and myself, practically every concept hinted there (and in many other places) stems from a shared vision (even if not explicitly stated). It is consequently a pleasure for me to state here how much I am indebted to Prof. Filippone for her invaluable support in our common research (and in my life). Special thanks are also due to my former pupil, Dr. Gerardo Barbera, for important unpublished materials from the Bashkardi area.

4 This terminology is according to Lakoff (1987, 276). Conceptual Metaphor Theory, sometimes called Cognitive Metaphor Theory, was developed by researchers within the field of cognitive linguists. Recent developments within this field are treated by Kövecses (2002, 2005) and Evans and Green (2006).

The body-part lexicon overlaps in many points with those of other conceptual domains. First of all, there is the lexicon used to describe the environment. Metaphorical mappings involving human (or animal) body parts as a source, and elements of the landscape as a target, are commonly found in most languages, including Balochi. In the framework of modern onomasiology, which operates in the light of cognitive linguistics, I concentrated on the "pathways" through which different concepts for parts of the landscape have been verbalized, going back (when possible) to the respective source concepts.

This article will describe a small selection of Balochi terms for parts of the human body, variously related to terms used to describe the landscape, which will be analyzed from an etymological and areal perspective. The most common Iranian terms having similar usage, such as *sar*, *pād*, *nyām*, are not included in this article since they have been at least hinted at – even if frequently in a simplistic way[5] – in the iconomastic studies in Iranological literature; a few relatively marginal terms – most of which are unknown even to scholars working in (Indo-)Iranian dialectology – will be briefly treated in order to give an idea of the methodological approach of the research.

2 Iconym: "parts of the body indicating the same relative position in the body as that of a single locational feature in a salient object of the landscape"

(1) Bal. ***barbūnz*** 'hillock' [= *sunṭ*] (Mitha Khan Marri and Surat Khan 1970: s.v.) ◆ EastBal. *barbūnz* 'peak, summit' (Ahmedzai n.d.: s.v.), cf. Psht. *wərb'uz*, Wan. *warbīz* 'muzzle, snout; spur of a hill' (according to Morgenstierne [2003]: s.v. Psht. *wərb'uz* < **fra-* + *poza-* 'nose'), Prs. *bar-pōz* 'the parts around the mouth', Bal. *būz*, *būnz* 'the animal's pointed mouse', with the same composition pattern as *bar-dast* 'shoulder-blade', *bar-čānk* 'hand, fist, hilt of sword', *bar-gaṛ* 'hole, pit', etc., either inner-Balochi, or Pre-Balochi. (Razzaq, Buksh, and Farrell 2001: s.v.) consider *sunṭ* (q.v.) as a synonym of *barbūnz* (East Balochi, from Mitha Khan Marri and Surat Khan 1970).

---

5 Surely not in the case of Wilhelm Eilers, who was an outstanding pioneer in this field of Iranology (cf. his research on the subject in the bibliographies contained in Eilers [1987, 1988]).

Here the pathway seems to go from THE LOWER TERMINAL PART (or perhaps THE POINTED MOUSE) of an animal head to THE TOP OF A MOUNTAIN, if the projected function of the pointed mouse is perceived as a SPUR (the SPUR metaphor is widespread in the mountain lexicon of many different languages, independently from the iconymical history of the term *spur* in each language). The origin of the metaphor could be located outside Balochi if one accepts Morgenstierne's (2003: s.v.) suggestion of a generalization of the iconym from THE PARTS AROUND THE MOUTH to WHAT IS NEAR A DOOR, MOUTH at least in Pashto (where some dictionaries give for *wərbʼuz* 'slave guarding a door', 'land in front or surrounding a gate').

(2) Bal. ***čūṭī*** 'hair, down on the head of a baby' (Mitha Khan Marri and
     Surat Khan 1970: s.v.); Bal. *čūṭī* 'summit, peak (of a hill or mountain)'.
     Razzaq, Buksh, and Farrell (2001: s.v.) give as synonym *sunṭ*, *ṭul*, 'peak of
     a mountain' (Mayer 1909, s.v.); Br. *čoṭī* 'top-knot, tuft. Crest, summit'
     (thus Bray 1934: s.v.), but are all meanings really documented? Cf. Urdu
     čōṭī 'a lock of hair left on the top of the head; crest of a bird; top; peak of a
     mountain etc.'; Si. *čōṭī* 'peak of mountain', 'crest', Sir. *čoṭī* 'peak' in Turner
     (1966: 266, no. 4883).

Four different bases are postulated by Turner (1966: 266, no. 4883) for this lexical family (possibly < Dravidian), but in any case Hindi/Urdu, Panjabi, Siraiki, Sindhi are rather homogeneous in preserving the vocalism *-o-* and meanings ranging from 'topknot, crest' to 'top (of a tree)', 'peak', etc.; Mayrhofer (1956, 3: 396) < Dravidian; Mayrhofer (1992, 1: 546) notes: "Nicht klar", but remarks that in case of Indo-Iranian origin *coḍa-* 'curl' (epic +) should be primary as contrasted with *cūḍa-* 'bulge on a brick' (Śatapatha-Brāhmaṇa +) ("*c°* nur vor urspr. Diphthong lautgesetzlich, *cū́ḍa-* usw. 'mit *ū* für *o*', AiGr I² Nachtr 14").

Here the pathway seems to go from ANY SALIENT/PROTRUDING FEATURE ON/ FROM THE HEAD OF A HUMAN/ANIMAL to THE TOP SECTION OF A MOUNTAIN. In consideration of Psht. *čoṭi* 'uncombed, disheveled', it is reasonable to assume an Indo-Iranian pressure in the iconymic process of the Indo-Iranian frontier languages, having originated in an area in which the focus was on HUMAN HEAD, a metonymic process toward the mountain lexicon.

(3) Bal. ***dīmag***, ***dōmag*** CoBal. *dūmmag* 'čammē ēkirr-o-akirr kuṭṭ itagēn hadd'
     (Hashmi Baloch 2000: s.v.), with the following example: *pōnzē dūmmag =
     pōnzē piḍḍ* 'nasal septum'); IrBal. *dumbag* 'tail', also (politely) 'bottom', an
     Archive informant from Iranshahr, but note that Bal. *dumbag* only means
     'tail fat (of sheep)' (thus correctly Elfenbein [1990: s.v.] and Hashmi Baloch

[2000: s.v.]); cf. Bal. *dīm* 'back, hinder part' (Dames 1891: s.v.); Mayer (1909: s.v.); and one *Archive* informant from Sibi (cf. Filippone 1996: 307); possible etymological connections of Bal. *dīm* with the group of Bal. *dumb*, homogeneously recorded as 'tail', are treated in Filippone (1996: 307–308).

Geographical meanings: 'high place, ascent' (Mitha Khan Marri and Surat Khan 1970: s.v.), also *caṛhāī*, *buṛzay* (Razzaq, Buksh, and Farrell 2001: s.v.: no English meaning but glossed as Bal. *burzag*, Urdu *caṛhāī*) – *Archive* informants: Turbat-1 *dōmmag* perhaps 'foot of mountain'; Turbat-2 *dūmmag* 'ridge of *gwāš*' (foot, middle of a hill); 'that part of *gwāš* having a *šep* (slope) at both sides'; Bālgitar/ Turbat *dūmmag* 'mountain peak running to the plain'. Cf. Larestani *domaga* 'starting part of a valley'.

  Eilers (1988: 291–292) remarks that Dames (1913: 651a, 654, 657) connects the ethnonym *Dōmkī* to the toponym *Dōmbak* in Iranian Balochistan (with difficulties in explaining *ō* as against Bal. *dum*). In *Balochi Race*, Dames (1904: 54) connected the same ethnonym *Dom(b)kī* with the river *Dumbak*. If these ethnonyms/ toponyms have original short vowels, they might be connected with 'tail'; in the place of settlement of the Dombkī, nothing contrasts the association 'tail' with 'slope'. Also in the Pamir toponyms referring to mountain slopes containing *dum* 'tail' (*dumzoj* etc.) are known; cf. Junker (1930: 77–78, 96, 121).

  Here the pathway follows the common experience according to which if the mountain is conceived as a human/animal body, its CAUDAL SECTION is what lies at the foot of the mountain, i.e., its PIEDMONT SLOPE. The origin of the iconymic process may be pan-Iranian (cf. Larestani *domaga* 'starting part of a valley', in which 'starting' points to its lower layer), since Yaghnobi *dumzoj* quoted above is confirmed by Xromov (1975: 33 s.v.), but it seems isolated in East Iranian (all the remaining *dum*-toponyms quoted by Xromov are Tajik); cf. Ossetic *dymæg/ dumæg* in the translated meaning of 'kraj', konec' as stated in Abaev (1958: s.v.).

(4) Bal. *kaš(š)*, recorded as 'armpit' in Mayer (1909: s.v.) and Dames (1891: s.v.), mainly refers to the 'side of the body' or to the 'lateral area just under the ribs', as in Barker and Mengal (1969: s.v.), Elfenbein (1990: s.v.), Razzaq, Buksh, and Farrell (2001: s.v.). Notwithstanding its lexicographical attestations, it seems to be unknown among the East Balochi speakers (Filippone [1996: 311 and n. 80]; *Archive*).

Bal. *kaš* in the sense of 'beside' enters the series of locatives that prototypically refer to the human body sides and the area adjacent to them, i.e., *pahnadā/ bagalā/kašā*. While *pahnadā/bagalā* are found almost everywhere in Balochi,

kašā belongs only to the Southern Balochi lexicon (perhaps also accepted by Western Balochi speakers, even if not actively used by them; cf. Filippone [1996: 190]).

Bal. *kaš(š)* belongs to the lexical family of Av. *kaša-* 'armpit', and is commonly considered of Indo-Iranian origin, cf. Sanskrit (Atharva Veda- Saṁhitā) *kákṣa-* 'armpit' (Mayrhofer [1992: 288], Middle Iranian cognates in Bailey [1979: 56b]). In Balochi it is probably a loanword < Persian, as already stated by Geiger (1891: 453, no. 130).

In view of RaBal. *kač(č)* 'thigh' (Rzehak and Naruyi 2007: s.v.); Bshk. *kač* 'gluteus muscle', *kačak kert* 'to embrace' (syn. *bayal kert*) Barbera (n.d.: s.v.), Sist. *kač* 'thigh', and its geographical projections: *kačč* 'bank of river' (Elfenbein 1990: s.v.); EastBal. *kaččh* 'a piece of flat alluvial ground near the bank of a torrent below the rocks' (Dames 1907: 120); EastBal. *kaččh* 'cultivated land by the side of the river; an island' (Mitha Khan Marri and Surat Khan 1970: s.v.); RaBal. *kič* 'small pieces of land near the bank of a stream' (Ahmedzai (n.d.: s.v., = Brahui according to Ahmedzai, not in Bray [1934: s.v.]); IrBal. *kačč* 'meadow' *Archive* informant from Iranshahr, a (rather old) connection between the *kačč* and *kašš* families is highly probable, and it also possible that it originates in the epic Sanskrit variant *kacchā-* of classic Skt. *kakṣā-* 'girdle'. Skt. (MBh) *kaccha-* 'bank, shore, marshy ground' is continued in Pa. *kaccha-* 'marshy land', Pkt. *kaccha-* 'bank, flooded forest'. New Indo-Aryan (Turner 1966, no. 2618) has Si. *kaco* 'low alluvial lands lying below a bank or hill or lately thrown up by river'; Sir. *kachhā* 'land subject to inundation; alluvial low-lying land where tamarisk grows', *kachhī* 'the alluvial valley of the Indus', and other derivatives with similar meanings. Eilers (1988: 297, 368 n. 226) had already remarked that in the Balochi area, a series of geographical names containing *kač* (e.g., Bābarkač, Kačhī, Nīlīkač, Rūdīān Kač, Kacha Dāman) refer to 'depressions, lowlands'. Since Dzadr. *kackay* 'terrain se trouvant à proximité d'un cours d'eau' (Septfonds [1994, s.v.]) and Wan. *kucaṇá* 'armpit' (according to Morgenstierne [1930: 168; 2003, s.v.] < Khetrani *kucəṇī́* 'armpit') face up Psht. *kše̗*, Wan. *če* 'in', a specialization of *kač* for the geographical meaning and *kaš* for the body/locative lexicon may have arisen in the Balochi area from an older distribution in which the IA *kač*-outcomes spreading westwards along the Ocean coast superseded the Ir. *kaš*-outcomes.

The base denotes the SIDE OF A BODY, and in three-dimensional objects it refers to the two lateral surfaces (if an intrinsic axe is perceived) or to all the vertical axes (if there is none); in bidimensional objects it generically indicates a relationship of proximity, with an emphasis on the localization of the object in the area '(partially) encircled' by the ground.

3 Iconym: "parts of the body presenting functional similarity/similar shape to geographical features"

(5) Bal. ***kump*** 'hunchbacked/hump' (Hashmi Baloch 2000: s.v.); Ahmedzai (n.d.: s.v.) *kub* 'hunchbacked'; *kubbī* (a) 'bent'; (b) 'crookedness'; *kubbō wang* 'person with a bent back'; Brahui *kōmp* 'hump' (according to Elfenbein 1983: 199 < Bal.); Bal. *kumbīy* 'truffle' (type POT according to Morgenstierne [1973: 18], but probably type HUMP according to Rossi 2016: 217); Psht. *kūp* 'crooked, bent in the back'; *kūpaey, kūbaey* 'a hunchback'; *kwab* 'hump' – Indian words like Panǰ. *kubb*, Hind. *kub* 'hump', Panǰ., Sir. *kubbā* 'hump-backed' have influenced the Pashto forms according to Morgenstierne (1927: s.v.), while most recently Morgenstierne (2003: s.v.) separates the Iranian family of Psht. *kwab* (< Ir. **kaupa-*) from IA **kubba-* of Turner (1966, no. 3301); Parachi *kūmbū* 'shoulder' (as protruding from the body?) may (or may not) be connected. Geographically, we have Bal. *kump* 'hillock', also toponym indicating hillocks: *kōp* in *dokop, gwarkop* (from Makran Gazetteer) 'place-names' (Morgenstierne 1946–1948: 289) – probably some unenlarged form of *kōpag* 'shoulder', according to Morgenstierne – toponyms from Makran: Sarbāz *kopk* [=/kōpk/?], Sarāvān *Kupag* [=/kūpag/] (Spooner 1971: 527), with the following annotation: "Names of Baluchi origin – or at least fully Baluchized. These are almost exclusively names of natural features, e.g. rivers, streams, rocks, mountains [. . .], and small areas. These can be seen to suggest the toponymy of a pastoral, nomadic people").

Apparently Central-Iranian dialects have only the geographical metonymical projection:

(5) a. type *g/qomb*: Naini *gom*, *gomb*, *gombu* 'hillock, heap of earth'; Behdinani *qomb* 'clay vessel' (for which see Rossi [2016]) and 'raised ground'; Khunsari *qombeli*, *gombeli* 'relief'; *qombela* 'prominent, raised';

b. type *kope*: Judeo-Isfahani *kope* 'heap'; Judeo-Pers., Yazd *koppo* 'heap'; Khunsari *kopa* 'heap'; *keppeli* 'prominent, raised'; Kermanshahi *qopa* 'prominent, raised', *qomboli* 'prominent, raised'; Kurmanji *qov* 'hump'; Sorani *qubke* (1) 'protruding', 'dome' [= *kubk*]; (2) 'top of mountain';

c. Fārs, Lori and Southeastern coast: Bandari *gambel* 'hill'; Bakhtiari *gomboluk* 'prominent, raised'; Davani *kombor* 'peak, stone relief (hill or mountain)'; Jiroft-Kahnuji *kombar* 'earth hill'.

Here the pathway proceeds from the conception of a prototypically BENT HUMAN SPINE to ANY FEATURE OF THE LANDSCAPE BEING BENT AND THEREFORE STICKING OUT FROM THE HORIZON LINE; this is based on the assumption that the prototypical human body is conceived as lying on the horizon line.

(6) Bal. ***mōl*** Psht. *mowl* 'hump' (Hanley 1981: s.v.). Cf. also Bakht. *mol* 'hump of an oxen'; 'round hill', Kurdish Sor. *milik* 'hunch of camel', etc.; Kurd. Sul. *mol̲* 'piled, heaped'; Naini *koo-mol-kaǰa* 'mountain with crooked neck'; *mol* 'neck'; Shushtari *mol* 'hump'; Yazdi *mol* 'neck', the same in all Fârs dialects, etc.

Apparently a different word is Bal. *mōl* 'a particular Balochi fashion of binding turbans' (Mitha Khan Marri and Surat Khan 1970: s.v.); *mōl* 'a corner of a turban used to cover the face' (Elfenbein 1990: s.v.); *mōl* 'particular way of binding turbans' (Razzaq, Buksh, and Farrell 2001: s.v.); cf. Psht. *mōl* 'tip of a turban'; 'way of wrapping a turban so it covers one's ears' (Pashtoon 2009: s.v.), 'pan du turban avec lequel on se cache le nez et la bouche' (Kabir and Akbar 1999: s.v.); Br. *mōl* 'muffling of chin and ears against the cold'; Sir. *mōl* 'a pad placed on the top of the head for carrying weights'; Si. *mōṛu* 'cock's comb' both < Skt. *mukuṭa-* (also *mauli-*) 'tiara, crest', according to Turner (1966: no. 10144, Skt. < Drav.), "wohl drav". Mayrhofer (2001: s.v.); cf. Burrow and Emeneau (1984: 437, no. 4888).

   Geographically, Bal. *mōl* 'round hill' Mayer (1909: s.v.), cf. *Mol* 'a place' [location uncertain; The Farhang-e joɣrāfiāi-ye Irān (Teheran, 1330, 7: 225) lists a place called Mol near Lar (a city often mentioned in the Balochi epics)] (Barker and Mengal 1969: 270); *Molā* name of a famous pass, and the Western Iranian words such as *mil(e)* in all Kurdish and Luri dialects for 'pass', 'hillock', cf. Mokri (1997: 8–10). WIr. *mōl /mīl* < Ir. **mṛdu-*, SWIr. form ~ Av. *mərəzu-* 'vertebra'; cf. Christensen, Barr, and Henning (1939: 338); Eilers (1987: 14–16, 1988: 371); Mayrhofer (1996: 334, with further literature). Notwithstanding the many difficulties raised by the proposal of Bailey (1979: 337b) to trace to the same Av. base also Skt. *malhá-* 'mit Auswüchsen am Hals versehen' (on which see Mayrhofer [1996: 334]), there may be some connection and/or semantic influence between the Iranian and the Indo-Aryan lexical families.

   If the bases collected here are really connected, the Balochi geographical usages/denominations would prove the antiquity of the metonymy from the HUMP type in Iranian. The pressure from the Indo-Aryan PAD type seems not to have produced geographical metaphors in Indo-Aryan MPrs. *kōf* 'mountain', Av. *kaofa-* 'hump (only in compounds); mountain' is the Iranian parallel par excellence; see all HUMP New Iranian continuants in Hasandust (2011: 343).

(7a) Bal. ***pūnz***, ***pōnz*** RaBal. *pōz* 'nose'; 'protruding part of mountain or anything' (Hashmi Baloch 2000: s.v.); *pōz* 'nose' (Elfenbein 1990: s.v.), *pōnz* 'nose' (Gilbertson 1925: s.v.).

(7b) Bal. ***pūnzīg*** 'heel' (Elfenbein 1990: s.v.); RaBal. *pŭnzuk* 'heel' (Elfenbein 1990: s.v.); *pō(n)zag* 'protruding part' (Hashmi Baloch 2000: s.v.); also RaBal. *pūnz* 'heel' (Elfenbein 1990: s.v.); Co. *pīnz*, EHBal. *pīz*, *pīδ*; CoBal. *pīnz* 'heel' (Razzaq, Buksh, and Farrell 2001: s.v.); *pēnz*, *phēnz*, also *pūnz* 'heel' (Mitha and Surat 1970: s.v.).

Archive informants: *pūnzīk* 'heel' (Kharan), *pūnzuk* (Panjgur-1), *pū(n)z* 'heel' (Turbat-1, Turbat-2, Karachi-1, Karachi-3, Karachi-4 [*pīnz*], Dashtiari [*pīnz*]), cf. CoBal. *pādē pīnz* 'heel' (Karachi-1, Oman), RaBal. *pādpūnz* 'heel' (Kharan-1, Kharan -2), RaBal. *pādē būnz* 'heel'; *pādpūnz* 'heel' (Ahmedzai n.d.: s.v.).

See Br. *būz* 'snout, muzzle; kiss (vulgar); skirt of a hill' (Rossi 1979: 122, no. F22); note Balochi forms with *b°* and the following Western Iranian ones: Prs. *pā-bus* 'heel'; Az. *boz* 'heel'; Gil. *buz*, *buzi*, *pā-buz* 'heel'; Khor. *buzak* 'bone of a horse leg'.

Geographical meanings: RaBal. *pūnz* 'boulder, rock' (Elfenbein 1990, s.v.); *pōz* 'protruding part of mountain or anything' (Hashmi Baloch 2000: s.v.); RaBal. *kōhe pōzag*; syn. *sunṭ* (Hashmi Baloch 2000: s.v.); CoBal. *pūzak* (toponym) 'crest of a mountain of the Makran Range south of Nikshahr' (Pozdena 1978: 78). A Turbat informant (*Archive*) knows the geographical usage of *kōhē pūnz*, but does not know the exact meaning.

According to the common opinion, *pūnz*, *pōnz* (palatalized in *pīnz*, *pēnz* in Coastal Balochi and Eastern Balochi) 'heel' – and their derivatives in *-ag*, *-uk*, *-īk*, *-īg* – are original Balochi developments (cf. Geiger [1890: 142, no. 306], with doubts of Morgenstierne [1927: 57, 1932: 49, 2003: 63], Benveniste [1955: 300]), while Bal. *pōz*, *pūnz* 'nose' is a borrowing < Prs. *pōz* 'snout, beak', also 'mouth area', with secondary nasalization (thus, e.g., Korn 2005: 216, cf. Korn 2005: 203; Geiger 1890: 142, no. 310).

In any case, it seems hardly tenable (because of its isolation in Iranian, phonetic grounds, and semantic reasons) connecting **pauk-* KISS (documented only by Khotanese) and Prs. *pōz* 'snout, beak' as assumed by Bailey (1979: 250b); cf. also Korn (2005: 203 n. 139), or hypothesizing **faź-*, **fāź-*, **fauź-*: *fuź-* / **pauź-* : *puź-* or **fi̯auź-* / **pi̯auź-* LOWER PART OF FACE to explain Prs. *pōz* 'snout, beak', as assumed by Rastorgueva and Èdel'man (2007: 49–51); a connection of Prs. *pōz*, Bal. *pō/ūnz* 'nose', and Bal. *pūnz* 'heel' remains possible in view of their proto-typical shape/function (as assumed here). The series of labels for 'a broad surface of the body, the front, face and breast or the back', already pointed out by Bailey (1967: 179–180) and summarily treated by Rossi (1998: 407–409), possibly specialized in some Iranian languages as a protruding body part, seems to

belong to a network of terms characterized by the amplitude of the attested forms, many of which open to geographical transfers. To the Balochi geographical metaphors one could add Kurm. *poz* 'cape; headland'; Sarvestani *puze* 'spur of mountain'; Larestani *pūza* 'spur of mountain'; Psht. *poza*, Waz. *pēza*, Wan. *pīza* 'nose'; 'peak of a mountain' (Morgenstierne 2003: s.v.).

Here the pathway proceeds from ANY PROTRUDING PART OF THE HUMAN/ ANIMAL BODY towards ANY FEATURE OF THE MOUNTAIN LANDSCAPE APPEARING AS PROTRUDING from the massif. If one arranges in a scale of protrusion Bal. ***mōl***, ***pūnz***, ***sunṭ*** as projected onto the landscape, the coefficient of roundness decreases and that of pointedness increases.

(8) Bal. ***sunṭ, suṭ*** 'beak', 'sting' and 'chin' in Hashmi Baloch (2000: s.v.); RaBal. *sunṭ* (a) 'trunk', (b) 'beak' (Rzehak and Naruyi 2007: s.v.); MwBal. *sunṭ* 'beak' (Elfenbein 1963: s.v.), PrsBal. *sunṭ* (Spooner 1967: 68) 'beak, bill'; *sunṭ* 'beak, bill; sting (of a mosquito)' (Elfenbein 1990: s.v.), also *sunt*; *sunṭī* 'beaked, stinger', *sunṭīg* 'a fierce mosquito' in S. W. Makrān (Sarawani); Mirjave *sūnṭ* 'beak' ('animal's mouth') (Coletti 1981: s.v.); *sunṭ* 'beak' (Barker and Mengal 1969: 30); *suṭ*, *suṭh*, *sunṭ* 'elephant's trunk; snout; bank'; 'hillock' (s.v. *barbūnz*) (Mitha Khan Marri and Surat Khan 1972: s.v.); *sunṭ* 'peak, summit; beak, bill; the trunk of elephant' (Ahmedzai n.d.: s.v.); EastBal. *sut* 'spur of a mountain' (Gilbertson 1925: s.v.), EastBal. *sut* 'spur of a mountain running down into a plain' (Dames 1891: s.v.); *sut* 'spur of mountain run to plain' (Mayer 1909: s.v.); *sunṭ* 'peak; summit' (Razzaq, Buksh, and Farrell 2001: s.v.; give syn. *ṭul*); PrsBal. *sunt* 'Bergsporn' (Pozdena 1978: 78), in toponyms: Širuksunt 'Bergsporn des Chahbahar Plateaus in Tiskupan); Bal. *sunṭ* 'bottom of a hill sloping into a beak' (Ata 1968: 142).

Archive informants: *sunṭ* 'lip' (= *lunṭ*) (Iranshahr-1), 'upper lip' (Oman); 'mouth and chin' (Turbat-2); 'chin' (Karachi-1, also knows *zanūk*), Karachi-4 (=*zanīk*) 'chin' (the same as *zanuk*, but mostly referring to birds, considered impolite with reference to human beings) (Turbat-1); 'mouth area' (Dashtiari), 'chin' [Panjgur-1]; IrBal *sonṭṭ* 'lip'; Turbat-2: *sunṭ* 'top of mountain if not rounded' (in this case it would be *sar*, *ṭul*).

Compare Br. *sunṭ* 'beak, muzzle; projecting corner; bottom of a hill sloping into a beak' (Bray 1934: s.v.; Rossi 1979: 49, no. A357); cf. Sir. *suṇḍ*, Si. *sūṇḍḍhi*, Skt. *śuṇḍā* 'trunk, proboscis'; Sist. *sont* 'muzzle' (with reference to human beings only when distorted).

Mayrhofer (1996: 426) hesitates in attributing Skt. *śuṇḍā-* 'trunk, proboscis' to a common Indo-Iranian base. Tremblay (2005: 426) assumes a base **sundika-* 'fauces' (> Khot. *ṣumca-* 'beak', Waxi *šendⁱk* (Lorimer 1902: s.v.), *šənḍg* 'gums of mouth' (Steblin-Kamenskij 1999: s.v.); NPers. *šand* 'beak', Ved. *śunda-* 'tusk') and adds (Tremblay 2005: 426 n. 28):

> The following facts militate against a direct borrowing of the Sakan word from Indian: 1. The meanings diverge; 2. The word for 'beak' is attested in Persian; 3. It is enlarged by an -*ika*-suffix in Khotanese and Waxi; 4. The Khotanese word has ṣ, not ś. If the Iranian word were a borrowing from Indian, it must be a very early one. The Indian lexeme was later borrowed in Sogdian B *šnth* 'trunk', and through Dardic (Khowar *šūn*, Tir. *šuṇḍ* 'lip'), in Šughni *šand* < **šundā*, Parachi Pashto *šǔṇḍ* 'lip'.

As in the case of **_pūnz_** above, the pathway here also proceeds from ANY PROTRUDING PART OF THE HUMAN/ANIMAL BODY toward ANY FEATURE OF THE MOUNTAIN LANDSCAPE APPEARING AS PROTRUDING from the massif. The notion of CONNECTIVITY (from one part to another of the mopuntani slope as from one part to another of the mouth/nose area in the face/snout) seems residual in some scattered Balochi evidence; in any case no other Iranian language documents both the bodily and geographical meanings in living usage (NPers. *šand* 'beak' is doubtful).

While in diachronic cognitive onomasiology the main strategies that exist in a language sample for conceptualizing and verbalizing a given concept are investigated, with the aim of explaining them against a cognitive background in terms of salient perceptions, prominency, etc., the iconymic (motivational) sequence reconstructed in the few examples commented on above is, by definition, consolidated, being as it were crystallized in the name itself, just as a fossil is embedded in the surrounding matter. If its chronological span reaches some point in ancient history, we can be sure of its relative antiquity.

No one would devalue the potential of this approach to the reconstruction of the cultural landscape in an area of such intensive multilingualism as that of the Indo-Iranian Frontier languages.

Abbreviations

Av.	Avestan
Az.	Azari
Bal.	Balochi (CoastalBal., IranianBal./ PrsBal., EastHillBal., MarwBal., RakhshaniBal.)
Br.	Brahui
Bshk.	Bashkardi
Drav.	Dravidian
Dzadr.	Dzadrani
Gil.	Gilaki
IA	Indo-Aryan
Ir.	Iranian

Khot.	Khotanese
Kurm.	Kurmanǰi
Khor.	Khorasani
Pa.	Pali
Panǰ.	Panǰabi
Pkt.	Prakrit
Prs.	Persian
Psht.	Pashto
Si.	Sindhi
Sir.	Siraiki
Sist.	Sistani
Skt.	Sanskrit
Sor.	Sorani
Sul.	Suleimani
Wan.	Wanetsi

References

Abaev, Vasilij I. 1958–1989. *Istoriko-ètimologičeskij slovar´ osetinskogo jazyka*. 4 vols. (Moskva) & Leningrad: Izd. AN SSSR (vol. 1), Nauka (voll. 2-4)

Ahmedzai, Agha Nasir Khan. n.d. Unpublished notes for a Balochi-Brahui-Urdu-English dictionary. Rakhshani dialect (including dialectal variants), lists more than 15,000 words and phrases. A copy of the manuscript is at the Archives of the Comparative Etymological Balochi Dictionary Project (compiled in the 1970s and 1980s). On the first page of the manuscript is the following annotation: "checked by the late Mir Gul Khan Naseer 7/9/83".

Alinei, Mario. 1997. Principi di teoria motivazionale (iconimia) e di lessicologia motivazionale (iconomastica). In *Lessicologia e lessicografia, Atti del XX Convegno della Società Italiana di Glottologia* (Chieti-Pescara, 12–14 Ottobre 1995). Roma: Il Calamo. 9–36.

Ata Shad (ˤAṭā Šād). 1968. *Balōčināma* [in Urdu]. Lahore: Markazi Urdu Board.

Bailey, Harold W. 1967. *Prolexis to the book of Zambasta*. Cambridge: Cambridge University Press.

Bailey, Harold W. 1979. *Dictionary of Khotan Saka*. Cambridge: Cambridge University Press.

Barbera, Gerardo. n.d. Unpublished notes for a Molkigâl Baškardi-English dictionary, 1,750 words.

Barker, Muhammad A. & Aqil Khan Mengal. 1969. *A course in Baluchi*. Vols. 1–2. Montreal: Institute of Islamic Studies, McGill University.

Benveniste, Emile. 1955. Études sur quelques textes sogdiens chrétiens (I). *Journal asiatique* 243. 298–335.

Bray, Denis. 1934. *The Brāhūī Language*, part II – *The Brāhūī problem*; part III – *etymological vocabulary*. Delhi: Manager of Publications.

Burrow, Thomas & Murray B. Emeneau. 1984. *A Dravidian etymological dictionary*. Oxford: Clarendon Press.

Christensen, Arthur, Walter Henning & Kaj Barr. 1939. *Iranische Dialektaufzeichnungen aus dem Nachlass von F. C. Andreas*, Abhandlungen der Gesellschaft der Wissenschaften zu Gottingen, III/11. Berlin: Weidmannsche Verlagsbuchhandlung.

Coletti, Alessandro. 1981. *Baluchi of Mirjave (Iran)*. Roma: Edizioni A.C.

Dames, Mansel L. 1891. *A text book of the Balochi language, consisting of miscellaneous stories, legends, poems and a Balochi-English vocabulary*. Lahore: Punjab Government Press. Part IV, 1–117.

Dames, Mansel L. 1904. *A historical and ethnological sketch*. London: The Royal Asiatic Society.

Dames, Mansel L. 1907. *Popular poetry of the Baloches*. Vols. 1–2. London: The Royal Asiatic Society.

Dames, Mansel L. 1913. Balōčistān. In *The encyclopædia of Islam: A dictionary of the geography, ethnography and biography of the Muhammadan peoples*. Vol. 2: B–D. Leiden: E. J. Brill. 625–640.

Eilers, Wilhelm. 1987. *Iranische Ortsnamenstudien*. Wien: Verlag der österreichischen Akademie der Wissenschaften.

Eilers, Wilhelm. 1988. *Der Name Demawend*. Hildesheim, Zürich & New York: Olms Verlag.

Elfenbein, Josef. 1963. *A vocabulary of Marw Baluchi*. Naples: Istituto Universitario Orientale.

Elfenbein, Josef. 1983. A Brahui supplementary vocabulary. *Indo-Iranian Journal* 25. 191–209.

Elfenbein, Josef. 1990. *An anthology of classical and modern Baluchi literature*. II: Glossary. Wiesbaden: Harrassowitz.

Evans, Vyvyan & Melanie Green. 2006. *Cognitive linguistics: An introduction*. Edinburgh: Edinburgh University Press.

Farhang-e joɣrāfiāi-ye Irān. 1328–1332.10 vols. Tehran.

Filippone, Ela. 1996. *Spatial models and locative expressions in Balochi*. Naples: Istituto Universitario Orientale.

Filippone, Ela. 2006. The body and the landscape. Metaphorical strategies in the lexicon of the Iranian languages. In Antonio Panaino & Riccardo Zipoli (eds.), *Proceedings of the 5th conference of the Societas Iranologica Europaea held in Ravenna, 6–11 October 2003*. Vol. II. *Classical & contemporary Iranian studies*. Milano: Mimesis. 365–389.

Filippone, Ela. 2010. *The fingers and their names in the Iranian languages*. Wien: Verlag der österreichischen Akademie der Wissenschaften.

Geiger, Wilhelm. 1890. Etymologie des Balūčī. *Abhandlungen der Königlichen bayerischen Akademie der Wissenschaften*, I. Cl. 19/1. 105–153.

Geiger, Wilhelm. 1891. Lautlehre des Balūčī. *Abhandlungen der Königlichen bayerischen Akademie der Wissenschaften* 19/2. 397–464.

Gilbertson, George W. 1925. *English-Balochi colloquial dictionary*. Hertford: Stephen Austin & Sons.

Hanley, Barbara. 1981. *Concise English-Pushto dictionary*. Quetta: Afghan Publishers.

Hasandust, Mohammad. 2011. *A comparative-thematic dictionary of the Iranian languages and dialects*. Tehrān: Farhangestān-e zabān va adab-e fārsi.

Junker, Heinrich F. J. 1930. Arische Forschungen. Yaghnobi-Studien I: Die Sprachgeographische Gliederung des Yaghnōb-Tales. *Abhandlungen der philologisch-historischen Klasse der sächsischen Akademie der Wissenschaften* 41/2.

Kabir, Habib & Wardag Akbar. 1999. *Dictionnaire pashto-français*. Paris: L'Asiathèque.

Korn, Agnes. 2005. *Towards a historical grammar of Balochi. Studies in historical phonology and vocabulary*. Wiesbaden: Reichert.

Kövecses, Zoltán. 2002. *Metaphor: A practical introduction*. Oxford: Oxford University Press.

Kövecses, Zoltán. 2005. *Metaphor in culture: Universality and variation*. Cambridge: Cambridge University Press.

Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago.

Lorimer, John G. 1902. *Grammar and vocabulary of Waziri Pashto*. Calcutta: Office of the Superintendent of Government Printing, India.

Mayer, Thomas J. L. 1909. *English-Biluchi dictionary*. Lahore: Sheikh Mubarak Ali.

Mayrhofer, Manfred. 1956–1980. *Kurzgefasstes etymologisches Wörterbuch des Altindischen*. Vol. III, 1976. Heidelberg: Winter.

Mayrhofer, Manfred. 1992–2001. *Etymologisches Wörterbuch des Altindoarischen*. Vol. I, 1992; Vol. II, 1996; Vol. III, 2001. Winter: Heidelberg.

Mitha Khan Marri (Miṭhā Xān Marrī) & Surat Khan (Sūrat Xān). 1970. *Balochi-Urdu Lughat* [in Urdu]. Quetta: Balochi Academy.

Mokri, Mohammad. 1997. *Le nom de « vallée » dans les toponymes iraniens*. Paris-Louvain: Éditions Peeters.

Morgenstierne, Georg. 1927. *An etymological vocabulary of Pashto*. Oslo: I komisjon hos Jacob Dybwad.

Morgenstierne, Georg. 1930. The Wanetsi dialect of Pashto. *Norsk tidsskrift for sprogvidenskap* 4. 156–175.

Morgenstierne, Georg. 1946–1948. Balochi miscellanea. *Acta Orientalia* 20. 253–292.

Morgenstierne, Georg. 1973. *Irano-Dardica*. Wiesbaden: Reichert.

Morgenstierne, Georg. 2003. *A new etymological vocabulary of Pashto*. Compiled and edited by Josef Elfenbein, David N. MacKenzie & Nicholas Sims-Williams. Wiesbaden: Reichert.

Pashtoon, Zeeya A. *Pashto-English dictionary*. Hyattsville: Dunwoody Press.

Pozdena, Hans. 1978. *Das Dashtiari-Gebiet in Persisch-Belutschistan. Eine regional-geographische Studie*. Wien: Verlag A. Schendl.

Rastorgueva, Vera S. & Džoj I. Èdel'man. 2007. *Ètimologičeskij slovar' iranskix jazykov*. Vol. III. Moskva: Izdatel'skaja firma « Vostočnaja literatura » RAN.

Razzaq, Abdul, Mula Buksh & Tim Farrell. 2001. Unpublished manuscript *Balochi Ganj. Balochi-English dictionary*, 19,631 entries, fourth draft.

Rossi, Adriano V. 1979. *Iranian lexical elements in Brāhūī*. Naples: Istituto Universitario Orientale.

Rossi, Adriano V. 1998. Ossetic and Balochi in V. I. Abaev's Slovar'. In *Studia iranica et alanica. Festschrift for Prof. Vasilij Ivanovič Abaev on the Occasion of His 95th Birthday*. Rome: IsIAO. 373–431.

Rossi, Adriano V. 2016. Problemi di etimologia areale nel Mediterraneo orientale: gr. κύμβαχος nel suo retroterra asiatico. *Linguarum varietas. An International Journal* 5. 211–227.

Rzehak, Lutz & Bedollah Naruyi (eds.). 2007. Abdurrahman Pahwal, *Balochi Gâlband. Balochi, Pashto, Dari & English dictionary*. Kabul: Al-Azhar Book Co.

Hashmi Baloch, Sayad. 2000. *Sayad Ganj. The first Balochi dictionary* [in Balochi]. Karachi: Sayad Hashmi Academy.

Septfonds, Daniel. 1994. *Le Dzadrâni. Un parler pashto du Paktyâ (Afghanistan)*. Louvain-Paris: Peeters.

Spooner, Brian. 1967. Notes on the Baluchī spoken in Persian Baluchistan. *Iran* 5. 51–71.

Spooner, Brian. 1971. Notes on the toponymy of Persian Makran. In C. E. Bosworth (ed.), *Iran & Islam. In memory of the late Vladimir Minorsky*, Edinburgh: Edinburgh University Press. 517–533.

Steblin-Kamenskij, Ivan M. 1999. *Ètimologičeskij slovar' vaxanskogo jazyka*. Sankt-Peterburg: Peterburgskoe Vostokovedenie.

Tremblay, Xavier. 2005. Irano-Tocharica et Tocharo-Iranica. *Bulletin of the School of Oriental and African Studies* 68 (3). 421–449.

Turner, Ralph L. 1966. *A comparative dictionary of the Indo-Aryan languages*. London: Oxford University Press.

Xromov, Aleksandr L. 1975. *Očerki po toponimii i mikrotoponimii Tadžikistana*, 1. Dušanbe: Irfon.

Martin Schwartz

4 On some Iranian secret vocabularies, as evidenced by a fourteenth-century Persian manuscript

Abstract: A fourteenth-century manuscript in Tashkent gives, in the margins of five pages, an anonymous Persian treatise on secret forms of communication. The treatise comprises a vocabulary for an underworld Shi'ite argot; also a series of Arabic kennings used by a sect of worshippers of 'Ali; a description of an alphabetic-numerological code, with illustration of its use; and verses in the Shi'ite argot. With the exclusion of details on the alphabetic-numerological code, this article will discuss various aspects of the secret vocabularies provided, which are of interest for the broader linguistic and sociological history of Persian.

Keywords: Iranian argots, jargons, Persian codicology, Shi'ites, Judeo-Iranian, Jews, Loterai, Aramaic, Hebrew, Gypsies, Old Iranian, Middle Iranian, New Iranian, Iranian etymology, Persian lexicography, Proto-Indo-European, Banu Sasan, worship of 'Ali, logograms, Central Asia

1 Introduction

For my account of Iranian secret vocabularies I shall use, as a springboard to related bodies of data, a medieval Persian treatise called كتاب ساسيان بكمال *Ketāb-e Sāsīān bekamāl* 'The Book of Accomplished Grifters' (henceforth *KS*). It is written on the margins of five pages of an unrelated text in a manuscript miscellany dated 1344 CE, kept in the Albiruni Center for Oriental Manuscripts, Tashkent State Institute of Oriental Studies. (I thank Professor Aftandil Erkinov of the National University of Uzbekistan, Tashkent, for kindly providing me with relevant photoscans upon my request.) The manuscript is listed in *Sobranie vostochnyx rukopisej Akademii Nauk Uzbekskoj SSR* (Catalogue of Oriental Manuscripts of the Uzbek Akademy of Sciences of the Uzbek SSR), Vol. I (Tashkent 1952, 196–97); the relevant text is on folios 74–77.

The *Sāsīān* of the title is the plural of Persian ساسى *sāsī* 'grifter, beggar, parasite, rogue' from ساس *sās* 'bug, louse, flea', cf. the name of the old Arab underworld, the *Banū Sāsān* 'Sons of S.', ennobled by association with the

Martin Schwartz, University of California–Berkeley

eponym of the Sasanian dynasts. *KS* in fact takes its name from what may now be described as the first part of a four-part tract on exclusionary forms of communication, of which at least the second and third parts were based on the anonymous author's personal experience. Here then, for the first time, is the treatise's four-part scheme: Part One, in nine topical chapters (*bāb*), is devoted to an argot of Shiʿite *Sāsīān*, with each argot word glossed in Persian in a different color of ink. Part Two, again in different colors, gives glossed expressions of certain ʿAli worshippers. Part Three may here be merely described as an ابجد *abjad* '(alphabetic-numerological) code' for private oral communication. Part Four continues verses from the end of Part One in the Shiʿite *Sāsīān* argot. Only from Part One was linguistic material cited by Ivanow (1922), who gave random words he managed to note, having briefly borrowed the manuscript from an unscrupulous dealer in Qarshi. Troickaja (1948), who found the same manuscript archived in Tashkent, concentrated on correlating the Shiʿite argot of *KS* Part One with the Abdoltili argot of Uzbekistan's Gypsies, mendicant preachers, and itinerant reciters. Schwartz (2014) provided a most extensive number of citations of the argot terms and their glosses in Part One (and Part Four), mainly showing origination of many of the argot words (including those of twentieth-century itinerants of Iran and Central Asia) from the earlier form(s) of the exclusionary jargon of Iranian Jews, the modern remnants of which vocabulary were collected by Yarshater (1977).

2 Vocabulary of the على الله *ʿAlī Allāh*

I shall now give examples of the expressions used by the ʿAlī-worshippers (*ʿAlī Allāh*) according to *KS* Part Two. Like the Old Germanic kennings, this exclusionary vocabulary consists of enigmatic metaphorical phrases, in this instance in Arabic, with Persian glosses:

(1) عرش الشيطان *ʿarš al-šayṭān* 'Satan's throne' = كامه *kāme* 'desire'. Via association with *kāme* is juxtaposed:

(2) ابو *al-hāris* الهارس* *abū *al-hāris* 'father of the *grinder' = آب كامه *āb kāme* 'a digestive liquid' (whose initial preparation involved the grinding of a dried fermented bread [Hosseinzadeh et al. 2013]). The manuscript has *al-hāriš* or *al-hādiš*; for copyist's errors in *KS*, cf. Part One فتنه *fetne* 'sedition' for تشنه* *tešne* 'thirsty' (juxtaposed with *gorosne* 'hungry' and glossing argot بر ميا 'thirsty' < 'waterless' < Aramaic *bar* + *mayyā*); cf. below on ساوته *sāūte* 'old person' and تاز *tāz* 'boy, catamite'. Arabic *abū* 'father of-' indicating association, occurs in further examples, the next three of which require disambiguation of the governed Arabic noun:

(3) ابو العباس *abū al-ʿabbās* 'father of the frowner' (not '-of the lion') = سرکه *serke* 'vinegar'.

(4) ابو الفرج *abū l-faraj* f. 'of the aperture' (not '-of the vulva') = کلید *kelīd* 'key'.

(5) ابو الجامع *abū l-jāmiʿ* f. 'of the gathering' (not '-of the mosque') = سفره *sofre* 'table(cloth)'.

(6) شهید بن شهید *Šahīd bin Šahīd* 'Martyr, son of Martyr') = بره *barre* 'lamb'.

(7) نبات القش *nabāt al-qašš* 'vegetation of the straw' = خایه مرغ *xāye-ye morɣ* '(chicken) egg'. The latter kenning may in part be backgrounded in the origin of *nabāt* 'vegetation' < 'to germinate', but the kenning is (also) visual. The final example is visual, and in fact cartoonish:

(8) حمید الکوسج *Ḥamīd al-Kūsaj* 'Hamid the Swordfish' = نای زن *nāy-zan* 'flute player'.

2.1 *KS* argot and Iranian Jewish jargon

As for the first and last parts of the *Ketāb*, in Schwartz (2014) I gave a long and varied account of how the Ketāb's Shiʿite argot and the twentieth-century argots of itinerant groups of Iran and Central Asia in their vocabularies greatly reflect the exclusionary speech of Iranian Jews in its older phase, in which, inter alia, words of Jewish Aramaic origin predominated over words from Hebrew. I shall here revisit a small part of that material, but also bring out new observations, and introduce material bearing on Iranian (non-Jewish) aspects of the relevant vocabularies.

The term (ی)لوترا <lwtrʾ(y)> is attested in Persian literature from the tenth century onward for some kind of cryptic speech. The same term (vocalized *lŏterāʾī, lūtrāʾī*) with variants has been used by Jews of Iran and Afghanistan for their exclusionary vocabulary, only to keep gentiles (and children) from understanding. As is true for the gentile argots, the morphology is that of ordinary Iranian speech. Thus *KS* has بهز <bhz> (Pers. برو *borou*) second person singular imperative of هزیذن <hzyδn> *hezīδan* 'to go', parallel to the Jewish Loterāʾi (henceforth JLot.) of Nehavand *be-hez* 'id.', infinitive *hezidan*; similarly formed are *KS* بنوند <bnwnd> 'give!' (Pers. *be-deh*), JLot. Mashhad *be-nund* 'id.' Cf. further KS نتکن <ntkn> 'don't make' (Pers. *ma-kon*), JLot. Mashhad *be-teken* 'make, fix!' Note also, in the past tense, KS می دهلم <my dhlm> 'I fear' (Pers. می ترسم *mī-tarsam*) and occurring in a small group of verses appended to the first part of *KS*; cf. *mī-dahlad* 'he fears' in a regionally unspecified JLot., and Tajik Gypsy Jugigi argot *me-dahlom* 'I fear'.

The above citations of the *KS* are among the Aramaic-originated forms that constitute the majority of verbs in that work's Shiʿite argot. These verbs, and other words in the *KS* (and later itinerant argots), with their correspondents in JLot., support the derivation of the designation *Loterāʾī, Lūtrā(ʿī)*, etc., from Heb. *Lo-Tōrāh* 'not [in the chief language of the] Torah', i.e., 'not Hebrew' = 'Aramaic'. Further confirmation comes from the *KS* itself, which calls the language of the argot verses زبان سوری *zabān-e sūrī* 'the Syrian language', corresponding to the older Jewish designation for Aramaic.

The words from Aramaic that passed from Jewish exclusionary vocabularies into the gentile argots inter alia show a selection from the various regional forms of JLot. For example, *KS* <bhz> 'go!' agrees in its aspiration with JLot. Nehavand *hez-* vs. JLot. Golpayegan *ez-* < Aram. *ʿazl-* 'going'. Against both, for 'to go', JLot. Shiraz, which often differs from the more northerly JLot. has, *gāledan*, also found in *KS* in synonymy with <hzyδn>, as كاليدن <kʾlydn> /*gālīdan*/, < Aram. *galy-* 'going out'. *KS* <tkn->, variants <tkyn->, <tykyn-> 'to make', etc., corresponds to JLot. Mashhad *teken-*, Herat, Kabul, *tikin-* 'to fix, make', but JLot. Shiraz *taːn* 'id.' shows a different outcome of the Aram. etymon: *taqqen* 'to establish, fix' > JLot. Mashhad *teken-*, etc., with **q > k*, but via **tayen-* or the like, > JLot. *taːn*. *KS* <nwnd-> (inf. نونديدن, Pers. دادن *dādan*) 'to give' corresponds to *nund-* 'to give' in the aforementioned axis JLot. Mashhad-Herat-Kabul < OAram. **nudn-* 'a gift', but JLot. Shiraz *av-*, Borujerd *ab-* 'to give' from Aram. *haβ* 'give!', of which a JLot. form, still with *h-*, yielded in Tajik Gypsy argots Jugigi *hob-/how-* and "Samarkand Lʸuli" *hob-/hov-* 'to give'. *KS* بيسه <bysh> 'egg' (Pers. خايه مرغ *xāye-ye morɣ*) from Heb. *bēyṣāh* 'id.' contrasts with JLot. Shiraz *bika* 'id.', whose *k* reflects an Old Aramaic phoneme written as *qoφ* in the early Achaemenid period (later /ʕ/).

The *KS* selection of regionally diverse JLot. words is again shown from words of Iranian origin. *KS* جهستن <jhstn> 'to see' (Pers. *dīdan*), i.e., /*čehestan*/ *vel sim.* (with *-estan* as in Pers. *dānestan* 'to know'), بجه <bjh> /*be-čeh*/ *vel sim.* 'see!' (Pers. بين *bebin*) corresponds with **čeh-*, which underlies JLot. Kashan *če-(V-)/čā-(C-)* 'know, see', and goes back to Old Iranian *čaiθ-* (Old Avestan *cōiθaṯ*) 'to perceive'. The latter contrasts with its synonymous OIr. variant etymon *čait-*, reflected on one hand by JLot. Shiraz *čed-* 'know, understand', and on the other by JLot. Yazd *čer-* (for *-r- < *-t* Judeo-Yazdi *šer- < *šyuta-* 'gone'), also Kermani and further Isfahan 'see, know, understand, recognize'.

In the same semantic field we have, in the above noted easterly JLot. axis, Mashhad, Herat *ruj-* (and Golpayegan *rej-*) 'to see, know', Kabul *ruč-* 'to look', not reflected in the argots. Derivation from OIr. *rauč* as 'to be illumined' would phonologically parallel, deriving from OIr. **(-)hačaya-* the verbs JLot. Shiraz *āj-*, Herat *huj-* (**hoj-*) 'to come' from old middle voice form 'to lead oneself', cf. Avestan *hācaiia-* (*Yasna* 5.18, etc.) 'to lead, direct, persuade'.

From a causative of the JLot. verb (via *'make come' > 'bring forth, produce') comes North Iranian Gypsy argots *ajon-* 'to make' (and 'to come'), Astarabad Gypsy "*hedjonddan*" 'to make'. With the semantics of Av. *hācaiia-* 'to persuade' and *upaŋhācaiia-* 'to come to agreement with somebody', **hāčaya-*, with preverb **abi* (and **upari*?) are derivable JLot. Nehavand *viāj-*, Mashhad *velāj-* 'to sell, finish (a deal with someone)'.

2.2 *KS* argot and the Persian lexicon

We now turn to words attested in *KS* of which equivalent forms passed from argot into Persian, where they are attested in early lexica and poetry (note especially Qarī' al-Dahr, cited by Asadī). Examples (9) and (10) survive in Persian speech. Note that examples (14) and (15) are attributed to زبان آسیان, i.e., *zabān-e *sāsīān* 'language of the *Sāsīān*'. Of these two examples, (14) *ṣ'bwth* differs somewhat semantically from *KS *s'wth* as well as formally, showing a different background of transmission of forms, which parallel the diversity in the Gypsy argot forms. For twentieth-century argots of Persophone itinerants, see Schwartz (2014) in the references.

(9) *KS* نهور ‹nhwr›: (a) 'eye' (Pers. چشم *čašm*). (b) 'light': نهور تیکینه ‹nhwr tykynh› '(day)light-making' (Arab. فجر المنصور *fajr al-manṣūr* 'the victor's dawn') = Pers. پیروزی *pīrūzī* 'victory'; the starting point is probably *pīrūzī* (which is from OIr. *pari-aujah-*) wrongly associated with Pers. روز *rūz* 'day'. (c) 'blind' (Pers. کور *kūr* 'blind'). As to the twentieth-century argots of Iran and Central Asia: 'eye' is found for *nuhur* among the Gypsies and mendicant dervishes of eastern Iran, *náhur* among the Gypsies of Osof, Fars, and Kerman, as well as Gypsy Toshmal and Luti musicians, and for *nhūr* among the Arāk Gypsies. Both meanings, 'eye' and 'blind', are found among the Tajik Jugigi and "Samarkand Lʸuli" Gypsies. Finally *nuhur* means 'blind' in the argots of the Tajik Arabcha and Soghutrosh Gypsies, in Uzbek Abdoltili, and the Magati Gypsies. The co-occurrence of 'eye, light' and 'blind', the paradox of which suits an intentionally cryptic argot, is explained from Jewish usage: Aram. *nəhōrā* is 'light, eyesight'; and *saggī nəhōrā* 'hearing much light' is a euphemism for 'blind'. Close in meaning to the Aramaic are Pers. نهور *nohūr*, Taj. *nuhŭr* 'eyesight, eye'.

(10) *KS* شیدا ‹šyd'› 'insane' (Pers. دیوانه *dīvāne* 'insane'). Pers. *šeydā* 'crazed, infatuated, impassioned' is well attested (Ferdousī, Daqīqī, Farrokhī, Asadī et al.). From Aram. *šēdā* 'demon', cf. Pers. *dīvāne* < دیو *dīv* 'demon'. The

Hebrew form, *šēd*, occurs throughout Jewish languages (including Yiddish *šed* 'demon'). Note Judeo-Isfahani *šezim* 'demons' and most relevantly, JLot. Shiraz *šedd* 'to catch disease'.

(11) *KS* دخ <dx> 'good' (Pers. نیک *nīk*) paired with زیف <zyf> 'bad' (Pers. بد 'bad'). A poem by Sūzanī of Samarkand (twelfth century) contrasts "Lūtrā" <dx> 'fair' and <zyf> 'vile, ugly'. Throughout Iranian and Central Asiatic Gypsy and mendicant argots *dax* (Kavoli *dak*) means 'good', and in Persian dervish, Uzbek Abdoltili, Tajik Jugigi, Chistoni, and Arabcha argots *dax* also means 'clean, pure, right, correct', the latter proceeding semantically from Aram. *dəxē/daxyā* 'clear, pure, ritually correct'. Remarkably, Arāk Gypsy argot distinguishes *dax* 'good, right' from daxyā 'pure'. *KS* <zyf>, etc., goes back to Aram. *zayiφ* (*zēφ-*) 'ugly'.

(12) *KS* هادور <hʿdwr> 'a job, some work' (Pers. کار *kār*). Tenth-century *Qaṣīda Sāsāniyya* هادور/هاذور <hʿdwr/hʾδwr> 'the circle of [fortune-tellers and their shills in a street crowd] around which people congregate'. In Gypsy argots of Iran, *xodur*, and of Central Asia, *hodur* means 'beggar', and in Magati argot ādur is 'peddling', the chief occupation of the Sheikh Momadi Afghan itinerants. Pers. هادوری *hādūrī* (Sanāʿī) 'member of a society of persistent beggars'. Aram. *hādōr* 'circle', *hādōrā* 'peddler', root *h-d-r* 'to go around'.

(13) *KS* هار <hʾr> 'feces' (Pers. گو *gū*). Pers. *hār* 'id.' (Sanāʿī). Central Asiatic Gypsy argots *hor* 'id.' From Aram. *hārē* 'feces'.

(14) *KS* *ساوته <*sʾwth> (ms. ساونه <sʾwnh>) 'old (person)' (Pers. پیر *pīr*). Arabcha argot *sout*, Jugigi argot *sowut*, Iranian Gypsy argot *sobut* 'old man'. Asadī, *Loγat al-Fors*, defines صابوته <sʾbwth> as an old woman who has reached the age of seventy; cf. *Borhān-e Qāṭeʾ* <sʾbwth> 'old woman'. In form <*sʾwth> stands to *sout*, *sowut* as <sʾbwth> stands to *sobut*. The alleged reference of the latter alone to a woman may only be explained from the context of the poem by Qarīʿ al-Dahr, which Asadī cites (see 2.5). The Persian spelling with *ṣād* indicates oral transmission, in which the word was treated as though Arabic. *Borhān-e Qāṭeʾ* simplifies Asadī's definition as 'old woman'; the attribution to زند واست <znd wʾst>, as though to Zand and Avesta, probably proceeds from استا <ʾstʾ>, a misreading of Asadī's آسیان *Āsīān*. The argot forms are from a cross of Aram. *sēβūtā* 'old age' and *sāβā* '(old) man'. For abstract used as adjective, cf. Heb. *gālūt* '(the Jewish) exile' > JLot. Golpayegan *gālut* 'miserable', Djougi "*galout*" 'bad', Sam. Lʸuli argot *gohlut* 'ill', etc.; *KS* کالوت <kʾlwt> with contextually similar meaning in a verse. Further Heb. *hăβalūt* (abstract of *heβel* 'vanity' > JLot. Borujerd, etc., *hevalut* 'bad').

(15) *KS* دنه \<dnh\> 'woman' (Pers. زن *zan*). Asadī, *Loγat al-Fors*, defines \<dnh\> as 'noun for woman in the language of the *Āsīān* (*nām-e zan be zabān-e Āsīān*), and gives a verse for \<dnh\> from Qarī' al-Dahr, with an argot phrase \<dnh\>-*ue zīf* 'an ugly woman' (see above on *zīf*). Cf. Persian musicians *duneh*, etc., Kurdish Lūter-e Jāberī *dānu*, Pers. dervishes *danew*, *deneb*; Pers. Gypsies *denew;* Aboltili *danap*; Tajik Gypsy argots *danap*, *danam*, *danawak*; Persian musicians metathetic *nedew*, *nidu*, etc. The argots of itinerants point to early derivation from Arab. *ðanab* 'tail', with phonetic change usual in comparison with the conservatism usual in the *KS* for other forms from Arabic. For the argotic semantics 'tail' > 'woman', cf. the Tajik Gypsy forms in the next item. I now view the similarity of the forms *nVdV* to the Hebrew word for *niddāʰ* 'menstruant' as coincidental. Dr. Ali Ashoury tells me (orally) that in Laki, *dānu* is used for *bānu* (< Pers. *bānu*) 'lady, woman, wife'. I take this as a borrowing from Lūter-e Jāberi *dānu*, whose vocalism would reflect influence of the Persian word.

(16) *KS* هره \<hrh\> 'rump, buttocks' (Pers. کون *kūn*). Abū Dulaf's tenth-century CE *Qaṣīda Sāsāniyya* has *hurr* 'rump', cf. Aram. *ḥor* 'behind, posterior', whereas the later argots of Iran and Central Asia agree as to final vowel with *KS* \<-h\>. Pers. has هُرّه *horre(h)* 'rump, buttocks'. In Tajik Gypsy argots, *hurra* comes to mean 'vulva'; thus "Samarkand Lʸuli" *hurra*, Jugigi *γurra*, and Chistoni *ura* 'vulva' (cf. for **h-* "Sam. Lʸuli" *muhǔz*, Jugigi *muγǔz*, *KS* ماهوز \<mʾhwz\> 'city' < Aram. *māḥōzā* 'id.'; "Sam. Lʸuli" *hal*, Jugigi *γal* 'a piece'; Chistoni Tajik *ar* < *har* 'every').

(17) *KS* ناز \<nʾz\> 'boy' (Pers. γolām 'boy') is found in the midst of a series with Persian gloss دختر *doxtar* 'daughter, girl' toward the beginning and \<sʾwnh\> for *\<sʾwth\> (Pers. *pīr* '[old] man'[14] above) toward the end, which position supports the meaning 'boy' as per the gloss *γolām*. However, emending ناز to *تاز *tāz* yields the attested word for 'catamite' (see 2.3 below), a meaning also common for Pers. *γolām*. As for the position of **tāz* with its gloss *γolām* indicated above, note that the early Persian lexico-graphic tradition, alongside the usual explanation of *tāz* as 'catamite, passive homosexual', also gives for *tāz* the meaning 'beardless young man'. The nuance 'catamite' is made likely by the fact that *KS*, Part Four, ends with a difficult argot verse, most likely scurrilous, featuring indefinite *tāz-ī* and definite *tāz*. The etymology is unclear (cf. Pers. *tāze* 'fresh' or less likely *tāz* *'galloping'). I now fully doubt my earlier speculation that *KS*, Part One ناز is a miscopying of *نار **nār* < Heb. *naʿar* 'boy'.

2.3 Classical Persian poets and transmission of argot

It is clear that it was particularly satirists (Qarīʿ al-Dahr, Sūzanī, Sanāʿī) who were cited in early Persian dictionaries as sources of argot words in scurrilous contexts. Qarīʿ al-Dahr (or Qarīʿ al-Fors), late fourth and early fifth centuries, cited by Asadī, *Loγat al-Fors*, is especially relevant. In one verse we have the collocation of <dnh> 'woman' and <zyf> 'ugly' in <dnh-y zyf> 'an ugly woman'. Furthermore, *ṣābūte*, *horre*, and *tāz* are collocated in the following verse:

مرا که سال به شش و هفتاد رسید و رمید
دلم ز شُلّهٔ صابوته و ز هُرّهٔ تاز
marā ke sāl be šeš o haftād rasīd o ramīd
delam ze šolle-ye ṣābūte (v)o ze horre-ye tāz
[My years having reached seventy and six
My heart shuns an old (one's) *con* and a catamite's *cul*]

The context explains Asadī et al. having *ṣābūte* mean 'a septuagenarian woman'; rather, noun 'old person' or adjective 'old'. (I thank Mahmoud Omidsalar for helpful discussion of the foregoing verse.)

2.4 Fossilized nouns with possessive suffixes as parallel to aramaeograms

In the group of 2.2, *KS* هره <hrh>, Pers. هُرّه *horre(h)* vs. Abū Dulaf *hurr*, Aram. *ḥor*, exemplifies a type of diversity within the history of argot nouns, which is inherited from JLot., and constitutes a noteworthy oral parallel to the variable fixation of stereotyped forms with possessive suffixes in logograms of the Aramaic-originated Middle Iranian scripts. In the instance of <hrh>, etc., the termination goes back to an Aramaic possessive suffixation of *-eh* 'his' as is found vestigially in JLot. Khomein *ragle* 'foot' < Aram. *ragleh* 'his foot'. In the word for 'behind' > 'rump', the vestigial *-eh* merged with the reflex of Middle Persian *-ag*, i.e., Early New Persian *-a*, whereby *-a* in the Tajik Gypsy argot forms. Another *KS* word for a body part, لکته <lkth> 'finger' (Pers. انگشت *angošt*) probably reflects a JLot. form based on an Aram. *-eh* possessive of a slang noun from root *l-q-ṭ* 'to pick up'. The last word is mentioned (from Ivanow 1922) by G. Morgenstierne (1973:151) as "probably" belonging to a series of New Iranian words for 'finger' (Nushki Balochi *aŋgul*, etc.) which reflect **anguli-*. Of these "Persian Gypsy" *lekik* is best taken with the Aramaic-originated *KS* argot word, while Makrani Balochi *laŋkúk* is a diminutive of a metathesis of *aŋgul*, and Kumzari *linkit* a dissimilation of a form like the Makrani Balochi.

KS دكنى <dkny> 'beard' (Pers. *rīš*) would stand to Persian Gypsy *dagnā* 'area of the beard, mouth, etc.' as <hrh> is to Abū Dulaf *hurr*, but represented respectively Aram. *dagnī* 'my beard' and *dagnā* '(the) beard'. Alternation reflecting fossilized nouns *'my N' alongside 'his N', from underlying Hebrew words, is shown by *KS* كيمولو <kymwlw> 'camel' (Pers. اشتر *oštor*) = */gimōlō/ < Heb. *gəmallō* 'his camel' vs. JLot. Golpayegan *gamelli* 'camel' < Heb. *gəmallī* 'my camel'. Aram./Heb. *āβ(ī-)* is reflected with fossilized possessive 'thy' in *KS* ابيك <'byk> 'father' (Pers. پدر *pedar*).

Aram./Heb. *āβ(ī-)* 'father' is reflected with fossilized possessive 'thy', *-x(ā)*, in *KS* <'byk> 'father' (Pers. پدر *pedar*), JLot. Shiraz *abeq* 'id.' and 'my', *-ī*, in JLot. Khomein *ābi* 'id.'. This situation is remarkably parallel to that found for local variation in Middle Iranian Armaeograms, e.g., 'son': Pahlavi *BRH* in Persia < Aram. 'his son', Parthian and Sogdian *BRY* < Aram. 'my son'. Also remarkable, again in the realm of body parts, is the aforementioned JLot. Khomein *ragle* 'foot' < Aram. *ragleh* 'his foot' as parallel to Pahl. *LGLH* 'foot'.

2.5 The argots, East Iranian, and etymology

We may now examine argot words with an East Iranian perspective that are etymologically interesting.

(18) East Iranian forms are expectable a priori in consideration of the fact that among the cities for which *KS* provides argot names there are toponyms referring to Central Asia, and that the Arab Banū Sāsān poet Abū Dulaf had a career in the Central Asiatic court of the Samanids. The twelfth-century poet al-Ḥilli uses the term كنت <knt> for 'town'; the same word in the *KS*, glossed as Pers. شاره <š'rh> = *šahre* 'town' and compares with Sogd. *kanθ* 'city, town', etc. *KS* شكره <škrh> (with *k = g*) 'cat' (Pers. گربه *gorbe*) and Jugigi argot *šaγur* 'cat' respectively match Persian and Pamiric words for 'porcupine, hedgehog' cognate with Avestan *sukurəna-* (? < *suk-wṛHna-* 'having prickly wool' with root-stem *suk-* 'pricking, piercing', to Av. *sūkā*, Pers. *sūzan* 'needle'), the *KS* and Jugigi forms sharing the same semantic argotic change.

(19) Given the foregoing correlations, one may readily connect *KS* argot اكوچيدن <'kwčydn> 'to take' (Pers. گرفتن *gereftan*) with Tajik Gypsy Chistoni argot *yakučidan* 'to take, to buy' under **akōč-*; cf. Sogd. *ākōč* 'to hang up', further Sogd. *ptkōč* 'to catch fish' and Pamir cognates involving entanglement and being caught (Cheung 2007: 249; I thank N. Sims-Williams for online

discussion). Chistoni -*a*- cannot be from *-*ā*-, which would give *-*o*-; *KS* <kwč-> and Chistoni *yakuč*- would be reconciled by *akōč- with unstressed *ă- from *ākôč-. The *y* of Chistoni *yakuč*- is explainable from an imperative *bi-akuč- > *bi-yakuš*-; cf. Oranskij (1983: 175), where *yakuč* is listed, and, independently, Chistoni *yors*- 'to pass by' is reconciled with Jugigi, etc., argot *vars*-, *wars*- 'id.' via a suggested **bi-w/vors*- > *bi-yors- > *yors*-.

(20) Chistoni argot *nằr*- 'to make, to do' (Oranskij 1983: 169, 173) is relevant for the etymon of Proto-Indo-European h_2ner- 'male', for which Cheung (2007: 182–183) denies that there is evidence for a PIE verbal root behind this noun. Clearly Latin *neriōsus* 'strong' and Old Irish *nertum* are denominal, as may be Vedic *sūnara*- 'powerful, potent', although these presuppose 'strong, powerful'. While OPers. /hūnara-/ 'ability, skill' is semantically comparable with Parachi *nar*- 'to be able', nothing indicates that the latter is denominal, especially since a verbal root *(H)nar* seems to be reflected elsewhere in Iranian, in meanings, moreover, that cannot derive from 'male', and in some instances even from 'be strong'. A PIE primary verbal root emerges from suggested diachronic ordering of the data with a view to semantic development, as illuminated by words with parallel developments:

Ossetic *nærs*- 'to swell up, become fat', and *nard* 'well fed' would provide the semantic starting point; cf. PIE **t(e)uH* 'to swell' > 'be powerful, be able'. PIE *h_2ner- 'male, strong one' would simply be a root-stem. From an Iranian verb *nar*- 'to be strong, virile', Ossetic has *nart* 'heroism', most easily derived from *narθra-, whose deverbal suffix *-*θra*- presupposes a verb root *nar*. From 'be powerful' as 'have power over' > 'come to possess' (cf. Iranian root *xšay* 'rule, possess, be able' and further Gr. *krátos* 'power, rule', *kratéō* 'I hold'), one may account, with different preverbs, for Balochi *gīnār*- 'to hold, take possession of', Gabri *afnūrdan*, Yazdi *pe-nar-t* 'to take' and (via 'take hold of') Oss. *avnal*- 'to touch'. In view of the semantically similar PIE *magʰ (*meh_2gʰ) 'be able', whence Eng. *might* (verb and noun), *may*. The potential idea seen in the verbs *might* and *may* accounts for the meaning of the Lithuanian cognate *magù magéti* 'to want, to like', which parallels Lith. *nóras* 'desire'. The Greek cognate of the Germanic and Slavic words, μῆχος 'means, expedient', gets us beyond Parachi *nar*- 'to be able' (whose non-denominative development from 'be strong, potent' should now be self-evident) to Chistoni argot *nār*- 'to make', for which cf. semantically Gr. μηχάνομαι *mēkhánomai* 'devise, construct, bring about'.

   In Chistoni argot *nār*- we likely have local Central Asiatic preservation of an archaism, comparable to the preservation of early Iranian words in Jewish

jargon. In the instance of *nār-*, the preservation has a place in the greater history of the PIE root *h₂ner* 'be strong', etc.

Note: Throughout this article I have, for various reasons, transcribed Persian vowels as in Modern Iranian Literary Persian, rather than early Classical Persian. Thus I have, e.g., *horre-ye* rather than *hurra-yi*, *tāz-ī* rather than *tāz-ē*, *rūz* rather than *rōz*, etc.

References

Cheung, Johnny. 2007. *Etymological dictionary of the Iranian verb. Leiden Etymological Dictionary Series*, 2. Leiden & Boston: Brill.

Hosseinzadeh, Ayda, Ali Mehdizadeh, Arman Zargaran & Moham(m)ad M. Zarshenas. 2013. Abkama, the first reported antibiotic on gastritis and infection throughout history. *Pharmaceutical Historian*, August. 39–42. http://www.researchgate.net/publication/258999430 (accessed 2 January 2017).

Ivanow, W[ladimir]. 1922. An old Gypsy-Darwish jargon. *Journal (and Proceedings) of the Asiatic Society of Bengal* NS/10. 375–383.

Morgenstierne, Georg. 1973. Notes on Balochi etymology. In Georg Morgenstierne, *Irano-Dardica*, 148–165. Wiesbaden: L. Reichert.

Oranskij, Iosif Mixailovič. 1983. *Tadžikojazyčnie ètnografičeskie gruppy Gissarskoj doliny srednaja Azij. Ètnolingvistíčeskoe issledovanie*. Moscow: Akademija Nauk SSSR.

Schwartz, Martin. 2014. Loterāi: Jewish jargon, Muslim argot. In Houman M. Sarshar (ed.), *The Jews of Iran – The history, religion and culture of a community in the Islamic world*, 33–57. London, New York: I. B. Tauris. Revised from 2012 article. http://www.iranicaonline.org/articles/loterai (accessed 2 January 2017).

Sobranie vostochnyx rukopisej Akademii Nauk Uzbekskoj SSR. Catalogue of Oriental Manuscripts of the Uzbek Akademy of Sciences of the Uzbek SSR, Vol. I, 196–197. Tashkent 1952, 196–197.

Troickaja, Anna Leonidovna. 1948. Abdoltili – cexa artistov i muzykantov srednej Azij. *Sovetskaja Vostokovedenie* 5. 251–264.

Yarshater, Ehsan. 1977. THe hybrid language of the Jewish community of Persia. *JAOS* 97(1). 1–7. Cf. Yarshater, Ehsan, *Encyclopaedia Iranica* XV, Fasc. 2, pp. 156–160.

Agnès Lenepveu-Hotz

5 Specialization of an ancient object marker in the New Persian of the fifteenth century[1]

Abstract: In Early New Persian (tenth to eleventh centuries), there are two object markers: a postposition *rā* and a circumposition *mar...rā*, which appear to be equivalent. In some texts of the fifteenth century, *mar...rā* still exists, even if its use is sporadic. Nevertheless, the presence of this ancient object marker alongside the usual marker *rā* poses a question. Based on the occurrences found in a text written in 1484 and conserved in an autograph manuscript, this article aims to analyze and clarify the function of this circumposition in comparison with the postposition. We will see that adding the former preposition *mar* allows the author to avoid ambiguity in some uses of indirect object when *rā* tends to change its marking from an indirect and direct object to simply a direct object. The circumposition is also used to express other specific values, namely, external possession and focalization on an indirect object.

Keywords: New Persian, circumposition *mar...rā*, object marking, direct/indirect object, external possession, focalization.

1 Introduction

In the earliest Persian prose texts (tenth to eleventh centuries), authors used either the postposition *rā* or the circumposition *mar...rā* to mark the direct object (DO) or the indirect object (IO). Bossong (1985: 59) states that the morpheme *mar* generally fell out of use after the twelfth century, although, as we will see, it can still be found in some fifteenth-century texts. To my knowledge,

---

[1] This article is based on a presentation I made at the Third International Conference on Iranian Linguistics, held in Paris on 11–13 September 2009. Since then, I have focused on the circumposition *mar...rā* in two ways: the difficulty of detecting its use, a problem linked to that of the transmission of Persian manuscripts (Lenepveu-Hotz 2014), and its employ in poetry, with the example of Firdausī's *Shāhnāma* (Lenepveu-Hotz 2016).

Agnès Lenepveu-Hotz, University of Strasbourg

no grammar study has specifically addressed this circumposition. Salemann and Shukovski (1925: 27–28) and Jensen (1931: 44) present *mar…rā* as an equivalent to the simple postposition *rā*. For Phillott (1919: 57), "it is generally redundant but occasionally restricts the meaning to the case in point". Moreover, its use cannot be compared to that of Middle Persian, since *mar* did not exist in that stage of the language as a preposition. In Middle Persian *mar* was a noun, meaning 'number, account' (for this etymology, see Benveniste 1938: 459–460), and it had not yet been grammaticalized.[2]

In this article, I will analyze *mar…rā* and its use in a work of the fifteenth century: the *Rauzat al-ahbāb fī siyar al-nabī wa-'l-āl wa-'l-ashāb* (RA), composed by Amīr Jamāl al-Husaynī al-Daštakī al-Šīrāzī in 1484. It was transmitted through an autograph manuscript dating from 1497–1498 and preserved in the library Āstān-e Qods-e Razavi of Mashhad. This text details the life of the Prophet Muhammad and other prophets in a simple narrative prose. For these reasons it is relevant for a linguistic study.[3]

Throughout the one hundred pages I studied in RA,[4] there are twelve occurrences of *mar…rā*, showing a sporadic use of the marker. Yet, these occurrences are numerous enough to justify some analysis, especially as it does not seem to be a stylistic effect or a purposefully archaistic feature. It is important to clarify that the small number of occurrences does not imply that the morpheme has no linguistic value. For example, in French, the conditional used to express the future in the past does not appear frequently: there are only ten occurrences in one hundred pages from Honoré de Balzac's *Le père Goriot* (1835), for instance. No one interprets this to mean that this function of the conditional does not exist in French. The occurrences of the circumposition will be analyzed using the same perspective, that is to say that a small number of occurrences does not undermine the fact that *mar…rā* is used with a specific value.

The presence of the circumposition *mar…rā* in the fifteenth century is puzzling because it had been on the verge of disappearance. First, a comparison of RA with other texts from the same period will allow a view on whether the

---

2 For the prepositions and postpositions in Middle Persian, see Durkin-Meisterernst (2014: 298–359). Nevertheless, *mar* appears as a preposition in two occurrences on a Middle Persian papyrus dated from the seventh century and preserved in Berlin (Benveniste 1938: 459–460); it is used before the name of a day, implying a count in the calendar.

3 Occurrences in other works written in the fifteenth century will be used to confirm the analyses, but one must keep in mind that these other works could have been subjected to correction by the scribes (more likely a deletion than an addition), especially for this type of dialectal features (for this problem, see Lenepveu-Hotz 2014). Therefore, the main occurrences will be taken from RA.

4 These correspond to the first fifty pages (1a–26a) and the last fifty (298b–324b) of the 648-page manuscript.

morpheme *mar* was used in all of them (Section 2). Secondly, I shall examine the postposition *rā* on its own in RA (Section 3) in order to analyze the uses of the circumposition in comparison with the simple postposition. This will lead to an understanding of why the author sometimes chooses *mar...rā* instead of the more common *rā*. It will appear that sometimes *mar...rā* is used as an IO in ambiguous cases (Section 4), while in others it expresses external possession (Section 5) and focalizes on some IO (Section 6).

2 Dialectal specificity

To begin, I will discuss whether *mar...rā* occurs in all texts from this period or is a dialectal feature. Focusing on works from the fifteenth century, it is worth noting that *mar...rā* appears in some texts, but not in others. For the earliest Persian prose texts, Lazard (1963: 382–384) shows that *mar* was a dialectal feature of Transoxiana and eastern Afghanistan, and was rarely employed in texts written in Herat and in Khorasan. It thus can be presumed that this was also the case in the fifteenth century.

Let us look at several works from the fifteenth century, focusing primarily on the latter half:

- ‘Abd al-Razzāq Samarqandī (1413–1482): *Matla‘-i Sa‘dayn*.
- al-Daštakī al-Šīrāzī: *Rauzat al-Ahbāb* (RA), 1484.
- Davānī: *‘Arz-i sipāh-i Uzun Hasan*, 1476.
- Hāfiz-i Abrū: *Panj risāla-i tārīxī*, 1414–1436.
- Jāmī: *Bahāristān* (Ba), 1487.
- Mīr Xʷānd (1433–1498): *Tārīx-i rauzat al-safā* (RS).
- Xunjī: *Tārīx-i ‘Ālam-ārā-i Amīnī* (TA), 1490.
- Xunjī: *Mihmān-nāma-i Buxārā* (MB), 1508–1509.

All of the texts that feature the circumposition *mar...rā* were written in Herat[5] (Table 1): in the one hundred pages that I selected from each of the above texts[6] I extracted twelve occurrences in RA, six in Jāmī's *Bahāristān*, four in the *Tārīx-i rauzat al-safā* of Mīr Xʷānd, and three in the *Mihmān-nāma-i Buxārā* of Xunjī. According to Xʷānd-Amīr, Al-Husaynī al-Daštakī, the author of RA, lived in

---

5 There are many texts written in Herat from the fifteenth century because of the importance of the court of the Timurid Sultān Husayn Bāyqarā.

6 The selected pages are noted under "Corpus" at the end of the article. In the sections to follow, statements about a certain text do not refer to the whole text but only to the pages indicated.

Herat, taught in the Madrasa-i Sultānīya of this city, and preached in the Masjid-i jāmi' for several years.[7] It is well known that the name Jāmī is associated with Herat as well, and Mīr Xʷānd also spent most of his life in this city, under the protection and patronage of the vizier Mīr 'Alī Šīr Navā'ī. The fact that there is no occurrence of *mar...rā* in the *Matla'-i Sa'dayn* by 'Abd al-Razzāq Samarqandī, who was born in Herat, does not undermine this hypothesis. The author traveled frequently outside of the Herat area, particularly around India, and was eventually sent to an embassy in Gilan. Furthermore, if an author of Herat does not use this circumposition, this does not imply that it is not a dialectal feature of the area because an author could decide that a dialectal feature might hamper the universal success he hopes his work will achieve.

Indeed, there is no occurrence of *mar...rā* in the three other texts written by the authors who lived in other parts of Iran. Although Hāfiz-i Abrū was born in Herat, he was educated in Hamadan, traveled often, and died in Zanjan. As for Davānī, he spent all his life in Fars: born in Davan, near Kazarun, he worked as *qadi* of Fars, as a professor in a madrasa of Shiraz, and died in Fars. Xunjī also first lived in western Iran. He himself said that he was Xunjī by lineage, Šīrāzī "by birth and origin", and Isfahānī by residence. However, in his *Tārīx-i 'Ālam-ārā-i Amīnī*, there is one occurrence of *mar*, but it occurs only in one manuscript and this variant was not kept by the editor, who preferred the variant of the other manuscript: *har*, which suited the context better (*va har muslim rā* ... 'and for each Muslim ...', 88). Moreover, the initial variants of the letters *mim* (م) and *hā* (ه) can appear almost identical in old manuscripts.[8] Regarding the *Mihmān-nāma-i Buxārā*, Xunjī began writing it in Bukhara and finished it in Herat (see Storey 1927–1977, vol. 1: 372), and this book presents several occurrences of *mar*.

Table 1: *mar...rā* in some texts from the fifteenth century

Herat					Outside Herat		
al-Daštakī	Jāmī	Mīr Xʷānd	'Abd al-Razzāq	Xunjī (MB)	Hāfiz-i Abrū	Davānī	Xunjī (TA)
12	6	4	0	3	0	0	0 (1)

Consequently, there is evidence suggesting that the circumposition *mar...rā* is a dialectal feature of Herat, and possibly of other regions of the eastern Iranian domain, but it seems that *mar...rā* does not appear farther west.

---

7 For more information about the author, see Newman (1994).

8 See also Lenepveu-Hotz (2014) about the difficulty locating *mar* in manuscripts.

If in the fifteenth century the circumposition had actually disappeared in other regions where it had previously been used (further study would verify this point concerning earlier centuries), the movement of *mar…rā* from Transoxiana and eastern Afghanistan to Herat may be linked to the transfer of the Timurid court from Transoxiana to Herat during the fifteenth century. However, this hypothesis would need to be confirmed: when Timurids appeared (fourteenth century), was the circumposition still employed in Transoxiana, and only in this area? If the displacement of *mar…rā* to Herat occurred prior to the arrival of the Timurids in Transoxiana, it must have been linked to another factor.

To understand why this dialectal feature still appears in the fifteenth century, we will now turn to the use of *rā* in RA, i.e., to the question of whether *rā* is more of a marker of IO like in Early New Persian, or whether it has progressively become a marker of DO as in contemporary Persian. As we will see (Section 4), the result may have consequences on the interpretation of the use of the circumposition *mar…rā*.

3 Use of *rā* in the *Rauzat al-Ahbāb*

In Table 2, *rā* has become a marker of DO: 49.1 percent of the occurrences with a simple verb mark a DO. Adding the occurrences of the object of a complex verb, the result is 80.2 percent. Simple and complex verbs are given separately because the instances of compound verbs can also be analyzed as simple verbs with a real object and an IO marked by *rā*, as Key (2008: 240) and Paul (2008: 335) demonstrate. However, one could assume that at this period the complex verbs are lexicalized, so the third column of Table 2 presents them together.

Table 2: The use of *rā* in the first twenty-five pages of RA

DO	Object of a complex verb	Total of DO	IO	Other uses
109	69	178	39	5
49.1%	31.1%	80.2%	17.6%	2.2%

These figures confirm the general evolution of *rā*, and RA appears as an intermediate stage between the tenth and twelfth centuries (51.2 percent DO / 45.5 percent IO) and the twentieth century (82.3 percent DO / 6.3 percent IO) (percentages taken from Lazard 1970: 384).

In the earliest Persian prose works, the circumposition *mar…rā* could be used for the different functions of *rā*, either DO or IO, apparently without modifying the meaning in comparison with *rā* alone.[9] There is a DO in (1) and an IO in (2), both from the same text, the *Hidāyat al-mutaʻallimīn* (HM), written in 980:

(1) **mar** *zafān* **rā** *bi-jumbān-ad*[10]
     mar   tongue   *rā*   VAFF-move.PRS-3SG
     'It moves the tongue'.
     (HM 61, 7–8)

(2) **mar** *har* *šaxs-ē*   **rā** *mazāj-ē*   *buv-ad*
     mar   each person-INDF *rā*   constitution-INDF be.PRS-3SG
     'For everyone there is a constitution'.
     (HM 19, 14)

Contrary to the earliest state of New Persian, *mar…rā* always marks the IO in the occurrences found in RA as well as in the other works of the fifteenth century (Ba, RS, and MB). This is one of the reasons why I do not agree with the nuance of respect that Bahār ([1942] 1994: 1: 401) interpreted in *mar*. Why would an author want to mark respect merely with an IO? Moreover, there are sentences in which one expects a show of respect (with words like "God" or "the Prophet") and *rā* alone is found; see examples (3) and (4) below. There are also occurrences with a commonplace word used with *mar…rā*.[11] In addition, there are other conclusions to be drawn from this restriction to IO. First, the fact that the circumposition is not used for DO in the fifteenth century as contrary to its use for this purpose in the earliest texts may imply that it specialized in particular functions. Secondly, the circumposition cannot be analyzed as an equivalent to the postposition alone: if this was the case, why would it have been employed only to mark the IO and not, like the postposition, to mark both IO and DO?

The fact that in the fifteenth century *rā* tends to mark the DO more frequently than the IO suggests that the author chooses *mar…rā* when *rā* alone could be interpreted either as a DO (more typical in this state of language) or as an IO (the former primary function of *rā*).

---

9 Cf. Lazard (1963: 382) and Bossong (1985: 59). Regarding the difficulty in understanding its value, see Lenepveu-Hotz (2016).
10 In what follows, language material in Arabic characters is presented in transcription, not in transliteration.
11 See also notes 14 and 29.

4 *mar...rā* as a marker for the IO in ambiguous cases

In RA, *mar...rā* is employed with very different verbs, both simple and complex: once with the verbs *āfarīdan* 'to create', *guftan* 'to say, to tell', *xʷāstan* 'to want', *dilālat kardan* 'to indicate', *vāqa' šudan* 'to happen', twice with *du'ā kardan* 'to pray', and five times with *būdan* 'to be'.[12]

Concerning the first four verbs, the use of *mar...rā* instead of *rā* alone seems meant to avoid an ambiguity in non-prototypical sentences between DO and IO, which are both potential functions of *rā*.[13] The uses of the postposition and of the circumposition depend on the actancy of the verb – one-actant verbs or two-actant verbs (Section 4.1) – and on the animacy (Section 4.2). When the verb is a complex verb (e.g., *dilālat kardan* 'make an indication'), the presence of the nominal element poses another question (Section 4.3). Even if there are few occurrences in RA, it is possible to distinguish the function of *mar...rā* as a marker that allows the disambiguation of some constructions, where *rā* alone could be understood as a DO marker. For each of these four verbs I will first examine the use of *rā* alone in order to demonstrate the role played by *mar* in the occurrences I found in RA.

4.1 *mar...rā* for IO and *rā* for DO

In RA, two verbs illustrate the general theory of disambiguation: *āfarīdan* 'to create' and *xʷāstan* 'to want'. For both, if we compare all the occurrences with and without *mar...rā*, it appears that the author employs the circumposition for IO in occurrences where the postposition alone could have been interpreted as a DO, the main value marked by *rā* alone.

With *āfarīdan*, *rā* is employed only for a DO: to create something or someone, as in (3). However, in examples (4) and (5), the author wants to express both the thing created and for whom it was created.[14] As noted by Paul (2008:

---

12 In the other texts, *mar...rā* is mainly used with the verb *būdan* 'to be': five times in Ba (out of six occurrences of the circumposition), once in RS, and the three occurrences in MB. The other occurrences of RS are used with *dādan* 'to give' and two complex verbs (*sajda kardan* 'to prostrate' and *hāsil šudan* 'to be produced', and in Ba, with *dāštan* 'to have').

13 Meunier and Samvelian (1997: 212) show that *rā* can avoid an ambiguity in contemporary Persian too: "Quant aux groupes nominaux ambigus, elle [la présence de *rā*] en sélectionne une lecture définie et/ou spécifique" [For the ambiguous nominal groups, it (the presence of *rā*) selects a definite and/or specific reading].

14 Again the nuance of respect cannot be seen in the use of *mar...rā*: Adam appears in both (3) and (4), and this poses the question of why he deserves respect in (4) and not in (3).

334, with other examples), "the co-occurrence of a direct and an indirect *rā* is avoided".[15] This seems also to apply to *mar…rā*.

(3) va čūn ādam **rā** biy-āfarīd
   and when PN *rā* VAFF-create.PST.3SG
   'And when he created Adam'
   (RA 319b, 20)

(4) xudāy-i ta'ālā **mar** ō **rā** šahvat āfarīd
   God-EZ Almighty *mar* he *rā* desire create.PST.3SG
   'God, the Almighty, created desire for him'
   (RA 8a, 4)

(5) haqq-i ta'ālā **marā** barāy-i tu āfarīd-a
   truth-EZ Almighty I.*rā* for-EZ you create-PP
   'God, the Almighty, created me for you'.
   (RA 8b, 14–15)

In all the occurrences of the verb "to create" in RA, the DO appears with *rā* if it is animate (eighteen times) and without *rā* if it is inanimate (six times). Paul (2003: 182) and Key (2008: 244) state that in Early New Persian the presence or the absence of *rā* is linked to the animacy of the object. To some extent, this was still the case in the fifteenth century. In (4), one could say that the author may have used *rā* alone because the DO is unmarked. However, because *rā* tends to express a DO with *āfarīdan* in this text, the author may have thought that the sentence would be clearer with *mar…rā*. In (5), the personal pronoun *man* 'I' must be marked by *rā* for it is an animate DO and the IO can only be built with a preposition, e.g., *barāy* 'for', and no longer with *mar…rā*.

Even if there are fewer occurrences of *xʷāstan* 'to want' with an object (three with *rā*, one with *mar…rā*), the use of the circumposition is also a marker of disambiguation, like with *āfarīdan*. In the very similar sentences in examples (6) and (7), *mar…rā* marks the IO *dōst* 'friend'; the DO *sōxtan* 'the burning' is unmarked in the first sentence, whereas the postposition alone marks the DO *dōst* in the second.

---

15 In Early New Persian (tenth to eleventh centuries) – thus a former stage of the language – Lazard (1963: 380–381) finds some examples of such a co-occurrence, but says that in the usual pattern the DO is built with *rā* and the IO with the preposition *ba* 'to', or the DO is unmarked and the IO is marked by *rā*.

(6) *čūn*   *dōst*   **mar**   *dōst*   **rā**   *sōxt-an*   *xʷāh-ad*   *zīst-an*
     when   friend   *mar*   friend   *rā*   burn-INF   want.PRS-3SG   live-INF

     ravā   *nēst*
     allowed   NEG.be.PRS.3SG

     'When a friend wants his friend to burn (lit. 'wants the burning for [his]
     friend'), he shall not deserve to live'.
     (RA 19a, 17–18)

(7) *čūn*   *dōst*   *dōst*   **rā**   *xʷāh-ad*   *sōxt-an*   *ravā*
     when   friend   friend   *rā*   want.PRS-3SG   burn-INF   allowed

     nēst
     NEG.be.PRS.3SG

     'When a friend cares for (his) friend (lit. 'wants the friend'), he shall not
     deserve to burn'.
     (RA 19a, 18)[16]

Two other occurrences of *xʷāstan*, in which the DO is marked only by *rā* (8),
confirm the impression that *rā* is used for a DO while *mar…rā* is used for an IO.

(8) a.   *ibrāhīm*   *vai*   **rā**   *dar*   *vaqt-ē*   *xʷāst*   *ki…*
          PN   he   *rā*   in   time-REL   want.PST.3SG   when
          'Ibrahim asked for (lit. 'wanted') him when … '
          (RA 20a, 2–3)

     b.   *tā*   *tu*   *ō*   **rā**   *bi-xʷāh-ē*
          in order to   you   he   *rā*   VAFF-want.PRS-2SG
          'In order for you to want her'
          (RA 308b, 15)

4.2 The question of animacy

The question of animacy deals with the larger problem of the evolution of object
marking. For *āfarīdan* 'to create' (Section 4.1), we saw that a DO is more often
marked if it is animate. For example, in the *Qābūsnāma*, a text of the eleventh
century, more than 80 percent of the human DO are marked with *rā* (Key 2008:

---

16 The context of (6) and (7) is the following: Nimrod has ordered that Ibrahim should be
burned. Angels ask Ibrahim whether he needs help. In (6) he explains why he did not ask for
help and the angel Jibril answers in (7).

232). However, the DO tends to be marked progressively with *rā* if it is a definite DO, even if it is inanimate. The fifteenth century, and thus RA, which was written in this century, present an intermediary stage between marking dependent on animacy (as in the tenth and eleventh centuries) and marking dependent on definiteness (as in contemporary Persian).

This is why *mar...rā* is employed in non-prototypical sentences with the verb *guftan* 'to say, to tell', again because the postposition *rā* alone could be ambiguous between a DO and an IO. In the unambiguous cases in which *rā* alone marks the IO (seven occurrences), i.e., the person to whom the subject says something, the marked noun or pronoun is always an animate one, such as in (9a).[17] When *rā* marks a DO,[18] it is with an inanimate noun related to speaking, like *kalamāt* 'words', *suxan* 'words or utterances', or *javāb* 'answer', i.e., for the contents of the speech (9b and four other occurrences). In the occurrence with *mar...rā* in (10), the IO *qalb va rūh* 'heart and spirit' also consists of inanimate nouns.[19] Even though without *mar*, *qalb va rūh* would not have been interpreted as the DO of *guftan*, the author prefers to clarify the function of this unusual inanimate indirect object, since the IO is prototypically human.

(9)   a.   *rōz-i   dīgar   kāhin-ān   nimrōd   **rā**   guft-and:*
             day-EZ   other   priest-PL   PN   *rā*   say.PST-3PL
             'The next day the priests said to Nimrod:'
             (RA 16b, 8)

      b.   *īn   kalamāt   **rā**   bi-gōy-ad*
             this   word.PL   *rā*   VAFF-say.PRS-3SG
             'He says these words'
             (RA 10b, 9)

---

17 In classical Persian it was very common for the IO of the verb *guftan* 'to tell' to be built with the postposition *rā*. Cf. Ovčinnikova (1956: 396) and Karimi (1990: 181). In RA, this structure is in competition with *bā* 'with' (twelve occurrences) and *ba* 'to' (four occurrences).
18 I do not discuss occurrences with a DO and an attribute of the object related to the meaning of naming, because in these sentences *mar...rā* is not employed – there is only the structure with *rā* alone. See for instance:

*ān-**rā**   duldul   mē-guft-and*
this-*rā*   PN   VAFF-say.PST-3PL
'They called him Duldul' (RA 320b, 3)

19 See Key (2008: 232): "The classification 'animate' included inanimate objects, body parts, places, and abstractions". Other occurrences show that *qalb* 'heart' and *rūh* 'spirit' are treated as inanimate nouns. Indeed, we must remember that the notion of humanness is scalar (cf. Lazard 1982: 185–186).

(10) *nafs-i   vai   **mar**   qalb   va   rūh   **rā**   guft:*
     soul-EZ   he   *mar*   heart   and   spirit   *rā*   say.PST.3SG
     'His soul said to (his) heart and (his) spirit:'
     (RA 17a, 22)

4.3 IO vs. DO with a complex verb: The case of *dilālat kardan* 'to indicate'

With complex verbs, one can assume that the situation of marking is different. When an object is marked with *rā*, it is difficult to interpret it as a DO or an IO because the nominal element of the complex verb can be interpreted as the first object of the verb. If we consider that in this stage of the language the complex verbs are lexicalized, the marked object can be either the DO of the complex verb or its IO. The uses of the postposition and the circumposition show us the same use of disambiguation as those we saw with simple verbs. The IO of *dilālat kardan* in (11) is marked by *mar...rā* while *rā* alone marks the DO of the verb in (12). This is especially important as the meaning of the verb is different: "indicate something about or related to someone" in the first example and "lead someone" in the second.

(11) *ahādīs-i   sahīha   dilālat   bar   subūt-i   īn   xasīsa*
     hadith.PL-EZ   true.PL   expressing   on   evidence-EZ   this   characteristic
     *mē-kun-ad   **mar**   ān   sarvar   **rā***
     VAFF-do.PRS-3SG   *mar*   that   lord   *rā*
     'The sound hadith indicate the evidence of this characteristic about the Prophet'
     (RA 315b, 16)

(12) *dilālat   kun-am   šumā   **rā***
     guiding   do.PRS-1SG   you   *rā*
     'I guide you'
     (RA 310a, 11)

In fact, *šumā rā* in (12) is on the same level as *subūt* and not on that of *mar ān sarvar rā* in (11).[20] These two sentences can be defined as "make an indication for you" in (12) and "make an indication of the evidence of . . ." in (11). Looking

---

20 So, even if the complex verb must be interpreted as a simple verb and its unmarked object, the distinction of the different level is the same.

more closely at *mar...rā* in (11), one could also analyze this marking in a different way: as the expression of the external possession.

5 Expression of external possession?

In many European languages the dative (or the IO) is also employed to express external possession,[21] as shown by König and Haspelmath (1998). They note examples in German, *Die Mutter wäscht dem Kind die Haare* 'The mother washes the hair of the child', or in French, *On lui a tiré les oreilles* 'He had his ears boxed'. The phenomenon concerns inalienable property: the possessor is animate, specifically human, and the possessed is a part of the body or clothes. The verb cannot be a state verb or a verb of perception but, according to the terminology of König and Haspelmath (1998: 533), it is a "verb of contact or change". As we will see, it is not this exact form of external possession in our text because of the meaning of the verb and because the complement built with *mar* still depends on a noun rather than on the main verb. However, it functions like this structure, i.e., *mar...rā* serves as a substitute for the usual possessive pattern with the *ezāfe* (or with a personal suffix) in some examples. For instance, (11) is almost equivalent to *īn xasīsa-i ān sarvar* 'this characteristic of the Prophet'.

Because *rā* never marks external possession in the one hundred pages I studied in RA, *mar...rā* can also avoid ambiguity. In (13), if *rā* had been employed, the reader could have interpreted the meaning to be that Adam is the recipient of the calling whereas in actuality he is the beneficiary of the prostration. The circumposition clarifies that *mar ādam rā* is linked to *sujūd* 'prostration' and not to the main verb *vāqa' šuda* 'took place' (the context demonstrates that the recipient of the calling is Iblīs). Moreover, (13) does not exhibit the usual order for this type of existential sentence (Indirect Object-Subject-Verb),[22] but the order Subject-Indirect Object-Verb.

---

21 For the term *external possession*, see König and Haspelmath (1998: 526): "le possesseur n'a cependant pas besoin de faire partie du même constituant de phrase que le possédé, mais peut aussi sous certaines conditions former un constituant de phrase distinct" [the possessor does not necessarily need to be part of the same phrase as the possessed but under certain conditions it may also constitute a distinct phrase].

22 For example, in:

*ibrāhīm **rā** mu'āraza bā nīmrōd vaqt-ē vāqa' šud ki...*
PN *rā* quarrel with PN time-REL placed become.PST.3SG when
'A quarrel took place between Ibrahim and Nimrod (lit. 'for Ibrahim with Nimrod') when ...'
(RA 18a, 7)

(13) *xatāb-i   sujūd   **mar** ādam **rā** bā   malāyika vāqa' šud-a*
     calling-EZ prostration *mar* PN   *rā* with angel.PL placed become-PP
     'The calling for prostration before Adam took place with the angels
     (for Iblīs)'
     (RA 7b, 19)

One can observe a similar phenomenon in other sentences.[23] The structure with *būdan* 'to be' and the IO generally marks the possession, like the Latin *mihi est* type. However, in (14), the author compares Muhammad's feet to pumice and does not say that Muhammad has no feet. In this case, we do not find the common expression of possession with *būdan* 'to be', which can be seen in other examples. In fact, with *mar…rā* and *būdan*, none of the five occurrences present the meaning 'to have', contrary to *rā* alone.[24]

(14) *va   **mar** ō **rā** pāy-hā na-būd-a   ast   čunān ki*
     and   *mar* he *rā* foot-PL NEG-be-PP be.PRS.3SG as
     qayqāb rā mē   bāš-ad
     pumice *rā* VAFF be.PRS-3SG
     'And his feet were not like pumice'
     (RA 302b, 10–11)

Unlike external possession, the complement built with *mar…rā* does not function as the beneficiary or recipient of the main verb but maintains its relation to the noun. The possessor is not the experiencer of the action. König and Haspelmath (1998: 567) state that in external possession, "le possesseur est affecté par l'action qui concerne son possédé" [the possessor is affected by the action that concerns its possessed]. However, like external possession, "the possessor is treated as an additional argument of the clause" (Payne and Barshi 1999: 5).

---

23 Also in Ba: *va dar tahniat-i fath-i vai **mar** hindūstān **rā** qasīda-ē dārad* 'and in congratulations for his victory against India there is a panegyric poem' (Ba 127). Even if the phrase *mar hindūstān rā* 'against India' is syntactically independent, it is semantically linked to *fath-i vai* 'his victory'.

24 This is not the case with the occurrences in the other texts, in which *mar…rā* is used with *būdan* 'to be' with the meaning of 'to have'. However, this does not contradict my assertion, because in all of these occurrences, there is a specific use of this structure: all appear in sentences in which the subject or the object follows the verb, giving the impression that there is a value of focalization, as we will see in Section 6 with other examples.

In place of external possession, one could have considered examples (11) and (14) as left dislocation (14) or right dislocation (11), but in (13) the words built with *mar...rā* are placed between the subject and a complement. In addition, in dislocation, the complement is not always in a genitive relation with a noun from the sentence. Nevertheless, one could say that in many examples the placement of *mar...rā* depends on a dislocation that strengthens the use of the circumposition instead of the usual structure with the *ezāfe*.

The complement marked by *mar...rā* cannot be a benefactive IO because of examples such as (13), in which Adam is not the beneficiary of the calling (we are sure of that meaning thanks to the context and the historical background), and examples such as (14), in which *būdan* 'to be' built with an IO does not have its usual possessive meaning. Therefore, it is more relevant to compare this structure with external possession.

Moreover, even if the use of *mar...rā* is linked to the original meaning of *rā* 'with regard to, for the sake of', the pattern may be compared to one of external possession. We find the same use of this structure in the German and French examples quoted above: a dative or IO (*mar ō rā*, *dem Kind*, *lui*) represents the possessor of a noun phrase unmarked by a possessive marker (*pāy-hā*, **die Haare**, **les** oreilles and not *pāy-hā-aš*, **seine** Haare, **ses** oreilles). Therefore, this is not exactly the same pattern as in contemporary Persian, as seen in sentences like *aqā kučik* **rā** *čeqadr xarj-e tahsil-eš kardam* [Aqā Kučik, how much I spent on his studies!] given by Lazard (2006: 180): there is not only the postposition *rā* but also the possessive marker *-eš*. However, compared with the usual structure with the *ezafe*, the use of *mar...rā* to express possession may be a nuance of focalization, like *rā* in contemporary Persian.[25]

6 Focalization

The use of a morpheme as a marker of focalization is quite common in other languages. Pottier (1968: 92) states that we see or do not see the preposition *a* in Spanish according to the speaker's purpose. Such an analysis can be applied to the uses of *mar...rā*. This is most likely the case for the occurrences in which the circumposition is employed to avoid ambiguity: the sentence in (10), for example, may be understood as 'To (his) heart and (his) spirit, his soul said:'. There is further evidence of this role. For instance, in (13), with *mar...rā*, the

---

25 This was most likely the role of *mar* alone in Early New Persian (Lazard 1963: 450–451), and perhaps also of the circumposition (Lenepveu-Hotz 2016).

author can focus on an unexpected prostration before Adam instead of the usual prostration before God. The focalization is also tangible in other examples, like in (15).[26]

(15) *agar*   *sajda*   **mar** *haqq* *rā* *būd-ē,*   *iblīs*
      if   prostration   *mar*   truth   *rā* be.PST.3SG-VAFF   PN

      takabbur   *na-kard-ē*
      pride   NEG-do.PST.3SG-VAFF

      'Had the prostration been for God, Iblīs would not have had a disdainful attitude'
      (RA 7b, 10–11)

Again, the order of the words is unexpected in such an existential sentence and the author stressed the focalization by using *mar…rā*. Indeed, in this context, we understand that Iblīs agrees to prostrate himself only before God (*mar haqq rā*) but not before Adam as God commands the angels to do. This will have a considerable consequence: Iblīs becomes the devil and is sent to hell. God (*haqq*) is then the most important piece of information in the first clause and is focalized by the circumposition, itself strengthened thanks to the placement of the complement.[27]

Regarding *du'ā kardan* 'to pray', even though the occurrences are not numerous enough to draw conclusions (two with *mar…rā* and only one with *rā*), one might conclude that the position of the object again plays a role in its marking. In fact, the two sentences with this verb and *mar…rā* do not contain the same order as the sentence with *rā* alone (16). There may be a marked order with *mar…rā* in (16a) and an unmarked order with the simple postposition in (16b).[28]

---

26 See another example: *va **mar** allāh-i ta'ālā **rā** muvahhid nabūda* 'And towards God, the Almighty, he was not monotheist' (RA 17a, 7).

27 We see the same phenomenon with *vāqa' šudan* 'to happen'; see (13) above: in this type of existential sentence, the order with *rā* is unmarked (Indirect Object-Subject-Verb) and marked with *mar…rā* (Subject-Indirect Object-Verb). An unusual order is also found in two other occurrences of *mar…rā* with *būdan* 'to be', in which the IO appears after the verb: *va ādam qabla būd **mar** ān sajda **rā*** 'And Adam was the direction for that prostration' (RA 7b, 8) and *qasm-i avval ki vājibāt ast hikmat dar taxsīs-i ānhā ba hazrat ziyādatī zulfā va husūl-i darajāt-i 'olā ast **mar** ō **rā*** 'The first part which is the compulsory religious prescriptions is for him the wisdom to devote these things to the Lord, to be much closer and reach the highest degrees (of spirituality)' (RA 313b, 11).

28 The other texts also present an unusual order SVO in the occurrences with *mar…rā*; see note 24 and this example in RS: *čirā sajda namēkunand **mar** xudāy **rā*** 'why they do not prostrate themselves before God' (RS 432).

(16) a. Subject- Verb-   *mar* Object *rā*

     b. Subject- Object *rā*-   Verb

Furthermore, thanks to the study of the context of these three sentences, it is clear that in the two occurrences with *mar...rā* the author focuses on this complement. In (17), he opposes all the dead at the beginning of the prayer to the man who has died, *mar mayyit rā*, at the end of the prayer. In (18) *ō* is focalized because the sentence is aimed at someone who has just killed the man represented by this *ō* 'he'.

(17) *va   dar takbīr-āt-i   dīgar du'ā   kard-ē   **mar***
      and   in   takbir-PL-EZ other prayer do.PST.3SG-VAFF mar

      *mayyit **rā** ba maqfirat   va   rahmat*
      deceased rā   to   absolution and   mercy
      'And in other *takbir*, he would pray for the deceased for (his) absolution and mercy'
      (RA 298b, 13)

(18) *čirā du'ā-i   barikat   na-kard-ē   **mar** ō   **rā***
      why   prayer-EZ   blessing   NEG-do.PST-2SG   mar   he   rā
      'Why did you not pray for his blessing? (lit. 'do the prayer of blessing for him')'
      (RA 313a, 2)

Contrary to (17) and (18), in the sentence with *rā* alone (19), there is no focalization in the context. This concerns any tribe Muhammad eats with and not one tribe in particular. Thus, in this sentence *ān qaum* 'that tribe' is not focalized.

(19) *čūn   nazd-i   qaum-ē   ta'ām xʷard-ē   ān   qaum   **rā***
      when   near-EZ   tribe-INDF   food   eat.PST.3SG-VAFF   that   tribe   rā

      du'ā   kard-ē
      prayer do.PST.3SG-VAFF
      'When he would eat with a tribe, he would pray for that tribe'
      (RA 303b, 17–18)[29]

---

29 This occurrence again contradicts the nuance of respect that *mar* could have expressed: even if the tribe is respected and honored (the Prophet would pray for them all), it is only marked with *rā*.

One may argue that there are also some occurrences in which *rā* alone marks the focalization. Indeed, such occurrences, such as (20), can be found with the postposition *rā* alone, but here it is a focalized subject and no longer an IO.

(20) *ān*   *nūr*   ***rā***   *dar*   *pīšānī-i*   *vai*   *būd*
       that   light   *rā*   in   forehead-EZ   he   be.PST.3SG
       'Regarding that light, it was on his forehead'
       (RA 4b, 11)

Therefore, one can assume that the author focuses on an IO with *mar…rā* and on other functions with *rā*.[30] There is a similar distinction in Wolof or Banda-Linda (Feuillet 2006: 627), where there are two different markers depending on the function of the focalized group: for example, Banda-Linda uses the morpheme *kà* for the subject and the morpheme *dà* for the object. A parallel can also be made with contemporary Persian, since *rā* can focus (or topicalize) on different complements, but never on the subject (Meunier and Samvelian 1997: 212–230).

7 Conclusion

To conclude, although there are few occurrences of *mar…rā* in the texts of the fifteenth century, it is evident that it is a dialectal feature of Herat, always employed with an IO. This circumposition is part of a continuum of the "zone objectale" [object zone] – according to Lazard's expression (1994: 95; translated in 1998: 91) – in which there are different degrees of marking, as seen in (21). It depends on the nature of the object, its degree of animacy, and on its precise function, i.e., a DO or an IO.

(21) unmarked   direct object   indirect object   indirect object
      direct object   marked by *rā*   marked by *rā*   marked by *mar…rā*

      ←——→

As for the focalization, we can only speculate about the phenomenon because of the limited number of occurrences. However, it is a likely hypothesis, as with

---

30 The use of *mar* as emphasis was already noted (see, for example, most recently, Bābāsālār 2013: 184, with examples in poetry; and Moayedi and Lotfi 2013: 112). However, no study has highlighted the difference between the emphasis expressed by the circumposition and the emphasis expressed by the postposition alone.

mar...rā the author can focus on an IO more intensely than when using *rā* alone or an *ezāfe* (in occurrences linked to external possession). The groups with other functions, like the subject, could be focalized with the postposition alone. Therefore, as in other languages, different markers may exist.

Consequently, more than a survival of an archaism, there is a restriction, a specialization of the former morpheme *mar* to distinguish some uses of IO from the more general uses with *rā* alone. This phenomenon took place at a time in which the postposition *rā* tended to mark the DO more and more restrictively. Thus, the circumposition helps avoid ambiguity in non-prototypical sentences. It is not an exact rule that authors strictly observe but it is one way in which Persian has renewed itself in order to express the distinction between direct and indirect object, when the postposition *rā* was in a transitory step between the old marking of indirect object and the new one of direct object.

8 Abbreviations in the glosses

EZ – *ezāfe* (linker between the head noun and its modifiers or the head noun and its possessor); INDF – indefinite; INF – infinitive; NEG – negation; PST – past; PL – plural; PN – proper noun; PP – past participle; PRS – present; REL – introducer of a relative clause; SG – singular, VAFF – verbal affixes (modal or aspectual).

Corpus

‘Abd al-Razzāq Samarqandī, Kamāl al-Dīn. *Matla‘-i Sa‘dayn va majma‘-i bahrayn* [The rising of the two stars of fortune and the confluence of the two seas], 33–133. ‘Abd al-Husayn Nawāī (ed.). Tehran: Mo’assesse-ye motāle‘āt va tahqiqāt-e farhangi, pažuhešgāh, 1993.

al-Axavainī al-Buxārī, Abū Bakr Rabī‘ b. Ahmad. *Kitāb hidāyat al-muta‘allimīn* [The guide book for students of medicine], 13–112. Jalāl Matini (ed.). Mashhad: Čāpxāne-ye dānešgāh-e Mashhad, 1965. Abbreviated HM.

al-Daštakī al- Šīrāzī, Amīr Jamāl al-Dīn ‘Atā’ Allāh b. Fazl Allāh al-Husaynī. *Rauzat al-Ahbāb fī siyar al-Nabī wa-’l-Āl wa-’l-Ashāb* [The garden of the friends in the footsteps of the Prophet, his family and his companions], 1a–26a, 298b–324b. Preserved in the library Āstān-e Qods-e Razavi of Mashhad (Nr. 4109). Abbreviated RA.

Davānī, Jalāl al-Dīn Muhammad. *‘Arz-i sipāh-i Uzun Hasan* [Presentation of Uzun Hasan's army], 41. Iraj Afshār (ed.). Tehran: Dāneškade-ye adabiyāt, 1956.

Hāfiz-i Abrū. *Cinq opuscules de Hāfiz-i Abrū*, 68. Félix Tauer (ed.). Prague: Editions de l'Académie tchécoslovaque des Sciences, 1959.

Jāmī, ‘Abd al-Rahmān b. Ahmad Nūr al-Dīn. *Bahāristān va rasā’il* [The spring garden and epistles], 65–169. ‘Alāxān Afsahzād, Mohammad Jān ‘Omarof, & Abu Bakr Zohur al-Din (eds.). Tehran: Mirās-e maktub, 2000.

Mīr Xʷānd, Muhammad b. Xāvandšāh. *Tārīx-i rauzat al-safā* [The garden of purity], 419–519. Jamshid Kiyānfar (ed.). Tehran: Asātir, 2001.

al-Xunjī al-Isfahānī, Abul-Xayr Fazl Allāh b. Rūzbihān. *Tārīx-i 'Ālam-ārā-i Amīnī* [The world-adorning history of Amini], 21–120. John E. Woods (ed.). London: Royal Asiatic Society, 1992.

—. *Mihmān-nāma-i Buxārā* [The guest-book of Bukhara], 1–50, 306–356. Manučehr Sotudeh (ed.). Tehran: Bongāh-e tarjome va našr-e ketāb, 1962.

References

Bābāsālār, Asghar. 2013. Kārbordhā-ye xāss-e *rā* dar barxi motun-e fārsi [Special uses of *rā* in some Persian texts]. *Adab-e fārsi* 3 (1). 181–196.

Bahār, Mohammad Taqi. 1994 [1942]. *Sabkšenāsi. Tārix-e tatavvor-e nasr-e fārsi* [Stylistics. History of the evolution of Persian prose], 3 vols. Tehran: Amir Kabir.

Benveniste, Emile. 1938. Sur un fragment d'un psautier syro-persan. *Journal asiatique* 230. 458–462.

Bossong, Georg. 1985. *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Gunter Narr.

Durkin-Meisterernst, Desmond. 2014. *Grammatik des Westmitteliranischen (Parthisch und Mittelpersisch)*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Feuillet, Jack. 2006. *Introduction à la typologie linguistique*. Paris: Honoré Champion.

Jensen, Hans. 1931. *Neupersische Grammatik*. Heidelberg: Carl Winter.

Karimi, Simin. 1990. Obliqueness, specificity, and discourse functions: *rā* in Persian. *Linguistic Analysis* 20 (3–4). 139–191.

Key, Gregory. 2008. Differential object marking in a Medieval Persian text. In Simin Karimi, Vida Samiian & Donald Stilo (eds.), *Aspects of Iranian linguistics*, 227–247. Newcastle: Cambridge Scholars Publishing.

König, Ekkehard & Martin Haspelmath. 1998. Les constructions à possesseur externe dans les langues d'Europe. In Jack Feuillet (ed.), *Actance et valence dans les langues de l'Europe*, 525–606. Berlin and New York: Mouton de Gruyter.

Lazard, Gilbert. 1963. *La langue des plus anciens monuments de la prose persane*. Paris: Klincksieck.

Lazard, Gilbert. 1970. Etude quantitative de l'évolution d'un morphème: la postposition *rā* en persan. In David Cohen (ed.), *Mélanges Marcel Cohen*, 381–388. The Hague-Paris: Mouton.

Lazard, Gilbert. 1982. Le morphème *rā* en persan et les relations actancielles. *Bulletin de la Société de Linguistique de Paris* 77 (1). 177–208.

Lazard, Gilbert. 1994. *L'actance*. Paris: Presse Universitaire de France.

Lazard, Gilbert. 1998. *Actancy* [English translation of *L'actance*]. Berlin and New York: Mouton de Gruyter.

Lazard, Gilbert. 2006. *Grammaire du persan contemporain*. Tehran: Institut Français de Recherche en Iran.

Lenepveu-Hotz, Agnès. 2014. Manuscrits persans et linguistique diachronique: la loi du provisoire? De la difficulté à saisir un morphème dialectal, *mar*. In Nalini Balbir and Maria Szuppe (eds.), *Lecteurs et copistes dans les traditions indiennes, iraniennes et centrasiatiques*. *Eurasian Studies* XII (1/2). 101–120.

Lenepveu-Hotz, Agnès. 2016. L'emploi de *mar… rā* chez Firdausī: simple raison métrique ou cause linguistique? In Céline Redard (ed.), *Des contrées avestiques à Mahabad, via Bisotun. Études offertes en hommage à Pierre Lecoq*, 259–276. Neuchâtel and Paris: Recherches et publications.

Meunier, Annie & Pollet Samvelian. 1997. La postposition *rā* en persan. Ses liens avec la détermination et sa fonction discursive. *Les cahiers de grammaire* 22. 187–232.

Moayedi, Mona & Ahmad Reza Lotfi. 2013. Barresi-e sāxt-e do mafʻuli dar motun-e adab-e fārsi [Double object construction in the Persian literary texts]. *Majalle-ye pažuhešhā-ye zabānšenāsi* 5 (1). 101–119.

Newman, Andrew J. 1994. Daštakī, ʻAṭā-Allāh, (d. 1506, 1511, or 1520), a scholar of Hadith in Khorasan in the late Timurid and early Safavid periods. *Encyclopædia Iranica* VII (1). 100. http://www.iranicaonline.org/articles/dastaki (accessed 22 January 2017).

Ovčinnikova, I. K. 1956. Ispol'zovanie posleloga *rā* v proizvedenijax tadžikskix i persidskix klassičeskix avtorov [The use of the postposition *rā* in the works of classical Persian and Tajik authors] (XI–XV vv.). *Trudy Instituta Jazykoznanija* (*Moskva*) 6. 392–408.

Paul, Ludwig. 2003. Early Judaeo-Persian in a historical perspective: The case of the prepositions *be*, *u*, *pa*(*d*), and the suffix *rā*. In Ludwig Paul (ed.), *Persian origins*: *Early Judaeo-Persian and the emergence of New Persian*, 177–194. Wiesbaden: Harrassowitz.

Paul, Ludwig. 2008. Some remarks on the Persian suffix *-rā* as a general and historical linguistic issue. In Simin Karimi, Vida Samiian & Donald Stilo (eds.), *Aspects of Iranian linguistics*, 329–337. Newcastle: Cambridge Scholars Publishing.

Payne, Doris & Immanuel Barshi. 1999. *External possession*. Amsterdam and Philadelphia: John Benjamins.

Phillott, Douglas Craven. 1919. *Higher Persian grammar*. Calcutta: University Press.

Pottier, Bernard. 1968. L'emploi de la préposition *a* devant l'objet en espagnol. *Bulletin de la Société de Linguistique de Paris* 63 (1). 83–95.

Salemann, Carl & Valentin Shukovski. 1925. *Persische grammatik mit literatur, chrestomathie und glossar*. Berlin: Reuther und Reichard.

Storey, Charles. 1927–1977. *Persian literature. A bio-bibliographical survey*, 2 vols. London: Luzac and Co.

Lutz Rzehak

6 Fillers, emphasizers, and other adjuncts in spoken Dari and Pashto

Abstract: This article deals with spoken language or more precisely with sponta-neous speech in colloquial Dari and Pashto. The focus is on stereotyped conver-sation fillers or "filled pauses" that often have been regarded as "disfluencies" since they constitute a delay in the flow of speech. Lexical items and other ele-ments that are used to fill hesitation pauses are syntactically omissible and structurally dispensable. It is argued that many of these elements can be assigned pragmatic functions, such as signaling an upcoming focused word, or the speaker's intention to plan or code his or her speech. The article studies the questions that lexical items, in general, can be used as adjuncts of that kind in colloquial Dari and Pashto, how they are inserted in fluent speech, and which pragmatic functions they fulfill. Furthermore, the article will study the question as to what extent the choice of a particular lexical item to be used as filler follows common patterns in a speech community, and to what extent the choice is based on individual preferences.

Keywords: Dari, Pashto, spoken language, fillers, communication strategies, pragmatics

1 Introduction

The following phrases were said during an interview that was taken by the Fed-eral Office for Migration of Switzerland with an asylum seeker who claimed that she came from Afghanistan. In this part of the conversation the asylum seeker speaks about her migration history.

Lutz Rzehak, Berlin Humboldt-University

(1) Female Hazara from Helmand province, about thirty-two years old
(I = interviewee, R = respondent)[1]

I:	*eh xō awwal bār-e awwal aft sāl bodin da irān bār-e dowwom čand sāl se sāl*	Okay. First, the first time you were seven years in Iran and the second time how many years? Three years?
R:	*se sāl ah se sāl / aft sāl awwal / bād az aft sāl dige raft- se sāl čiz-ē kam se sāl afġānestān zendegi kadēm **masalan** elmand-u kābol / bād az u āmadēm se sāl da irān*	Three years. Yes. Seven years first, then after seven years we went – we lived in Afghanistan for almost three years, for example, in Helmand and Kabul. After that we went to Iran for three years.
I:	*xō besyār xub*	Okay, very well.
R:	***masalan** koll-e az iyā mēša sēzda sāl*	And [Or: Well,] all of it makes thirteen years.
I:	*sēzda sāl*	Thirteen years.

In this part of the conversation the respondent uses the lexical item *masalan* twice. The first time this word is used in its original meaning and function as an adjunct for adducing examples when the speaker mentions some places of residence in Afghanistan. In this case *masalan* can be translated as 'for example', 'for instance'.

For the second usage of *masalan* such a meaning cannot be admitted and a translation as 'for example', 'for instance' would not make any sense. In this case the lexical item *masalan* can be given several meanings or functions. It can be understood as a coordinating or joining conjunction and one can translate it as 'and', correspondingly. But since *masalan* is used here immediately after a response by the interviewer, we may also assume that here *masalan* is used by the interviewed person as a discourse connective for retaking the turn of the conversation. Hence *masalan* can be given the meaning of 'yeah', 'um', 'well', 'but', or something similar in this case. From a syntactical and semantic point of view, the second *masalan* can be removed or discarded without compromising the core idea of the utterance. In written language the lexical item

---

1 Here and in all following cases no punctuation marks are used in the transcription of free speech. A slash </> marks noticeable prosodic breaks that are accompanied by a pause and last approximately up to half a second. Aborted lexical units are marked by a hyphen at the end. In this interview the interviewer uses the colloquial standard of Dari that is based on the old dialect of Kabul. Most linguistic features of the respondent can also be assigned to this variety. For a description of its phonology, morphology, and syntax, see Farhadi (1955).

masalan would hardly ever be used in the way it appears the second time in this part of the conversation. Here its usage, undoubtedly, is a special feature of spoken language or, more precisely, of spontaneous speech as realized in the course of an oral face-to-face conversation.

Face-to-face communication follows special linguistic rules. Structural elements are the opening and closure of conversation, strategies for taking the turn, holding the turn, and yielding the turn of the conversation, procedures for the production of meaning and the assurance of understanding (paraphrase, repair, etc.). Typical traits of spoken language that are instrumental in the organization and contextualization of conversation are short, often incomplete sentences, mixing of sentence structures, and a more frequent use of discourse particles. Other traits are caused by the fact that in spoken language the language production process is much more compact and compressed as compared to written language. The language planning process includes planning the utterance with regard to what to say, retrieving the words and integrating them into a sentence, articulating the sentence, and monitoring the output.[2] Features related to that are, among others, hesitation phenomena, pauses, speech errors, unexpected discontinuities in the expression of ideas, difficulties with word retrieval, and self-repair.

Among these features of spoken communication, this article deals with stereotyped conversation fillers. Often fillers are considered as "filled" or "non-silent pauses" or as "disfluencies" since they constitute a delay in the flow of speech.[3] Fillers can be non-lexical items like "eh", "em", "hmm", as well as lexical items such as *masalan* 'for example', 'for instance' in example (1) given above. It is a general feature of fillers that they are often syntactically omissible and structurally dispensable. This study is based on the idea that often fillers are assigned pragmatic functions, such as signaling an upcoming focused word, or the speaker's intention to plan or code his or her speech. In many cases they are associated with referential meaning or they are used for some additional highlighting. The focus of this article is on lexical fillers. The article studies which lexical elements, in general, can be used as fillers in colloquial Dari and Pashto and how they are inserted in fluent speech. Furthermore, the article will address to what extent the choice of a particular lexical item to be used as filler follows common patterns in a speech community, and to what extent the choice is based on individual preferences.

The main focus of this article is on Dari. For this study 105 sequences in Dari from thirty-one informants were analyzed. Speakers of Dari belong to a multi-

---

2 See Bussmann (2006: 651–652, 1115).

3 For disfluencies in spontaneous speech, see Brennan and Schober (2001: 274–275) and Horne (2006: 265).

lingual society with Pashto as another prevalent language and numerous phenomena of language contact.[4] On these grounds twenty-nine sequences in Pashto from six informants were also included in the analysis for purposes of comparison. I take my material from recordings of colloquial speech that I have made in various parts of Afghanistan during the last twenty years and from recordings of interviews that were taken by the Federal Office for Migration of Switzerland with migrants from Afghanistan. These interviews were made anonymous and kindly provided to me for research.[5] These interviews are structured as dialogues between two native speakers (no interpreter was involved) which allows study of the mentioned phenomena as part of conventional discourse strategies.

2 Fillers in spoken Dari

At a first glance, it becomes evident that most fillers can be considered hesitation pauses that have been "filled" for planning the utterance. Many lexical items are used in that function:

Table 1: Lexical elements that are used as fillers in spoken Dari

Lexical item	Meaning	Number of usage in 105 sequences
ba hesāb	'so to speak' (lit. 'calculated')	10
čiz	'thing'	9
masalan	'for example'	9
če	'what'	7
b-estelāh	'so to speak', 'quasi'	7
maqsad	'what I mean' (lit. 'purpose', 'intention')	5
xō	'yet', 'but'; 'after all'	4
xolāsa	'in summary', 'briefly'	3
ba ġoul-e ma'ruf	'as the word is', 'like they say'	3
ami	'that'	2
kam-i	'a bit'	2
baxšeš begōyam	'I apologize for saying'	2
diga	'else'	1
mesl-e	'like', 'as'	1
če begom	'what shall I you say'	1
rāst-eš	'in reality', 'in fact'	1

---

4 On Pashto-Dari bilingualism in Afghanistan and contact-induced language phenomena in these languages, see Kiselova (1982) and Rzehak (2012). On the theory of language contact studies, see Riehl (2009) and Thomason and Kaufman (1988: 35–64).

5 I thank the Federal Office for Migration of Switzerland for allowing me to use these interviews for this study.

2.1 Filled hesitation pauses

The most commonly used filler is *ba hesāb* (lit. 'calculated', 'counted') that in colloquial language is often used in the meaning of 'quasi', 'so to speak', 'in a manner of speaking', 'in a way', or 'sort of'. It can also be used with *ezāfa* in the form of *ba hesāb-e*. However, when inserted into spontaneous speech as filler, none of the mentioned meanings, obviously, seems to be appropriate. Mostly it is used for filling a hesitation pause that was inserted for planning the following parts of the utterance.

(2)   Male Tajik from the city of Herat, thirty-nine years old[6]

*mā amu manteye-yi ke **be hesāb** zendegi mikadim kōče-ye dudālun **be hesāb** bein-e e pāasār-u **be hesāb** čārsu… unğā hamām-i bud hamām-e azizi… bād?? **be hesāb-e** be rū be rū az ū **be hesāb-e** čiz bud kūče-i bud dāxel-e az ū kūča mā bodim*	That neighborhood where we so to speak lived [was] the Dudalun-street, in a manner of speaking between Pa-ye Asar and Charsu, before Charsu.… There was a bath-house, the Azizi bathhouse.… And so to speak across from it was quasi what, there was a street. We were in this street.

The second most common lexical item used for filling hesitation pauses is *čiz* 'thing'. This lexical item can be considered a temporary *placeholder*, i.e., it must be replaced by something else when the utterance is completed. Usually the speaker has an idea about the syntactic structure of the utterance, but he or she is not yet certain about a particular word or expression. This unclear word or expression is temporarily replaced by *čiz* until the necessary word comes to his or her mind. In the following examples *čiz* is translated as 'thing' or 'what'.

---

6 The sign /ɣ/ stands for a voiceless uvular fricative that is typical of the dialect of Herat. It corresponds to the voiceless uvular plosive /q/ and voiced uvular fricative /ġ/ of the colloquial standard of Dari. It is more voiced than /q/ but less voiced than /ġ/. The actual pronunciation of /ɣ/ can vary between a half-voiced uvular plosive and a voiced fricative, depending on the position within a word. See Ioannesyan (1999: 32).

(3)   Male Tajik from the city of Herat, thirty-three years old[7]

I:	*šomā xod-etān arōsi kadin*	Are you married?
R:	*mā **čiz** kadom eh nekāh kadom diga waxt ba arōsi našod moškel dištom āmadom birun*	I made the thing, the marriage ceremony but no time was left for the wedding. I had problems and I went abroad.

In this sequence *čiz* stands for *nekāh* 'marriage ceremony'. In example (4) it stands for *momtāz* 'talented'.

(4)   Male Hazara from Behsud province, fifteen years old[8]

I:	*rāğe ba ami maktab če yād-et ast če mitānin bogōyin az daure-ye maktab*	What do remember about that school, what can you tell about the school days?
R:	*darsā-ye mō ke pāyin bod az u diga kelāsā az senfā-ye bālātar yak nafar ke hamintur **čiz**-e kelās bod momtāz bod u-ra rawān mǔkadand da kelās-e az mō bar-e mō dars mǔgoft*	In my lower classes a person from the higher classes who was the thing, [the most] talented [of his class] was sent to our class and he gave classes to us.

When we say that *čiz* is a placeholder, this does not necessarily mean that the syntactic structure must be completed in exactly the same way it was planned originally. In example (5) the usage of *čiz* shows that the speaker originally had intended to create another structure of his utterance, maybe with an adjective like *saxtgir* 'strong', but he changed his mind and gave a periphrastic description of his father's behavior.

---

7 In this sequence the past stem of *dāštan* 'have' appears as *dišt*, which is also a characteristic feature of the dialect of Herat (my observation). Furthermore, in colloquial Dari for reasons of modesty, the personal pronoun for the first person singular *ma(n)* can be replaced by the pronoun for the first person plural *mā* (*pluralis modestiae*), in these cases the verb can also occur in the form of the first person plural or in the form of the first person singular as in example (3).

8 Example (4) was given in the dialect of the Hazaras (Hazāragi), where /ō/ corresponds to /ā/ of the colloquial standard. The sign /ǔ/ denotes a long and stable back vowel between /u/ and /ō/, which is characteristic of most varieties of Hazāragi. It is more closed than /ō/ and more open than /u/. It appears mainly in the verbal prefix of many verbs (cf. *mi-/mē-* in standard Dari). See Efimov (1965: 12).

(5)   Male Qizilbash from the city of Ghazni, eighteen years old

> I: *xub diga da xāna kodām kumak mikadi kār mikadi čiz mikadi*
>
> Well, did you help at home; did you work there or something like that?

> R: *ma xō xord-e xāna bodim xord-e xāna mēgan sag-e xāna as har če ke mēgoftan mēkadim ma ham xub padar-am bāz **čiz** bod namimānd ke ziyād birun berom*
>
> After all I was the youngest in the house. And people say that the youngest in a house is like the dog of the house. I did everything what they told me. Well, and my father was what, he didn't allow me to go outside often.

In a similar function the interrogative *če* 'what' or sometimes the rhetoric question *če mēgan* 'how it is said?' can be used as placeholder.

(6)   Male Dari speaking Pashtun from Paghman province, nineteen years old

> I: *xod-e maktab-etān da koǧā bod*
>
> Where was your school?

> R: *maktab-e mā da **če** kōte-ye sangi amu taraf-e **če** / **če mēgan** da pēš-e dēwānbēgi*
>
> My school was what, in Kote-ye Sangi in the direction of what, how it is called, in front of Dewanbegi.

The literal meaning of a filling expression often falls aside. This can be seen by the fact that a particular expression may be pronounced in a strongly reduced form, like, for example, *b-eslā* (> *ba estelāh* 'quasi', 'so to speak') in example (7).

(7)   Hazara from Jaghori district of Ghazni province, over forty years old

> I: *zabān-e mādari-yetān če ast*
>
> What is your mother tongue?

> R: *zabān-e mādari-ye mā ami zabān-e azāragi **b-eslā** dari mitānim begōyim*
>
> My mother tongue is Hazaragi, so to speak, we can say Dari.

2.2 Fillers as connectors

In some cases it is difficult to define exactly in which function a particular expression is used as filler. Some lexical items are inserted not only for filling hesitation pauses but for connecting words or phrases or for expressing the meaning relationship between one utterance and the following one. In these

functions, according to my data, the lexical items *xolāsa* 'in summary', 'briefly', *b-estelāh* 'quasi', 'so to speak', *maqsad* 'what I mean', or *masalan* 'for example', 'for instance' can be used. Thus in example (8) *xolāsa* can be considered not only a filled pause. From the point of view of semantics, it combines two clauses, but it could be displaced without doing serious harm to the idea of the utterance.

(8)   Male Hazara from Qarabagh province, sixteen years old

> I: *sabzimaidān če qesm ğāy ast*   What kind of place is
>   Sabzimaidan?
>
> R: *sabzimaidān az xod-e darwāze-ye*   It is one or two stations above the
>   *tehrān yak du istgāh bālātar miša*   Teheran gate. What I mean, there
>   **xolāsa** *unğā ke hast harruza*   bazaar is every day.
>   *bāzār a*

2.3 Fillers used for highlighting

Example (8) indicates another important pragmatic function that fillers can fulfill. They can be inserted for highlighting a particular element of an utterance. In example (8) the speaker might have said *unğā harruza bāzār a* 'There bazaar is every day'. But he didn't say that. Instead he said *unğā **ke hast** harruza bāzār a*. The commonly used expressions *ke hast* 'which is' (present tense) or *ke bod* 'which was' (past tense) are often inserted for highlighting a preceding element of an utterance. It could be translated as 'as for ...' or 'as far as ... is concerned', but such a translation is not compulsory as examples (9) and (10) show.

(9)   Male Hazara from Jaghori district of Ghazni province, age unknown[9]

> I: *bāz nazr-e bibi goftin da u če as*   Then you mentioned [the custom]
>   *nazr-e bibi*. What is [in] that?
>
> R: *nazr-e bibi **ke asta** amčunān čiz a*   As for Nazr-e bibi, this is a similar
>   *maqsad yak nazr mǔkona amu-ra*   thing. What I mean: People give
>   *amu nazr-a xānā ke nazr kada bāz*   offerings and these offerings –
>   *ğam mǔša mǔxōra amu rōz xoš tēr*   those houses that have given
>   *mǔna diga*   offerings gather and they spend
>   that day well.

---

9 Notice the use of *maqsad* 'what I mean' as connector as described above.

(10) Male Hazara from Jaghori province of Ghazni province, age unknown[10]

I: *sāl-e nau goftin sāl-e nau če mikonan*	You mentioned the New Year. What do people do at New Year?
R: *sāl-e nau masalan amu id amu sāl-e nau **ke asta** diga ke nauruz mŭša nauruz mŭgōyi diga amu-rā nauruz-rā migirand*	New year, for example, this festival, what concerns the new year, it becomes Nauruz, people call it Nauruz, and people celebrate this Nauruz.

2.4 Fillers as expression of uncertainty

Some lexical items that are inserted for filling hesitation pauses are clear symptoms of a speaker's uncertainty and of his indecisiveness about what to say and how to say it. In this function a great variety of lexical items can be used, and some of them one would hardly expect to be fillers like, for example, *dobāra* 'again' in example (11).

(11) Male Hazara from Jaghori district of Ghazni province, sixteen years old[11]

*yā ṭrēkṭar kerāya mŭkadi yā masalan če gāw yā xar-mar kerāya mŭkadi balde če-ši balde u-ra ke **dobāra** čiz mŭkadi xarmanikōbi mŭkadi u-ra **dobāra** da bād midādi i čizā*	Either one rented a tractor or, for example, one rented a cow or a donkey or so for what, for again doing, what, for threshing at the barn-floor and for what, winnowing and so.

2.5 Fillers as discourse markers

Often fillers are used as discourse markers in order to address the interlocutor directly and to maintain the conversation. These are often expressions of the Persian *ta'ārof* (Persian form of civility emphasizing defense and social rank) or rhetoric questions.

---

10 Notice that in the dialect of the Hazaras the verb form of the second person singular can be used for expressing impersonal forms, i.e., *mŭgōyi* stands here for 'people say', 'one says'.
11 In this sequence the retroflex /ṭ/, the preposition *balde* 'for' (cf. standard *barāye*) and the enclitic pronoun for the third person singular and plural *-ši* (cf. standard *-aš, -eš or -ašān, -ešān*) must be mentioned as special features of the dialect of the Hazaras.

(12)   Male Hazara from Jaghori district of Ghazni province, twenty-one years old

I:	*če kešt-u kār mēša unjā*	What is cultivated there?
R:	*unjā masalan amin gandom masalan kešt mikona diga **baxšeš begōyim** kačālu kešt mikonan piyāz kešt mikonan dehāt ai diga*	There, for example, that wheat, for example, is cultivated, else, I apologize for saying, potatoes are cultivated, onion is cultivated; after all it is a rural place.

(13)   Male Hazara from Jaghori district of Ghazni province, over forty years old

I:	*kodām mazāmin-a mixāndin da maktab*	Which subjects did you learn at school?
R:	*da maktab-e uskōl da maktab-e uskōl xō mā mazāmin-e moxtalef as wale mā az dari dāštim riyāzi dāštim **xedmat-e šomā arz konom** ke pašto dāštim qorān-e šarif dāštim*	In the school? In the school we had different subjects. We had Dari, we had mathematics – and I make so bold as to say you – we had Pashto, we had the Holy Koran.

(14)   Male Tajik from Qarabagh province, seventeen years old

*mādar-am har birun ke miraft kati čādar miraft az xāter-i ke **če mēgand če raqam bar-etān begom** az xāter-i ke polis u iyā šak nakona sar-e mā*	Whenever my mother went outside she wore a veil in order to – how it is said, how shall I say this to you – so that the police and these people wouldn't have any doubt on us.

The rhetorical question *fāmidi / pāmidi* 'Got that?' is also often inserted into fluent speech as a discourse marker and for filling a hesitation pause. An Arab speaker, whom I interviewed in Mazar-e Sharif, used it twenty times within a speech sequence with a length of twenty minutes. When a speaker asks in every second sentence *fāmidi* 'Got that?' as this Arab did, he does not really expect an answer. This qualifies this rhetorical question as filler. An answer to this rhetorical question is neither expected nor given. The question is inserted to address the interlocutor and to produce a pause for planning the rest of the utterance. It is worthy of mention that the lexical item *fāmidi* 'Got that?' was used with a similar frequency by other Arab speakers whom I met in Balkh province. This qualifies this filler as a special feature of the Persian dialect of the Arabs of Balkh province.

(15) Male Dari-speaking Arab from Balkh province, fifty years old

ma az kalime-ye xān bad mibarom / ***fāmidi*** */ hič waxt ma ami andēwālā-ye xod-a xān namēgom hič*	I don't like the word Khan. Got that? I never call my friends Khan.

(16) Male Dari-speaking Arab from Balkh province, fifty years old

bače-ye gōsfand-a bara mēgim / bače-ye boz-a bozġāla / boz-e nar-a taka mēgim / ***fāmidi*** */ boz-e nar-a taka mēgim / boz-e nār-ē ke mā xāyē-š-a kašida bāšim u-ra ke da u diga nari nakona u-ra bāz serka mēgim /* ***fāmidi*** */ boz nar as walē xāyē-š-a kašida ke diga i diga mardi nadāšta bāša / una serka*	We call the young of a sheep *bara*, the young of a goat *bozġāla*. We call a male goat *taka*. Got that? We call a male goat *taka*. A male goat the testicles of which were extracted so that it would not cry anymore, we call it *serka*. Got that? It is a male goat but its testicles were extracted so that he would not have manhood. This is a *serka*.

In this respect religious formula must also be mentioned. The expression *wallāh* 'I swear by God!' is often used at the very beginning of an utterance. It can be regarded as an expression that is inserted for retaking the turn of the conversation.

(17) Tajik from the district of Soltan-Saheb of Ghazni province, twenty-four years old

I: *da afġānestān az kodām qesmat-eš hastin*	Which part do you come from in Afghanistan?
R: ***wallāh*** *mā az ġazni-stim woloswāli-ye soltān sāheb*	I swear by God! I am from Ghazni, from the district Soltan-Saheb.

It is due to the semantics of this formula that *wallāh* 'I swear by God!' is often used when negative answers are given. In the culture of Afghanistan negative answers are regarded as something bad and people try to avoid such answers. The religious formula, obviously, shall underline the truthfulness of a statement that cannot be avoided.

(18) Male Hazara from Jaghori district of Ghazni province, age unknown

I: *aw dāštin*	Did you have water?
R: ***wallāh*** *ziyād aw nadāštim čun xošksāli šoda*	I swear by God! We didn't have much water because there was a drought.

(19) Tajik from the city of Ghazni, twenty-four years old

    I: *diga zabānhā-rā ham balad hastin*   Do you understand other
        masalan englisi bāša paštō bāša   languages, for example, English
                                                  or Pashto?

    R: **wallāh** *nē paštō-rā yād nadārom*   I swear by God! No, I do not know
        englisi-rā ham yād nadārom farz-e   Pashto, I do not know English
        mesāl fārsi-rā mēfāmom   either, but I know, for instance,
                                                  Farsi.

2.6 Filling longer hesitation pauses

If we admit that hesitation pauses are inserted for planning the rest of an utterance, it is a consequential assumption that longer pauses provide more time for planning. In order to fill longer hesitation pauses, filling words that are semantically close to each other can be combined. Typical combinations are *taqriban hodud-e* 'approximately about', *masalan farz-e mesāl* 'for example, for instance' or *wallāh hodud-e* 'I swear by God! About'.

(20) Male Qizilbash from the city of Ghazni, eighteen years old

    I: *wa i maktab če qesm bod maktab-e*   What was this school like? Was
        majāni bod yā pul mipardāxtin   the school free of charge or did
                                                  you pay?

    R: *maktab xō maktab-e afġāniyā bod*   The school, after all the school
        amu nafar bod ke **masalan mesāl**   was a school for Afghans. There
        diplōm gerefta bod az irān az   was a person, for example, for
        maktab-e irāniyā ba mā dars   instance, who got his diploma
        midād   from Iran, from an Iranian school.
                                                  He taught us.

Example (21) is of special interest because three fillers are combined with each other.

(21) Male Tajik from Ghazni province, twenty-four years old

    I: *če qadr waqt mēša ke āmadin*   How long has it been that you
                                                     arrived?

    R: **wallāh odud-e ya'ni** *az xāne-ye*   I swear by God! About, I mean,
        mā ke arakat kadim yak panj šaš   since I have started from my home
        māh mišawa   it is about five or six months.

2.7 Iranian influence

Some Afghans who spent some years in Iran have a special sympathy for the Persian language of Iran. Consequently they may also use some expressions for filling hesitation pauses that are not common in colloquial Dari but are in colloquial Persian of Iran. In example (22) this is the expression *ba ġoul-e ma'ruf* 'as the word is', 'like they say'. According to Dari phonology, this expression should be pronounced *ba qaul-e ma'ruf* (with the consonant /q/ instead of /ġ/ and with the diphthong /au/ instead of /ou/), but even in this way of pronunciation this expression is not common in Dari. It is rather popular in Iran and must be considered an Iranian influence on the linguistic behavior of these speakers.

(22)   Afghan Hazara, who was born in Iran and went to school there for five years, twenty-three years old

I: *kār mikadin če qadr barāyetān midādan*

You worked. How much did they pay you?

R: *pul bastagi dāšt diga **ba ġoul-e ma'ruf** u zamān masalan ke mā kār můkadim bača bodom u zamān xeili kam bod masalan ruz-i panǧsad tuman kār mikadom xub da kārxāna ba'd az in ke **ba ġoul-e ma'ruf** bozorg šodim xō bištar kār mikadom ruz se hazār tuman čār hazār tuman*

The money depended like they say. At that time when I was working I was a child. That time the money was very few, maybe 500 tomans per day. Well, in the factory after, like they say, I had grown up, I worked more, per day 3,000 tomans or 4,000 tomans.

3 Fillers in spoken Pashto

Dari and Pashto are very different languages from an historical-genetic point of view. Dari is a west-Iranian language with a rather analytical structure, while Pashto is an east-Iranian language with a synthetic structure. However, both languages coexist in the multilingual society of Afghanistan and many contact-induced linguistic phenomena are apparent. It is no surprise, therefore, that Dari and Pashto to a large extent share a common vocabulary in the field of fillers.

Most Pashto fillers are used the same way as in Dari. Most of them are omissible from the point of view of semantics and grammar. The Dari filler *ba hesāb* 'so to speak' has the Pashto equivalent *pəhisāb* that is used similarly. The speaker in example (23) is bilingual. Pashto is his first language, but he also

Table 2: Lexical items used as fillers in Pashto with Dari equivalents

Pashto	Dari equivalents	Meaning
pə hisāb, hisāb-kitāb	*ba hesāb*	'so to speak' (lit. 'calculated')
misāl	*mesāl*	'for instance', 'for example'
no	*diga*	'then', 'so', 'hence'
če dəi / da	*če (h)asta*	'what is'
poh šwei	*fāmidi*	'Got it?'
xo	*xō*	'yet', 'but'; 'after all'
wallāh		'I swear by God!'

knows Dari on the level of a second native language. This may have influenced his lexical choice for filling hesitation pauses in Pashto.

(23) Pashtun from Kunduz province, thirty-five years old

I: *tāse yāstəi də afğānistān də kom qismat na*

Which part do you come from in Afghanistan?

R: *zə pə asl ki də xānābād də ahtāğ yəma pə qaum niyāzai yəma **pə hisāb** wilāyat-e kunduz rādʑi də xānābād munga **pə hisāb** də ahtāğ yāstu də kunduz*

I am by origin from Khanabad, from Ahtaj, by tribe I am Niyazai, so to say, this belongs to Kunduz province, from Khanabad, I am, so to say, from Ahtaj, from Kunduz.

(24) Pashto-speaking Pashtun from Khost province, seventeen years old

I: *musāferi tsənga teregi*

How is your journey going on?

R: *musāferi **xo wallāh** ḍer pə muškilāt bānde teregi*

My journey is after all – I swear by God! – with many difficulties.

A Pashto equivalent of the Dari expression *fāmidi* 'Got that?' can be seen in example (25). This speaker is also bilingual, but his knowledge of Dari as a second language has hardly influenced his lexical choice in Pashto because no Arabs can be found in the province of Kunduz where this speaker comes from. The expressions *poh še* 'Got it?' is used by numerous speakers of Pashto from various places and it can be considered an original Pashto filling expression here:

(25) Pashtun from Kunduz province, fifteen years old

> I: *hamdəlta paidā šəwai ye hamdəlta*
> *loy šəway ye hamdəlta de žwand*
> *kəṛai*

Were you born there and did you grow up there? Did you live there?

> R: *ho dəlta paidā šəm pə čārdaro*
> *ki paidā šəwai yəm **poh še** də*
> *kunduz čārdaro ki u aməlta paidā*
> *šəwai yəma **poh še***

Yes, I was born there. I was born in Chardara. Got that? In Chardara from Kunduz. There I was born. Got that?

The Pashto equivalents of the Dari expression *ke asta* 'which is', used for highlighting, are *če dəi* (masculine) or *če da* (feminine). As in Dari, they are inserted for highlighting a preceding element of an utterance. They must not necessarily be direct copies of Dari, although the speaker in example (26) comes from a district where Dari is widely used as first language even by many Pashtuns. But the expression *če dəi* (masculine) or *če da* (feminine) can also be encountered in the speech of numerous Pashtuns from other places with no knowledge of Dari.

(26) Pashtun from Surkhrod district of Nangarhar province, about twenty-four years old

> I: *də tāso kəli ki ka na kəli pə bāra*
> *tsə mung ta mālumāt rākawe tsə*
> *tārif kawe*

As far as your village is concerned, what do you say about your village, what can you describe?

> R: *emunga kəlai **če dəi** kalāgāni di /*
> *har saṛai xpəla kālā lari*

As for our village, there are fortresses. Everyone has his fortress.

A filler that can be considered an exclusive feature of Pashto against Dari is the expression *ka na* 'isn't it?', 'isn't that so?' which usually is placed after the word or expression it refers to. It can be considered a rhetorical question inserted for highlighting.

(27) Male Pashto speaking Pashtun from Kunduz province, fifteen years old

> I: *yau dāktar dəi*

Is there one doctor?

> R: *dāktarān ləg xo na di **ka na** ziyād*
> *di **ka na***

Doctors are not few, are they? There are many, aren't they?

(28)   Male Pashtun from Nangarhar province, seventeen years old[12]

  I: *tsə karəl kegi tāso ʤāy ki*   What is cultivated at your place?

  R: *har tsə kegi xo biyā ziyātar xalk*   There grows everything, but
       məxki tso kāluna məxki taryāk tsə   previously, some years ago, most
       *kāwə **ka na** taryāk ziyāt kedə*   people produced what, much
       opium, didn't they, opium was
       much cultivated.

The popularity of the filling expression *ka na* 'isn't it?', 'isn't that so?' can be seen in the fact that both the interviewer and the interviewee use it in the following part of the same interview.

(29)   Male Pashtun from Nangarhar province, seventeen years old

  I: *də kom kəli yāstəi*   Which village do you come from?

  R: *šerzād markixēl*   Sherzad Markikhel.

  I: *au markixēl če dəi **ka na** dā cə šai*   And as for Markikhel? What is
       dəi   that?

  R: *markixēl ləka če šērzād yau ʤāy*   Markikhel was a place like
       *wu **ka na** misāl xabara yau qariya*   Sherzad, didn't it? It is so to speak
       da   a village.

The fact that *ka na* is used not only for highlighting but also for filling a hesitation pause in order to plan the utterance becomes evident in example (30) in which *ka na* is added to the conjunction *ʤəka* 'what' is really unusual.

(30)   Male Pashtun from Paktia, sixteen years old

       *həlta xo ṭilifun nəšta ʤəka **ka na***   There are no telephones because,
       həlta dāse hisāb da če žranda kār   evidently, the mills (for electricity)
       nakawi   do not work.

4 Summary

The analysis of 105 speech sequences in Dari from thirty-one speakers and twenty-nine sequences in Pashto from six speakers brought to light a great variety of lexical items and expressions that in one way or another are used as

---

12 Notice that in example (28) the interrogative pronoun *tsə* 'what' is used as a placeholder in a similar way as *če* in Dari.

fillers or emphasizers. Fillers in Dari and Pashto show many similarities due to the close neighborhood of the speech communities in the multilingual society of Afghanistan and mutual linguistic influences.

Generally speaking, all speakers of Dari and Pashto can resort to these pools of lexical items and expressions when inserting fillers in their speech, but some preferences are evident. Some lexical items or expressions proved to be the individual preference of a particular speaker. This does not mean that other speakers would not use these lexical items and expressions at all, but these particular speakers used them strikingly more often than other speakers.

The expression *ba hesāb* 'so to speak' was most often used by a thirty-nine-year-old Tajik from Herat. The expression *b-estelāh* 'so to speak', 'quasi' was used by an over forty-year-old Hazara from Jaghori much more often than by other speakers. The lexical item *xolāsa* 'in summary', 'briefly' was preferred over other lexical items by a sixteen-year-old Hazara from Qarabagh. The expression *ba ġoul-e ma'rūf* 'as is known' proved to be an exclusive feature of the language of Afghans who lived in Iran for a while and who show special sympathy to the Persian language of Iran. The rhetorical question *fāmidi* 'Got it?' was used by speakers who belong to the group of Persian-speaking Arabs in northern Afghanistan, but by no other speaker. It can be considered a dialect feature of the Persian dialect of the Arabs in northern Afghanistan. In Pashto, most popular are constructions with *ka na* 'isn't it?' and with *če dəi / če da* 'as for ...', 'as far as ... is concerned', which are used for highlighting a particular element of an utterance. Its structural and semantic Dari equivalent *če (h)ast* 'which is' is similarly popular in colloquial Dari. The interrogatives *če* (Dari) and *tsə* (Pashto) 'what' are likewise used as placeholders.

The idea that fillers can be considered filled pauses for planning an utterance is not wrong. But the analysis shows that at the same time the described lexical elements can fulfill various functions from the point of view of pragmatics and discourse strategies. They can be used as discourse markers, as connectors, for highlighting an element of an utterance, or for covering up uncertainty and indecisiveness of a speaker regarding what to say and how to say something.

References

Brennan, Susan E. & Michael F. Schober. 2001. How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language* 44 (2). 274–296.

Bussmann, Hadumod. 2006. *Routledge dictionary of language and linguistics.* Translated and edited by Gregory Trauth and Kerstin Kazzazi. London & New York: Routledge.

Efimov, Valentin Aleksandrovič. 1965. *Yazyk afganskix xazara: yakaulangskiĭ dialect* [The language of the Afghan Hazaras: The dialect of Yakaulang]. Moskva: Nauka.

Farhadi, Abd-ul-Ghafûr. 1955. *Le Persan parlé en Afghanistan. Grammaire du Kâboli*. Paris: Centre National de la Recherche Scientifique.

Horne, Merle. 2006. The filler EH in Swedish. *Lund University, Centre for Languages & Literature, Dept. of Linguistics & Phonetics, Working Papers* 52. 65–68.

Ioannesyan, Julij Arkad'evič. 1999. *Geratskiĭ dialect yazyka dari sovremennogo Afganistana* [The dialect of Herat of the Dari language of modern Afghanistan]. Moskva: Vostochnaya literature RAN.

Kiseleva, Lidiya Nikolaevna. 1982. Dvuyazyčie pašto-dari v Afganistane [Pashto-Dari bilingualism in Afghanistan]. *Narody azii i afriki* 6. 94–99.

Riehl, Claudia Maria. 2009. *Sprachkontaktforschung. Eine Einführung*. Tübingen: Gunter Narr Verlag.

Rzehak, Lutz. 2012. How to name universities? Or: Is there any linguistic problem in Afghanistan? *ORIENT. Deutsche Zeitschrift für Politik, Wirtschaft und Kultur des Orients* 53 (II). 84–90.

Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization and genetic linguistics*. Berkeley, Los Angeles, & London: University of California Press.

Youli Ioannesyan

7 The historically unmotivated *majhul* vowel as a significant areal dialectological feature

Abstract: The article focuses on the *majhul e* /ê/ in modern Tajiki and most of the Afghan Persian (Dari) dialects, which occurs as result of the historically short /i/ changing into /ê/ in a closed syllable when followed by pharyngeal consonants /h/ or /'/. Because /ê/ in such cases does not go back to historically long /ē/ and is fully determined by environmental factors, I define it as "historically unmotivated". However, since this phenomenon does not occur in Persian varieties, including the Khorasani Persian dialects of western Afghanistan, it should be regarded as an important distinguishing feature marking, along with other characteristics, the linguistic border between the Persian dialects of Iran (in a broader sense), on the one hand, and the Tajiki and Afghan-Persian varieties, on the other. The article is based on various published (mostly Russian) studies on Tajiki, Afghan Persian, and Iranian Persian and on the field materials I collected in Afghanistan.

Keywords: Persian, Afghan Persian and Tajiki dialectology, Iranian linguistics, Iranian studies, classification of Persian dialects

The three closely related languages – Iranian Persian, Afghan Persian (Dari), and Tajiki – form a vast continuum of varieties, stretching from western Iran to Afghanistan and Central Asia (Tajikistan, partly Uzbekistan). Because it is not easy to draw a geographical border between the dialects of Persian proper, those of Afghan Persian, and those of Tajiki based on purely linguistic factors, as these varieties overlap and merge into one another, it is therefore reasonable to conceive of this whole area as a single linguistic continuum within which three major groups can be defined, namely, western (western and central Iran), central (north-eastern Iran and north-western Afghanistan), and eastern (central and northern Afghanistan, Tajikistan, and parts of Uzbekistan).[1]

What is implied by *majhul* vowels in Iranian linguistics is long /ē/ and /ō/ (with possible variations) going back to classical Persian. This article deals with the former vowel with regard to the abovementioned linguistic continuum. This

---

[1] The central group consists of Khorasani (type) dialects. For more information on these three groups, see my articles: Ioannesyan (1995, 2007).

Youli Ioannesyan, Institute of Oriental Manuscripts of the Russian Academy Sciences

historically long (*majhul*) /ē/ has become /i/ in modern Iranian Persian,[2] but is preserved in Afghan Persian (and Tajiki) as a long /ē/: دير *dēr* 'late'. In this article, I use bold for historic phonemes (in classical Persian). Slant brackets / / indicate Iranian/Afghan Persian and Tajiki phonemes, square brackets [ ] enclose particular phonetic realizations (allophones) of Iranian/Afghan Persian and Tajiki. It is also worth noting that /e/ in Tajiki ("e" in the Cyrillic-based original Tajiki script) always stands for *majhul* /ē/ (IPA /eː/). For the Herati dialect, I also need to distinguish between the close mid /e/ and open mid /ɛ/ and between these two and /ə/. The latter symbol in a final position implies a variant of æ and stands for a near open front unrounded vowel, but slightly reduced and, for this reason, with a less clear articulation.

One of the most remarkable features of Tajiki, distinguishing it from Iranian Persian, is the changing of the historic short and historic *ma'ruf* long vowels into *majhul* vowels in a closed syllable when followed by glottopharyngeal consonants [h] or [ʔ]. This phenomenon as a universal characteristic of Tajiki merits special attention. According to V. S. Rastorguyeva (1964: 29), the changing of /i/ (which in Tajiki combines respective short and long vowels) into [e] (phoneme /ē/) is most common, for it is typical of the literary language and of all the local varieties. As will be shown later, this shift also occurs in a vast majority of Afghan Persian dialects.

This article focuses on the changing of historic short and long ĭ, ī into *majhul* /ē/. Because the latter in such cases does not go back to historic long ē and is fully determined by environmental factors, I define it as "historically unmotivated". However, since this phenomenon does not occur in Persian varieties, including the Khorasani Persian dialects of western Afghanistan, it should be regarded as an important distinguishing feature marking, along with other characteristics, the linguistic border between the Persian dialects of Iran (in a broader sense), on the one hand, and the Tajiki and Afghan Persian varieties, on the other hand (or between the western and central groups of dialects, on the one hand, and the eastern, on the other hand).

Let us now consider a few examples from Tajiki dialects disregarding some possible minor qualitative differences in the articulation of [e] (/ē/) between certain varieties of this language[3]:

---

2 With the exception of some dialects.

3 While in some Tajiki dialects this vowel has a very close variant that sounds halfway between [i] and [e], it may be realized as a diphthongoid in some others (Sokolova et al. 1952: 164; Sokolova 1949: 10). This led researchers engaged in field studies to transcribe it differently. The existence of various phonetic forms does not, however, obscure the general picture, for all the variants of *majhul* /ē/, regardless of qualitative differences, represent the same phoneme, which occupies a distinct place in the vowel system of the respective dialect without merging with /i/. I changed the original Cyrillic script for Tajiki to Latin for the sake of convenience. However, whenever necessary, I will indicate the Cyrillic spelling of a word in literary Tajiki in a note.

Table 1: Table 1: Samples from Tajiki dialects illustrating the "historically unmotivated" *majhul*

e(h) < **ī**h	e(h) < **ī**h	e(ʔ), e(h)[4] < **ī**ʔ
Badakhshani[5] /gire/[6] 'knot'	Badakhshani /saheh/[7] 'true'	Darvazi /toleh/[8] 'fortune'[9]
Kulabi /gᵊre/[10]		Panjakenti /toleh/[11]
Qarataqi /gireh/[12]		Kulabi /tobeʔ/[13] 'subject'
Qarategini /farbe/[14] 'fat'		Darvazi /tobeh/[15] 'district'
Ashti /farbeh/[16]		Kulabi /vose/[17] (personal name)
Kassansayi /farbeh/		Kulabi /neʔmatulo/[18] (personal name)
Leninabadi /farbeh/		Kulabi /badfeʔl/[19] 'evil-doer'
Kanibadami /farbeh/		
Badakhshani /menat/[20] 'labor'		
Kulabi /mehnat/[21]		
Varzabi /mehnat/[22]		
Farghanayi /mehnat/		
Sokhi /mehnat/		
Chusti /menat/		
Badakhshani /de(h)qon/[23] 'peasant'		
Qarategini /dehqun/[24]		
Darvazi /dekun/[25]		
Varganza[26] /deqon/[27]		
Kulabi /be(h)tar///be/[28] 'better'		
Goroni[29]/betar/[30]		
Panjakenti /beh/[31]		

4 Arabic /ʔ/ in borrowed words is regularly replaced with /h/ in some southern Tajiki varieties.

5 The Tajiki varieties of the Badakhshan area on the Tajiki side of the border between Tajikistan and Afghanistan.

6 Rosenfeld (1982: 80). Here and below in the column the word in literary Persian, in Arabic script: گره, in literary Tajiki (in Cyrillic script): гиреҳ.

7 Murvatov (1982: 140 n. 966). The word in literary Persian, in Arabic script: صحیح, in literary Tajiki: саҳеҳ.

8 Rosenfeld (1956: 201). Here and below in the column the word in literary Persian, in Arabic script: طالع, in literary Tajiki: толеъ.

9 The examples illustrating the shift e(h) <**ī**h and e(h) <**ī**ʻ could have largely increased if religious terminology had also been available as in the case of Afghan Persian varieties (see below). However, Soviet scholars whose works I use here rarely focused their attention on that terminology for ideological reasons.

10 Sokolova et al. (1952: 158).

Despite the fact that very few dialects of Afghan Persian have so far been studied, there is convincing evidence that the above phonetic phenomenon is not confined just to Central Asia and is largely shared by Afghan Persian varieties. For instance, it regularly occurs in the dialect of Kabul:

Table 2: Samples from the Kabuli dialect illustrating the "historically unmotivated" *majhul*

Kabuli dialect	The spelling of the literary form in Arabic script	English translation or meaning
/dē/ (</deh/)[32]	ده	'village'
/bē/ (</beh/) //bētar/ (</behtar/)	بهتر / به	'better'
/gerē/[33] (</gereh/)	گره	'knot'
/dēqɒn/ (</dehqɒn/)	دهقان	'peasant'
/ʧēra/ (</ʧehra/)	چهره	'face'
/mēr/ (</mehr/)	مهر	'love'
/tazbē/ (</tasbih/)	تسبیح	'prayer beads'
/tarɒbē/ (</tarɒwih/)	تراویح	extra prayers performed by Muslims at night in the month of Ramadan
/ʃēr/ (</ʃeʔr/)	شعر	'verse, poetry'
/ētebɒr/ (</eʔtebɒr/)	اعتبار	'trust; validity'
/tɒlē/ (</tɒleʔ/)	تالع	'fortune'
/samē/ (</samiʔ/)	سمیع	personal name
/wɒqē/ (</wɒqeʔ/)[34]	واقع	'happening'
/ētemɒl/ (</ehtemɒl/)	احتمال	'probability'

---

11 Ivanova (1956: 282).

12 Uspenskaya (1956: 222).

13 Murvatov (1982: 140 n. 966). Here and below in the column the word in literary Persian, in Arabic script: تابع, in literary Tajiki: тобеъ.

14 Murvatov (1982: 54 n. 2207). Here and below in the column the word in literary Persian, in Arabic script: فربه, in literary Tajiki: фарбеҳ.

15 Rosenfeld (1956: 201).

16 Here and below in this column unless otherwise indicated: Rastorguyeva (1963: 196).

17 Nemenova (1956: 100 n. 543). The name in literary Persian, in Arabic script: واسع, in literary Tajiki: Восеъ.

18 Nemenova (1956: 91 n. 244). In the cited source, the word is used with the object marker. The name in literary Persian, in Arabic script: نعمت الله, in literary Tajiki: Неъматулло.

19 Nemenova (1956: 95 n. 406). The word in literary Persian, in Arabic script: بدفعل, in Tajiki: бадфеъл.

20 Rosenfeld (1982: 124). Here and below the word in literary Persian, in Arabic script: محنت, in Tajiki: меҳнат.

21 Murvatov (1982: 175).

Isolated examples of this shift are also recorded in the Yakaulangi dialect of the Hazara: /dēɣů/ 'peasant', /ʧēra/ 'face' (Efimov 1965: 22).[35]

A different situation is observed in western Afghanistan, which is an area of Khorasani type varieties. In that part of the country, historic ĭ immediately preceding [h] or [ʔ], is realized as a more open vowel than [e], which normally corresponds to historic ĭ in other cases.[36] Thus, the form /dæɣoni/ 'agricultural labor' was recorded by me in 1981 from a native of the province of Badghis and from a native of Ghor. The latter person provided another sample illustrating the same shift: /dɛʋt-æ digə/ 'other villages'. However, as is already clear from this example, a more open articulation is not confined in his speech just to positions considered in this article. This is also attested by another sample recorded from the same dialect speaker: /dɛl bɛ dɛl/ 'heart to heart'.

Vowel /e/ (corresponding to historic ĭ) is often but not regularly realized as a more open sound in the position before [h] or [ʔ] in the varieties of Herat province that also belong to Khorasani type dialects.[37] The samples in which

---

22 Here and below in this column unless otherwise indicated: Rastorguyeva (1964: 135).

23 Rosenfeld (1982: 66). Here and below the word in literary Persian, in Arabic script: دهقان, in Tajiki: деҳқон.

24 Murvatov (1982: 98 n. 839).

25 Rosenfeld (1956: 202).

26 Varganza is a small town in Uzbekistan whose population consists mostly of Tajiks.

27 Mahmudov (1978: 13).

28 Murvatov (1982: 256 n. 2297, 188 n. 369, 211 n. 541). Here and below the words in literary Persian, in Arabic script: بهتر / به , in Tajiki: беҳтар/ беҳ.

29 Goroni is one of the Tajiki varieties in Gorno-Badakhshan Autonomous Province in southern Tajikistan spoken in the small mountainous town of Goron.

30 Bogorad (1963: 56).

31 Ivanova (1956: 318).

32 Here and below unless otherwise indicated: Farhâdi 1955: 19–20, 38 (original transcription changed).

33 In some dictionaries, published in the former Soviet Union, the word is transcribed as *gera*, which is not correct. The latter form originally occurred in Bogdanov's (1930) Stray Notes on Kabuli Persian. Bogdanov, who was a pioneer of Afghan Persian studies, did not distinguish between the local dialects of Afghan Persian, which was due to the beginning stage of these studies at the time. The above form of the word, in case it had been recorded correctly, could not have come from a native speaker of the Kabuli dialect.

34 Here and below unless otherwise indicated: Pakhalina (1964: 30, 32, 58) (original transcription changed).

35 The widespread occurrence of this phenomenon in Afghan Persian varieties is reflected in the transcription of the folk quatrains from different parts of Afghanistan, published by Asadollah Sho'ur (1974: 409).

36 What is implied here is [e] in such words as [del] دل 'heart', [kɒmel] کامل 'complete', etc.

37 I have done a special study on the Herati dialect, which has resulted in a series of publications, including a monograph.

the change occurs at the end of the word most clearly illustrate this tendency. It is important to point out that since in the latter case it concerns the final stressed syllable, which is the most favorable position for preserving all the qualitative and quantitative characteristics of the vowel in question, the change cannot be caused by such factors as reduction. Here are a few examples from the Herati dialect: /gʊrə/[38] 'knot', /*zærapuʃ*/ 'armored personnel carrier', /matʃetdʒɒma///mastʃetdʒɒmɛ/ 'great mosque' (the corresponding forms in literary Afghan Persian are /gereh/, /zerehpōʃ/, /masdʒed-e dʒɒmeʔ/ (<**dʒāmĭʔ**).

While historic **ĭ** in the described position undergoes change (which I will discuss in more detail later), historic **ī** in the same position preserves its quality in Herati unlike Tajiki, Kabuli, and some other dialects, where it has changed into *majhul* vowel (see above): /tespi/ 'prayer beads', /tafri//təfri/ 'recreation, amusement', /piy/ 'tallow' (the corresponding forms in literary Afghan Persian are /tasbih/, /tafrih/, /pih/).[39]

The changing of [e] into a more open vowel before [h] also occurs in the Persian varieties of the Iranian part of Khorasan, including the dialect of Mashhad: /mæhræbani/[40] 'kindness', /bædæ/ – imperative of the verb "to give", /dæ/ 'village' (Massé 1925: 93, 101, 73), /mæh/ 'mist', /motavadʒdʒæh/ 'attentive'[41] (the corresponding forms in literary Iranian Persian are /mehrebɒni/, /bedeh/, /deh/, /meh/, /motavadʒdʒeh/). Examples illustrating the same shift are also found in W. A. Ivanow's materials from northern (Iranian) Khorasan (Ivanow 1925: 269 [quatrain number 52], 292 [quatrain number 171]): /bæytærum ku/ 'make me better', /tʃahræ/[42] 'face' (the corresponding forms in literary Persian are /behtær-æm kon/, /tʃehre/).

The fact that in Khorasani dialects on either side of the Iran-Afghanistan border, the vowel corresponding to historic **ĭ** in the position before [h] or [ʔ] often assumes a very open articulation changing into [a], [æ], [ə], [ɛ], excludes any possibility of it being classed as a variant of the *majhul* because wherever the latter is preserved in Khorasani varieties (it occurs occasionally as an archaic

---

38 This symbol in a final position implies a variant of æ (a low or near-low front unrounded vowel) with a less clear articulation.

39 Literary Afghan Persian does not reflect the changing of historic **ĭ, ī** before [h] or [ʔ] into *majhul* /ē/. Therefore, I contrast Herati (and Khorasani varieties in general) with the dialect of Kabul where this change occurs (see examples in one of the passages above) and not with literary Afghan Persian. Literary Tajiki, on the other hand, does reflect it: [tasbeh] (тасбех) 'prayer beads', [tafreh] (тафрех) 'recreation, amusement', etc.

40 In Khorasani varieties [ɒ] in an unstressed syllable tends toward losing its rounded quality and merging into [a]. This phenomenon occurs often but not regularly.

41 The latter two examples were recorded by me from a native of Mashhad.

42 Original transcription changed.

relic in them), it is normally a very close [e] ranging between close [i] and close-mid [e]. For further proof of the validity of this argument, let us turn to Herati. In this dialect, [ē] is not a separate phoneme because of a progressive tendency toward gradual merging together of [ē] and [i] into one phoneme. This is best illustrated by the existence of interchangeable doublet forms: /ʃēr///ʃīr/ 'lion', /dēr///dir/ 'long', etc.[43] The [ē], which is but an allophone of one phoneme /[i]/[ē]/ here, shows a clear tendency to a more close articulation and is not realized as more open vowels ([a], [æ], [ə], [ɛ]). A very close realization of *majhul* [e] (/ē/) is observed even in Tajiki varieties (see above) and the Kabuli dialect[44] in which it is preserved as a separate phoneme. With regard to Tajiki, this is best illustrated by the alternation [e]/[i] in the transcription of this vowel, used by scholars for Tajiki varieties: /gire///giri/ 'knot' (Rosenfeld 1982: 80, 1971: 179, 1962: 129), /farbe///farbi/ 'fat' (Murvatov 1982: 7, 54), /vose///vosi/ – a personal name (Nemenova 1956: 100 n. 543, 123 n. 1219), etc.

The described situation in Tajiki and Kabuli dialects contrasts sharply with that in Herati. In the latter, the common variants corresponding to historic ĭ, in the position before [h] or [ʔ]: [a]/[æ]/[ə]/[ɛ] are not within the range of allophonic variation of historic *majhul* ē. In fact, they fall within the acoustic range of other phonemes, such as /e/ and /a/, merging into their allophones. (It is hard to define more precisely to which of the two phonemes a particular vowel in this position belongs because of the neutralization of phonemic distinction between /a/ and /e/ ([ɛ]) due to the common alternation [a]/[ɛ] (<a) in the final position.)

It will not be correct to assume that the more open variants ([a]/[æ]/[ə]/[ɛ]) in Herati are special positional realizations of the *majhul* vowel restricted to a final open syllable because linguistic data indicate otherwise. Thus, whenever historic *majhul* occurs (as an archaic relic) in a final open stressed syllable, it preserves its typical quality of close [e], which merges into [i]:[45] [nē] / [ni] (<**nē**) – a word of denial or refusal, [bi] (<**abē**) 'without'. On the one hand, the phonological opposition between the vowel variants corresponding to historic

---

43 See also Ioannesyan (1995: 226, 2007: 260, 2009: 2).

44 In Kabuli, /ē/ (often transcribed as /ê/) is realized as a close-mid vowel. See Pakhalina (1964: 46).

45 Had it been otherwise, it would have run counter to the general trend of the evolution of the phonological type of Iranian languages. According to D. I. Edelman, "historic long front close-mid vowels (like **e** and **o**) have been preserved with no qualitative changes for a long period in some [Iranian] languages, while in others they show a tendency to shifting to closer sounds: **o**> /ů/ in Tajiki, **o**> /u/ in the Iron dialect of Ossetic, and in Persian. However, **e** and **o** are nowhere realized as more open vowels" (Rastorguyeva and Edelman 1975: 64; translation mine).

majhul ē and those corresponding to historic **ǎ** in this position is preserved: [bi] (<**abē**) – [bə]/[bɛ] (<**pad**) – a prefix and a preposition with different grammatical meanings, while [a]/[æ]/[ə]/[ɛ] corresponding to historic **ǐ** in a final syllable, in the position before [h] and [ʔ] fall within the allophonic range of **ǎ** in a final syllable and merge with it. Thus, [ə] in the word /gʊrə/ (literary Afghan Persian: /gereh/) 'knot' is qualitatively no different from [ə] in /ʃurə/ (literary Afghan Persian: /ʃura/) 'saltpeter', [ɛ] in /masʧetʤʊmɛ/ (literary Afghan Persian: /masʤed-eʤʒomeʔ/) 'great mosque' is identical with [ɛ] in /ʤɛmɛ/(literary Afghan Persian: /ʤomʔa/) 'Friday' and in /moʤasɛmɛ/ (literary Afghan Persian: /moʤassama/) 'statue'.

The positional variants of the vowel (corresponding to historic **ǐ** before [h] and [ʔ]) are also remarkable. Its most open variant [a] is observed when it precedes a word or a morpheme starting with [a], while its most close allophone [e] occurs before [y]: /pʊk maʧetʤʊma aste/ '(this) is a complete great mosque', cf. /maʧetʤʊme-ye erɒt/ 'the great mosque of Herat', /maʧetjʊme-ya/ '(this) is a great mosque'. The phoneme corresponding to historic **ǎ** shows a tendency to the same allophonic distribution (at the end of a word, its open variant occurs if the following word or morpheme begins with [a], while the close allophone is observed before a word or a morpheme starting with [y]):

Table 3: Samples demonstrating the allophonic distribution of the phoneme corresponding to historic ǎ, which depends on the sound that follows it

Open variant	English translation of the sample	Close variant	English translation of the sample
/goʃnaam astom/	'I am also hungry'	/madgöu guʃne-ye/	'the cow is hungry'
/deraxt-e pɛstaam aste/	'there is (also) a pistachio tree'	/pəste-ye xandɒn/	'a partly open pistachio nut'
/bəm tay-ye kisaam bud/	'it was in my purse'	/se kise-ye sadtumani/	'three purses with a hundred *tomans* in each'
/barra am piʃ-e az in-æ/	'there is even a lamb with it'	/bare-ye dige/	'another lamb'
/ʧel tʊ korram (<korraam)ba rad-e az i/	'forty colts are even (running) after her'	/korre-ye pari/	'a fairy colt'
/rame-ye gusfand-a/	'it is a flock of sheep'	/migæ i rama az kin-a?/	'he says: "whose is this flock?"'
/vazifa anʤʊm mide/	'he carries out his responsibility'	/vazife-yi-râ anʤʊm midaam/	'I carry out the responsibility'

It should be noted that in Herati, the word "knot" and the word "crying" may sound the same as the result of a vowel reduction under the influence of [r] occasionally leading to a complete vowel drop: /gir/(</gɪrə/</gereh/) – /gir//ger/ (</gɪrɛ/</gerya/). This shows that [ə] corresponding to [eh] in literary Afghan Persian is subject to the same phonetic alterations as [ɛ] corresponding to [(y)a] in the final position.

The analysis of the linguistic factors above indicates that the vowel corresponding to historic ǐ before [h] and [ʔ] in Khorasani varieties is not *majhul*/ē/. Since all the variants of the former represent one phoneme realized as different allophones, this conclusion holds true for all its acoustic variants, open and closed alike. As has been illustrated above, historic ī does not become [ē] in this position either. Consequently, in Khorasani, historic short and long ǐ, ī do not change into *majhul* /ē/ before glottopharyngeal consonants in contrast to the eastern group of dialects.

It is worth noting an interesting phenomenon. In the dialects where the changing of historic short and long **i** into *majhul* /ē/ before glottopharyngeal consonants takes place, the same *majhul* vowel also occurs in the final position of the two words: "three", "Saturday" (as well as in composites containing these words), which appears etymologically unjustified (cf. Iranian Persian /se/, /ʃembe]/: /se (sē)/ 'three' in Varzabi, Farghanayi, Shahristani, and other varieties of Tajiki (Rastorguyeva 1963: 175, 1964: 71), cf. /sē, sēyom/sēwom/ 'three, third' in Kabuli (Farhâdi 1955: 19, 61–62) and /se (sē)/ 'three' in the Yakaulangi dialect of Hazara (Efimov 1965: 26); /yakʃambe (-ē)/ 'Sunday', /tʃorʃambe (-ē)/ 'Wednesday' in Kulabi (Nemenova 1956: 89 n. 198, 148 n. 3); /tʃorʃanbe (-ē)/ – a form common to all southeastern Tajiki varieties, /payʃanbe (-ē)/ 'Thursday' in Badakhshani (Rosenfeld 1982: 53, 145), /ʃambe (-ē)/ 'Saturday', /seʃambe (-ē)/ 'Tuesday' in Varzabi, Farghanayi, Ura-Tyube, and Shahristani varieties; /ʃæmbe (-ē)/ 'Saturday' in Rishtani (Rastorguyeva 1963: 176, 220),[46] cf. /pain-ʃambē/ 'Thursday' in Kabuli (Farhâdi 1955: 31).

In the Khorasani dialects, on the other hand, these words are characterized by alternation [e/ɛ/æ] in the same position, which, as has been pointed out above, excludes the possibility of the vowel being considered a variant of *majhul* /ē/: /se//sæ/, /sæyyom/ 'three, third' in the dialect of Mashhad (Massé 1925: 85, 89, 90, 93),[47] /sæ/ 'three' in the dialect of Qayen (Zomorrodian 1974: 90), /ʃæmbæ/ 'Saturday' in the dialect of Sabzivar (Ivanow 1925: 292 n. 167), /sɛ//sæ/

---

46 The vowel in both words in Tajiki is definitely *majhul* [ē]. Had it been otherwise, the words would have assumed the forms /sa/, /ʃamba/, because there is no [e] in Tajiki other than *majhul*.

47 Here and in the examples below, original transcription changed.

'three', /ʃombe/ / /ʃombɛ/ 'Saturday' in the dialect of Herat. In the latter dialect, there is also a composite word /sætabaɣai/ 'three story (building)', which also confirms the fact that the vowel in /sɛ///sæ/ is not *majhul* /ē/. The fact that the vowel in question is short is reflected in the transcription of words containing the numeral in a Herati dialect vocabulary: سپایه 'tripod', سپولی '*sepuli*' (name of a coin), سلا 'threefold' where the Herati forms are contrasted with Kabuli: سه پایه ،سه پولی (Fikrat 1976: 99). Here are more examples from Kabuli illustrating the shift /se/ > /sē/: /sē-rōza/ 'three day long', /sē-pɒya/ 'tripod', /sē-gɒna/ 'triple, of three kinds', /sē-barɒbar/ 'three times [larger or more]' (Farhâdi 1955: 63, 103, 123). It is highly remarkable that the word /sē-rōza/ rhymes with /fērōza/ 'turquoise' and /mē-sōza/ 'is burning' in popular Afghan Persian poetry.[48]

The above facts indicate that the difference in the realization of the vowel in these two words, on the synchronic level, may also be considered an important distinctive feature between the eastern type varieties and the central (Khorasani type) group, while the seemingly unexplainable changing of historic ĭ into *majhul* in these words in the eastern group dialects, where it normally occurs under the influence of glottopharyngeal consonants, may provide evidence for the fact that the vowel in the above words might have originally been followed by [h]. That the final *majhul* in the two words must have occurred at some point in history due to environmental factors only is obvious from the fact that had it been otherwise, the corresponding variants in western and central group varieties would have ended in /i/ and /[i]/[ē]/ respectively (these phonemes in the mentioned varieties regularly correspond to /ē/ in eastern dialects) assuming the following forms: /si/ 'three', /ʃambi/ 'Saturday' in the western group and /si/sē/, /ʃambi/ʃambē/ 'Saturday' in the central. However, this is not the case. The influence of [h], which with respect to changing the vowel into *majhul* is confined to the eastern group, could well have explained the situation. Historically, there is no ground to reject such a possibility, which makes this conclusion even more likely. Let us consider some historic and phonetic reasons for this hypothesis.

The Middle Persian numeral "three" is spelled with **h** in Manichaean texts: **sh**[49] which is very remarkable, given the fact that Manichaean spelling was less conservative or historical and corresponded more closely to contemporary pronunciation than Pahlavi Middle Persian spelling. It is highly unlikely that the final **h** in the word could have been a silent letter, an orthographic feature representing a terminal **e** or another vowel. The most convincing proof that it

---

48 See Sho'ur (1974: 219 n. 3).
49 See MacKenzie (1986: 74), Durkin-Meisterernst (2004: 306, 423).

was not the case is the Manichaean spelling of the words ***bē/bě*** 'but, out, except' : ***b'***, ***kē*** 'who, which': ***ky*** (MacKenzie 1986: 18, 50), which clearly indicates that a terminal **e**, whether long or short, was denoted by symbols other than a silent letter **h**. That final **h** in Manichaean Persian writing system is not a mute symbol is also illustrated by ***dah*** 'ten' (in transliteration: ***dh***,[50] which is analogous to ***sh***), ***rěh/rah*** 'chariot' (in transliteration: ***ryh***, ***rh***),[51] etc. In some early New Persian classical texts, according to Lazard (1963: 216), the word "three" was also spelled as ***sih*** (in transliteration: ***syh***). These facts confirm the viewpoint of those scholars (Lorimer 1922: 69, 135; Wolff 1935: 534; Tedesco 1921: 198) who trace back /se/ 'three' in New Persian to an older form ***sih*** (***seh***).[52]

Similarly, there are historic grounds for assuming that the word "Saturday", at least in one of its variants, might have originally ended in **h,** which eventually dropped. Not surprisingly, such a variant is found in very few (mostly Persian-Persian) dictionaries, though missing in the majority of others.[53] Thus, ***ʃanbĭh***, in pronunciation: ***ʃambĭh*** (where **h** is not a mute sign) is listed among the phonetic variants of the word "Saturday" in the dictionaries of Dehkhoda and Mo'in. The first gives the following forms: [شَمْ ب / بِ / بِ هْ] شنبه (Dehkhoda, relevant entry), which correspond to ***ʃamba/ʃambĭ/ʃambĭh*** (!) in romanization. The latter three variants are also represented in Mo'in's dictionary in his modern Persian transcription: /šanba(-e), šanbeh/ (Mo'in 1977–1978, 2: 2081).

Let us now consider some linguistic and historic grounds for the existence of the form with a final [h]. In classical Persian, the word often occurred as ***ʃanbĭd*** or ***ʃanbĭð***, spelled in Arabic script as شنبد or شنبذ[54] respectively. In Firdowsi's *Shāhnāme*, one comes across a composite word: یکشنبدی ***yakʃanbĭdī*** 'Sunday' (Wolff 1935: 575, 882). Interestingly, forms /seʃembet/ and /seʃambet/ 'Tuesday' were recorded by Zhukovski (1922: 261) at the beginning of the twentieth century in the Shiraz region, in the villages of Sivand and Abdu respectively, while /ʃæbbat/ 'Saturday' (with [a] denoting historic **ā**), /yeyʃæbbat/ 'Sunday', /tʃarʃæbbat/ 'Wednesday', etc., are found in the Judeo-Persian variety of Isfahan (Abrahamian 1936: 67, 68, 70, 72).[55]

All the above examples from classical Persian and the twentieth-century dialects testify to the fact that the word "Saturday" did have a variant ending in

---

50 See MacKenzie (1986: 23).

51 See Durkin-Meisterernst (2004: 295).

52 Cf. also Farhâdi (1955: 19).

53 It does not occur in Russian or Tajiki otherwise comprehensive dictionaries, including those focused on classical Persian.

54 These forms occur in classical texts and dictionaries (Ivanow 1923: 18; Vullers 1864: 469; Mo'in 1977–1978, 2: 2081).

55 Original transcription changed.

a consonant: **d**, **t**, or **δ**.[56] Alternation between dental **d**, **t**, **δ** / **z**, and **h** in intervocalic and postvocalic positions in some words is reflected in early classical Persian texts: **Harāt/Harāh** 'Herat' (city), **bāzu/bāhu** < **bāδūɣ** 'arm' (Lazard 1963: 148, 171), etc.

Classical Persian literature provides evidence for the existence of the form **ʃanbǐh**. In the "Seven Beauties" (*Haft Paikar*) by Nizāmī, there is an instance of the word combination: **ʃanbǐh bud** rhyming with **bǐh bud**:

شنبه آنجا که قسم شنبه بود
و آن دگرها چنان کز آن به بود
ʃanbǐh ānjā kǐ ɣǐsm-ǐ ʃanbǐh bud
v-ān dǐgarhā tʃunān k-az ān bǐh bud[57]
[On Saturday the place prescribed for it,
And on the other days as it was meet[58]]

But the most remarkable and convincing evidence in favor of the variant **ʃanbǐh** (**ʃambǐh**) is provided by the expression **ʃambǐhī** (**shambihī**) **kardan** 'to keep a Sabbath, to sabbatize', occurring in some dictionaries (Johnson 1852: 765; Dehkhoda (n.d.). The form **ʃambǐhī** could only be derived from **ʃambǐh**. This, combined with the other facts we considered, conclusively testifies to the existence of **ʃanbǐh** (**ʃambǐh**) 'Saturday', in which **h** is not a mute sign with no phonetic value.[59]

The final conclusions of this article can be summarized in the following way. One of the distinguishing phonetic features of Tajiki and the majority of Afghan Persian varieties, which form the eastern group of the linguistic continuum of New Persian dialects, is the changing of historic short and long **ǐ**, **ī** into *majhul*

---

56 This is quite logical to expect given that the word had a consonant ending at the earlier stages of the history of west Iranian languages. Thus, its form in Middle Persian, according to 'Amid's dictionary, was **ʃunbat** ('Amid 1389: 710), while in Manichaean Parthian, it occurred as **ʃambat** (**ʃmbt**) (Durkin-Meistererernst 2004: 318, 418). Though the form **ʃambed** (**ʃmbyd**) suggested by de Blois for the word "Sabbath" in Manichaean Middle Persian, which would have matched **ʃanbǐd** or **ʃanbǐδ** of classical Persian (see above) perfectly, is questioned by Durkin-Meistererernst (de Blois 1998: 483; Durkin-Meistererernst 2004: 318), in my opinion, de Blois's viewpoint should not be discarded out of hand since the issue needs further investigation.

57 *Haft Paykar-e Nizāmi*; transcription mine.

58 Translated by C. E. Wilson and B. A. Lond (*The Haft Paikar*).

59 In an early Judeo-Persian translation of the Pentateuch from the Vatican collection, the word "Sabbath" normally occurs in transliteration as **ʃbt** (Paper 1965b: 99, 103–104). But there is an instance of it being spelled in a different form: **ʃmṭh** (Paper 1965b: 174). The form occurs in the combination: **s'l ʃmṭh** 'year of rest, sabbatical year' (Leviticus 25: 5). This might be another indication that the word had variants ending in **h**.

/ē/ before [h] or [ʔ]. This phenomenon does not occur in the western group em-bracing the Persian varieties of western and central Iran or in the central group, which consists of the dialects of northeastern Iran and northwestern Afghani-stan. In the latter group, though historic ĭ, ī before [h] can be realized as more open vowels, they do not match the acoustic characteristics of *majhul* /ē/ falling within the acoustic range of other phonemes. This feature (the absence of the described phenomenon) links the western and central groups together. Hence the linguistic boundary of the area within which the changing into *majhul* occurs is the line separating the two groups of dialects – the eastern and the central. The existence of this distinguishing phonetic feature enabled me to put forward the hypothesis that the final [ē] in the words /sē/ 'three' and /ʃambē/ 'Saturday' in the eastern group varieties, as opposed to the western and central groups in which the words end in [e] or in more open vowels, could well be ex-plained if we assumed that the *majhul* in the eastern group occurred under the influence of [h] (which might have been dropped at a later stage), a possibility for which there exists not only phonetic but also historic ground.

References

Abrahamian, Rouben. 1936. *Dialectes des Israélites de Hamadan et d'Ispahan et dialecte de Baba Tahir*. Paris: Librarie d'Amerique et d'Orient.

'Amid, Hasan. 1389. *Farhang-e Fārsi*. Tehran: Rah-e Roʃd.

Blois, François de. 1998. Review of: W. Sundermann (ed.), *Iranian Manichaean Turfan texts in early publications (1904–1934): Photo edition*. Journal of the Royal Asiatic Society of Great Britain and Ireland 8 (3). 481–485.

Bogdanov, Leonid. 1930. Stray notes on Kabuli-Persian. *Journal and Proceedings of the Asiatic Society of Bengal. N.S.* Calcutta, 1–124.

Bogorad, Y. I. 1963. Goronskiy Govor Tadzhikskogo Yazyka [The Goroni variety of Tajiki]. *Iranskiy Sbornik* (In honor of I. I. Zarubin), 44–59. Moscow: Izdatelstvo Vostochnoy Literatury.

Dehkhodā, Ali-Akbar. *Persian dictionary*. http://www.vajehyab.com/dehkhoda/%D8%B4%D9%86%D8%A8%D9%87-2 (accessed 12 June 2016).

Durkin-Meistererernst, Desmond. 2004. *Dictionary of Manichaean texts*. Vol. 3. Texts from Central Asia and China, edited by N. Sims-Williams. Dictionary of Manichaean Middle Persian and Parthian. Turnhout: Brepols Publishers.

Efimov, Valentin. 1965. *Yazyk Afganskikh Hazara: Yakaulangskiy Dialect* [The language of the Hazara people in Afghanistan: The Yakaulangi dialect]. Moscow: Nauka.

Farhâdi, Abd-ul-Ghafûr (Ravân). 1955. *Le Persan parlé en Afghanistan. Grammaire du Kâboli: Accompagnée d'un recueil de quatrains populaire de la région de Kâbol*. Paris: Centre National de la Recherche Scientifique.

Fikrat, Mohammad Asef. 1976. *Loghât-e Zabân-e Goftâri-ye Herât*. Ba Ehtemâm-e R. Farhâdi [A vocabulary of spoken Herati, with the assistance of R. Farhâdi]. Kabul: Baihaqi.

Haft Paykar-e Nizami (صفت بزم بهرام در زمستان و ساختن هفت گنبد) [The account of Bahram's feast in winter and the erection of seven domes]. http://www.nosokhan.com/library/Part/0NX (accessed 12 July 2016).

Ivanova, S. Y. 1956. Materialy po Penjikentskomu Govoru Tadzhikskogo Yazyka [Materials for a study of the Panjakenti dialect of Tajiki]. *Trudy Instituta Yazykoznaniya*, vol. 6: 281–342. Moscow: Izdatelstvo Akademii Nauk.

Ioannesyan, Youli. 1995. Mesto Geratskogo Sredi Dialektov Dari-Persidskogo Yazykovogo Massiva [The position of Herati among the dialects of the Dari-Persian linguistic continuum]. *St. Petersburg Journal of Oriental Studies* 7. 224–242. St. Petersburg: Peterburgskoye Vostokovedeniye.

Ioannesyan, Youli. 2007. Situating the Khorasani dialects within the Persian-Dari-Tajiki linguistic continuum. In Franklin Lewis & Sunil Sharma (eds.), *The necklace of the Pleiades. Studies in Persian literature*. Presented to H. Moayyad on his eightieth birthday, 259–270. Amsterdam & West Lafayette: Rosenberg Publishers & Purdue University Press.

Ioannesyan, Youli. 2009. *Afghan folktales from Herat. Persian texts in transcription and translation*. Amherst: Cambria Press.

Ivanow, Vladimir. 1923. Tabaqat of Ansari in the old language of Herat. *Journal of the Royal Asiatic Society of Great Britain and Ireland* 55. 1–34, 337–382.

Ivanow, Vladimir. 1925. Rustic poetry in the dialect of Khorasan. *Journal and Proceedings of the Asiatic Society of Bengal*, N.S. 21 (3). 233–313.

Johnson, Francis. 1852. *Dictionary, Persian. Arabic, and English*. London: Wm. H. Allen and Co.

Lazard, Gilbert. 1963. *La langue des plus anciens monuments de la prose Persane*. Paris: Librarie C. Klincksieck.

Lorimer, David. 1922. *The phonology of Bakhtiari, Badakhshani and Madaglashti dialects of modern Persian, with vocabularies*. London: Royal Asiatic Society.

MacKenzie, David Neil. 1986. *A concise Pahlavi dictionary*. London: Oxford University Press.

Mahmudov, M. 1978. *Lahjahoyi Tojikoni Rayoni Kitob* [The dialects of the Tajiki people of the Kitob region]. Dushanbe: Donish.

Massé, Henri. 1925. Contes en Persane populaire recueilles et traduits. *Journal Asiatique* N206. 71–157.

Mo'in, Mohammad. 1977–1978. *A Persian dictionary*. Vols. 1–6. Tehran: Sepehr printing house.

Murvatov, Ch. 1982. *Shevayi Janubiyi Zaboni Tojiki* [The Southern varieties of Tajiki]. J. 5. Dushanbe: Donish.

Nemenova, Roza. 1956. *Kulabskiye Govory Tadzhikskogo Yazyka: severnaya gruppa* [The Kulabi varieties of Tajiki: Northern group]. Stalinabad: Izdatelstovo Akademii Nauk Tadjikskoy SSSR.

Nizāmī of Ganja. 1924. *The Haft Paikar*. Translated from the Persian, with a commentary by C. E. Wilson & B. A. Lond. Vol. 1, Translation; vol. 2, Commentary. London: Probsthain. http://persian.packhum.org/persian/main?url=pf%3Fauth%3D176%26work%3D001 (accessed 12 July 2016).

Pakhalina, Tatyana. 1964. K Kharakteristike Kabulskogo Prostorechia [Concerning the characteristics of Kabuli colloquial speech]. *Indiyskaya i Iranskaya Filologia*, 55–59. Moscow: Nauka.

Paper, H. H. 1965a. The Vatican Judeo-Persian Pentateuch. *Acta Orientalia* XXVIII (3–4). 263–340.

Paper, H. H. 1965b. The Vatican Judeo-Persian Pentateuch. *Acta Orientalia* XXIX, 1–2. 75–181.

Rastorguyeva, Vera. 1963. *Ocherki po Tadzhikskoy Dialektologii: Tadzhiksko-Russkiy Dialectniy Slovar* [Essays on Tajiki dialectology: A Tajiki-Russian dictionary of dialects]. Vol. 5. Moscow: Izdatelstvo Akademii Nauk SSSR.

Rastorguyeva, Vera. 1964. *Opyt Sravnitelnogo Izucheniya Tadzhikskikh Govorov* [An experience of a comparative study of Tajiki varieties]. Moscow: Nauka.

Rastorguyeva, Vera & Joy Edelman. 1975. *Opyt Istoriko-typologicheskogo Issledovaniya Iranskikh Yazykov: Fonologiya. Evolutsiya Morfologicheskogo Tipa* [An experience of a historic and typological study of Iranian languages: Phonology. Evolution of the morphological type]. Vol. 1. Moscow: Nauka.

Rosenfeld, Anna. 1956. Darvazskiye Govory Tadzhikskogo Yazyka [The Darvazi varieties of Tajiki]. *Trudy Instituta Yazykoznaniya*, vol. 6: 196–272. Moscow: Izdatelstvo Akademii Nauk.

Rosenfeld, Anna. 1962. *Namunahoyi Folklori Darvoz* [Samples of Darvazi folklore], 2nd edn. Dushanbe: Nashrdavtodzh.

Rosenfeld, Anna. 1971. *Badakhshanskiye Govory Tadzhikskogo Yazyka* [The Badakhshani varieties of Tajiki]. Leningrad: Izdatelstovo Leningradskogo Universiteta.

Rosenfeld, Anna. 1982. *Tadzhiksko-Russkiy Dialektniy Slovar*: *govory yugo-vostochnogo Tadjikistana* [A dictionary of Tajiki dialects with a Russian translation: Southeastern Tadzhikistan varieties]. Leningrad: Izdatelstovo Leningradskogo Universiteta.

Sho'ur, Asadolla. 1974. *Tarânahâ-ye Kohsâr*. Ba Ehtemâm-e F. Faqiri [Songs of the mountains, with the assistance of F. Faqiri]. Kabul: Vezârat-e Ettelâ'ât-o Kaltur, Matba'a-ye Dawlati.

Sokolova, Valentina. 1949. *Fonetika Tadzhikskogo Yazyka* [The phonetics of the Tajiki language]. Moscow & Leningrad: Akademia Nauk SSSR.

Sokolova, Valentina, Roza Nemenova, Y. I. Bogorad, Vladimir Livshits & A. I. Farhadyan. 1952. Noviye Svedeniya po Fonetike Iranskikh Yazykov [New data concerning the phonetics of Iranian languages]. *Trudy Instituta Yazykoznaniya*. Vol. 1. Moscow: Izdatelstvo Akademii Nauk.

Tedesco, Paul. 1921. Dialectogie der Westiranischen Turfantexte. *Le Monde Orientale*. T. 15. 184–258. Uppsala: A.-B. AkademiskaBokhandeln.

Uspenskaya, Lyudmila. 1956. *Qarataqskiy Govor Tadzhiksogo Yazyka* [The Qarataqi variety of Tajiki]. Stalinabad: Izdatelstovo Akademii Nauk Tadjikskoy SSSR.

Vullers, Ioannis A. 1855. *Lexicon persico-latinum etymologicum*. Vol. 1. Bonnae ad Rhenum [New York?]: Impensis Adolphi Marci.

Vullers, Ioannis A. 1864. *Lexicon persico-latinum etymologicum*. Vol. 2. Bonnae ad Rhenum [New York?]: Impensis Adolphi Marci.

Wolff, Fritz. 1935. *Glossar zu Firdosis Schahname*. Berlin: Reichsdr.

Zhukovski, Valentin A. 1922. *Materiali dlya Izuchenia Persidskikh Narechiy. Part 2. Dialekty Polosy Goroda Semnana etc*. [Materials for a study of Persian varieties. Dialects of the region of Semnan, etc.]. Petrograd: Printing house of the Russian Academy of Sciences.

Zomorrodian, Reza. 1974. Le système verbal du Persan Parlé à Qâyen. *Studia Iranica*. Vol. 3, Fasc 1: 87–112. Leiden: Brill.

Zohreh R. Eslami, Mohammad Abdolhosseini, and Shadi Dini

8 Variability in Persian forms of address as represented in the works of Iranian playwrights

Abstract: In this study, a set of seven Persian plays was systematically searched for every instance of address forms. In total, a corpus of 500 address forms was collected. The data were analyzed within a shared framework from two classic studies on address forms: Brown and Gilman's (1960) second two-dimensional semantic model and Brown and Ford's (1961) "controlled induction" approach. The findings of the study indicated that Persian address forms can be classified into fifty-four types among which honorific title (HT), professional title (PT), generic noun (GN), first name (FN), general title (GT), and general title + last name (GT + LN) are the more frequent forms. Considering the semantic factors associated with the most common and other variant forms of address, it was found that in Persian the semantics of power should be expanded so as to include not only age and/or occupational differences but also religious status, ideologies, and political affiliations. Also, the results of this study question Brown and Ford's (1961) universalistic views about systems of address since the Persian system of address includes more address types compared to that of Western languages.

Keywords: address forms, Persian, playwrights, variational pragmatics, drama

1 Introduction

Address forms are words or expressions one uses to designate the person being talked with in the course of interaction. Address forms are socially driven and they signify the complex social relations of individuals (e.g., Dittrich, Johansen, and Kulinskaya 2011; Philipsen and Huspek 1985). By means of certain forms of address, interlocutors define their relationship to each other and identify themselves as part of a social group. As noted by different scholars (e.g., Dittrich, Johansen, and Kulinskaya 2011; Zhang 2011), the systematic use of variable address forms has sociocultural significance and conveys the attitude and dispositions of interlocutors toward each other.

Zohreh R. Eslami, Texas A&M University

Forms of address are also a reflection of cultural values and indicate political, social, and ideological changes. Much research on address forms has used Brown and Gilman's (1960) framework, which is based on two types of opposition: less formal pronouns (T) versus more formal pronouns (V) and between power and solidarity semantics. The power semantics stipulates that the nonreciprocal use of V by one interlocutor and T by the other expresses a power relationship and the reciprocal use of T or V marks a solidarity relationship. Although many researchers have confirmed Brown and Gilman (1960) model, some researchers have shown its limitations (e.g., Clyne et al. 2003; Hymes 1972; Kendall 1981; Nanbakhsh 2011). Clyne et al. (2003) argue that the dichotomy T/V can no longer simply be understood in terms of solidarity versus power. Nanbakhsh (2011) argues that the T/V model does not provide an accurate explanation for the mixed *šoma* pronoun use with the second person verbal agreement in Persian or the fact that *šoma* can be used as an address form among individuals who are intimate. Along similar lines, Hymes (1972) argued that power and solidarity models are not sufficient for the investigation of natural data. These researchers have questioned Brown and Gilman's (1960) dichotomous framework and universalistic claims due to the decrease in nonreciprocal pronoun use, detachment of T from the notion of solidarity in some social situations, dynamic and strategic use of address forms in intra- and interpersonal relationships, and the more complex and diverse markings of relationships as exemplified by the variety of address forms used in some societies versus others. As Braun (1988) stated, in their search for universal rules based on the abstract, idealized, and simplistic dimensions of power and solidary and of symmetry/asymmetry, researchers have neglected language-specific and culture-particular principles governing the proper usage of address terms. Further research is needed to expand the scope of the field by taking into account the dynamic use of address forms in different relationships and in societies where the relational role of interactants plays a significant and complex part.

This study aims to extend the investigation of address forms to Persian and to shed light on the existing literature on universality of address system. We examined the use of forms of address in a set of plays written by contemporary Iranian playwrights. Our aim was to investigate the applicability of Brown and Ford's (1961) universalistic claims and to determine whether or not the Persian system of address functions similarly to that of language(s) based on which Brown and Ford formed their universalistic claims. The findings of the study broaden the scope of research on address forms in different languages by including a less studied language (i.e., Persian).

2 Background

A brief review of the literature reveals that there are important cross-cultural and crosslinguistic variations in the way people use their address repertoire in different languages, and that cross-cultural differences in expectations of linguistic behavior, interpretive strategies, and signaling devices can lead to breakdown in interethnic communications. Obviously, without insight into the different social values and communication styles of diverse speech communities, one can easily form unwarranted value judgments. This, therefore, necessitates the need for a wider understanding of cultures and corresponding styles of communication.

Braun (1988) pointed out that Asian language address systems are marked with more variation in their repertoire of address terms than are languages in Western societies. Studies of address forms, from this perspective, show how clashes between different interaction styles can lead to intercultural miscommunications. For example, considering the exchange of first names between two interlocutors, Brown and Ford (1961) indicate that for American interlocutors to exchange such a term of address, i.e., first names, there should be only a very small increment of intimacy – as small sometimes as five minutes of conversation – whereas in many cognate languages today first names are used only after a long period of friendship (Brown and Ford 1961).

Brown and Gilman (1960) suggested a two-dimensional system of power and solidarity. According to this model, the terms of address used by a social inferior to a social superior may be different from those between peers. In Persian, similar to several other languages (e.g., Spanish, Italian, French), speakers have to make a choice between two forms of "you", the familiar and the deferential. The familiar "you" is used when an intimate relationship exists between the speaker and the addressee, or when the addressee is in a subordinate position. The deferential "you", on the other hand, is used when the relationship between the speaker and addressee is an asymmetrical relationship and the addressee is in a superior social position or when the speaker and addressee do not have a close personal relationship.

Persian address forms have attracted the attention of several authors (e.g., Eslami and Eslami-Rasekh 2007; Keshavarz 1988, 2001; Nanbakhsh 2011; Sharifian 2009). Keshavarz (1988) conducted a sociolinguistic analysis on the forms of address in post-revolutionary Iranian Persian using questionnaires. His findings show a sudden shift from power to solidarity due to the Islamic revolution and as he states, due to the change, the forms of address have become simplified. The findings show that forms of address marking solidarity

(e.g., "brother" and "sister") have gained popularity, whereas the asymmetrical forms reflecting the complex social class structure of prerevolutionary Iran have declined. In another study, Keshavarz (2001), using questionnaire data, indicated that variation in forms of address is related not only to the age, sex, and social class of the language users, but also to the social context (setting, intimacy, and social distance). Eslami and Eslami-Rasekh's (2007) study used questionnaire and interview data to examine the use of address forms by two different groups of religious and secular women in Iran. Contrary to Keshavarz's (1988) findings, Eslami and Eslami-Rasekh (2007) found that there was a shift from the use of religious/revolutionary terms of address in the earlier post-revolutionary years to more neutral terms of address (e.g., *xānom* 'Mrs.' and *āghā* 'Mr.' versus *xāhar*/ 'sister' and *barādar* 'brother'. As stated by Eslami and Eslami-Rasekh (2007), social, political, and religious values and affiliations have a direct impact on the choice of address forms, and that Iranian female language use is an indication of differences in the ways they perform their gendered identities.

Sharifian (2009), in line with Beeman's (1986) elaborate discussions on Persian *taarof* system and the use of singular and plural distinction as an indicator of status, highlights the degrees of respect being identified in the use of pronominal, verbal, and lexical use in address forms in Persian rather than the dichotomous use of solidarity/deferential forms. His study shows the intricate system of Persian address forms and various choices Persian speakers have at their disposal to mark their relations to their interlocutor. Nanbakhsh (2011) discusses sociolinguistic functions of address pronoun switching and the mismatch construction. Nanbakhsh emphasizes that a sociolinguistic variable may index a variety of social and pragmatic functions with a change of stance in discourse, and that address pronoun switching is not constrained by age or gender of interlocutors but by their interactional goals, their evolving social relationship in discourse, and stance taking in interaction (Nanbakhsh 2011).

The studies discussed above have focused on specific aspects of Persian forms of address (i.e., singular versus plural pronouns, the role of intimacy and distance). Our study uses an extensive corpus of 500 address forms from Persian plays to shed more light on the richness and variety of forms of address used in contemporary Persian. Furthermore, this study used dramatic representation of address forms in contemporary Persian plays. As asserted by different researchers (e.g., Moreno 2002; Moyna and Ceballos 2008), dramatic texts are sources of language use reflecting the speech of the time, and plays can represent a wide range of language use choices in a period of interest to the author.

3 Method

3.1 Data

The data in this study were collected from a set of seven plays written by different contemporary Iranian playwrights. This kind of data, in spite of their advantages, seem not to be extensively used by the researchers studying forms of address in different languages. The plays were chosen to provide a broad spectrum of social settings by different playwrights.

It should be noted that in particular cases, literary works can provide the investigator with more reliable data than actual situations because, according to Brown and Ford, "much of the everyday behavior that is governed by social dimension is difficult to record and involves an uncertain number of contrasts" (1961: 375). To reach a sufficient level of naturalness, playwrights prefer to write their plays using the features of the spoken rather than the written language. As stated by Moreno (2002), dramatic texts are sources of the best information available on the speech of the time and plays can represent the full range of society in a specific period of time.

As for the fitness of plays in sociolinguistic research, Brown and Ford (1961) point out that there are more instances of address terms in plays than in any other form of literature. Secondly, as stated by Brown and Ford (1961), studying social structure in everyday life, due to its being governed by a diverse amount of social dimensions, is difficult to record, and involves an uncertain number of constraints. Furthermore, as stated by Labov (1966), an informant's tendency toward standardizing increases when a greater degree of attention is directed to language. In contrast, in plays, since the non-prestigious forms are to be uttered by the characters of the plays rather than by the writer, playwrights feel free to reflect the actual norms of different social groups, even though they sound stigmatized or non-prestigious.

Several researchers have raised the shortcomings of data collected by interview and questionnaire. One major issue with a questionnaire is that the informant's impression of what an item intends to elicit might be different from that of the researcher who has designed the item. Also, in the questionnaire data, the language users have to imagine what they would say in different situations and this may not reflect what they actually say (Eslami and Mirzaei 2012, 2014).

Considering the feasibility of using plays, and the fact that they seem to be closer to natural language use, they were selected as the source of data for the present study.

The data for this study were selected from seven Persian plays, including *Majles-e Divān-e Balx* [Balkh court house] *Gowharshād* [Gowharshad/Name of a person or a place], *Majhul-ol-hoviyye* [Unknown identity], *Āheste bā gol-e sorkh* [Slowly with the red flower], *Monji dar sobh-e namnāk* [Messiah in the moist morning], *Kasb-e bi-hāsel* [Earn the ungainly], and *Maxmal* [Velvet] written by different Iranian playwrights. From these plays, more than 500 instances of terms of address exchanged between play characters were extracted. Then the data were analyzed to determine what linguistic variation occurred as the effect of social variables of solidarity and power influencing the interlocutors' relationship varied.

3.2 Research Focus

The aim of this study was to identify the range of address types used in a corpus of address forms collected from seven plays. We also identified the most frequently used address types in the corpus, their specific features, and the characteristics of other variant address types in terms of the associated semantic factors. Moreover, this study discusses the major dyadic patterns that govern the choice of address forms with regard to the effect of the social variables of power and solidarity.

3.3 Data analysis

The data in this research were analyzed within a shared framework using the two-dimensional semantic model introduced by Brown and Gilman (1960) and the "controlled induction" approach developed by Brown and Ford (1961). Brown and Gilman (1960) and Brown and Ford (1961) – by providing the notions of "reciprocity"/"non-reciprocity" (symmetry/asymmetry), of "power" and "solidarity", and of T and V pronouns – formed a theory of address terms and claimed that the rules and models in this theory suffice to explain address usage in different speech communities.

Based on the framework provided by Brown and Ford (1961), we applied semantic analysis to the study of address forms in a selected set of seven plays written by Iranian playwrights. The purpose was first to describe the norms of address used in a set of Persian plays, and to point out major dyadic patterns in which the extracted address instances function. To do this, from among all variables affecting the relationship between speaker and addressee, the effects of two social variables of power and solidarity were considered. Although the recent literature has revealed the effects of more complex factors such as

gender/age (Archibald Afful 2010; Liu 2009; Mahnmoodan 2013), and attitude (Koven 2009; Nanbakht 2011; Mahnmoodan 2013) on the speakers' choices of address terms and in spite of the arguments against the power-solidarity model for its context-free analysis (Liu 2009), power and solidarity, as indicated in the classic works of the literature (Brown and Gilman 1960; Brown and Ford 1961), still seem to play a more principal role in language variation. Furthermore, our aim was to provide a comprehensive account of the different forms of address in Persian based on the power-solidarity framework and examine if the findings of this study support the universalistic views (i.e., reciprocal and nonreciprocal use of V and T) held by Brown and Ford (1961) about systems of address in different languages.

The semantic factors associated with each major dyadic pattern and other variant forms of address are explained. As explained by Brown and Gilman (1960: 253), semantics of a certain form of address is a "covariation between the pronoun and the objective relationship existing between speaker and addressee".

For each of the plays, every instance of address terms was recorded, together with an identification of the speaker and addressee so as to make it possible to decide about the presence or absence of the two variables, i.e., solidarity and power, in the interlocutors' relationship. Then, with regard to the specified composition of the variables for every instance of address form, and using Brown and Gilman's (1960) two-dimensional semantic model, the address terms were classified into different categories.

4 Findings

4.1 Specification of available forms of address

The address forms collected from the plays (500 instances) were classified into fifty-four types indicating the different combinations in which Persian terms of address occur. These address types can be further placed into two subgroups: namely, simple and compound forms of address. The former includes terms of address that occur singly, and the latter comprises those with a combination of two, three, or four members.

Table 1 illustrates the frequency and percentage of each type of address form identified in the data. The abbreviations used in the table are defined in the following sections as we provide explanations about the distribution of different forms of address.

Table 1: Frequency and percentage distribution of Persian forms of address

No.	Address type	Frequency	Percentage
1	PT + 0	30	6.0
2	RT + 0	3	0.6
3	OFT + 0	6	1.2
4	HT + 0	74	14.8
5	OT + 0	10	2.0
6	GT + 0	29	5.8
7	GN + 0	49	9.8
8	FN + 0	38	7.6
9	LN + 0	9	1.8
10	TA + 0	1	0.2
11	NN + 0	21	4.2
12	KT + 0	13	2.6
13	TE + 0	6	1.2
14	FKT + 0	26	5.2
15	IKT + 0	1	0.2
16	NNOT + 0	4	0.8
17	NNFN + 0	8	1.6
18	GT + TE	7	1.4
19	GT + LN	48	9.6
20	GT + PT	13	2.6
21	GT + RT	1	0.2
22	GT + OFT	2	0.4
23	GT + FN	4	0.8
24	GT + GN	1	0.2
25	GT + OT	1	0.2
26	GT + NN	1	0.2
27	HT + OT	3	0.6
28	HT + OFT	7	1.4
29	HT + PT	7	1.4
30	HT + GT	2	0.4
31	HT + FN	1	0.2
32	HT + LN	3	0.6
33	RT + FN	2	0.4
34	RT + GT	7	1.4
35	PT + LN	1	0.2
36	CT + KT	1	0.2
37	FN + CT	9	1.8
38	FN + TE	7	1.4
39	FN + GT	2	0.4
40	KT + FN	4	0.8
41	KT + TE	4	0.8
42	FKT + FN	5	1
43	FKT + RT	1	0.2
44	FKT + TE	8	1.6
45	IKT + FN	1	0.2
46	IKT + TE	4	0.8
47	GN + TE	8	1.6
48	LN + TE	1	0.2
49	TA + TE	1	0.2
50	TR + TE	1	0.2
51	GT + KT + TE	1	0.2
52	HT + GN + TE	1	0.2
53	HT + FN + CT	1	0.2
54	GT + FN + CT + LN	1	0.2

In addition to the vast variety inherent in Persian address forms in terms of their constituent parts, such terms occur with different frequencies. As indicated in Table 1, among the fifty-four types of address terms determined, the ones that had the highest frequency included: HT, GN, PT, FN, GT, and GT + LN.

To determine whether or not these most frequent address terms, which mainly occur singly (except GT + LN), hold the same high frequency if considered in terms of both simple and compound instances, the occurrences of each of them both in simple and compound cases were counted. For example, to determine the frequency of PT, i.e., such titles as *doctor* 'doctor', *mohandes* 'engineer', etc., the number of its occurrences both as PT alone and in such combinations as PT + LN, HT + PT, GT + PT, etc., were counted. Table 2 shows the frequency distribution of these terms of address in both simple and compound cases.

Table 2: The frequency of the most common address forms

No.	Address type	Simple		Simple + compound	
		Frequency	Percentage	Frequency	Percentage
1	HT	74	14.80	99	19.80
2	GN	49	9.80	59	11.80
3	PT	30	6.0	51	10.20
4	FN	38	7.60	75	15.00
5	GT	29	5.80	120	24.00

The result, at this stage of analysis, confirmed that such terms of address are among the most frequent ones as compared with the total set of address types.

Below we present the most common (most frequently used) forms of address divided into two categories: simple and compound forms of address.

4.2 Most common simple forms of address

As mentioned earlier, simple forms of address are address terms that occur alone. This category consists of such types as: HT + 0, GN + 0, FN + 0, PT + 0, and GT + 0.

4.2.1 Honorific title (HT + 0)

In classifying the extracted forms of address into fifty-four categories, HT + 0, occurred with the highest frequency. Perhaps this is the case because the

usage of this term is not restricted to a certain setting in terms of power and solidarity. In other words, one can, more or less, witness the exchange of such terms of address in various dyads, regardless of the social status of the characters exchanging them.

The reason why such an address term is used in different settings, which thus leads to its predominance over other terms of address in Persian, is perhaps because individuals in Iranian society are expected, under normal circumstances, to show a certain degree of humility and humbleness when they refer to themselves and to show respect for others when addressing them. To use such terms appropriately, one should take into account such variables as age, sex, social status, and the degree of intimacy or distance between the participants.

Honorific terms, e.g., *enāb-e āli* 'your Excellency', *hazræt-e āli* 'your Excellency', etc., used by play characters in various settings, were most frequently used by inferiors in contexts where they wanted to show respect to their superiors out of politeness or flattery. For example, in the play *Majles-e Divān-e Balx* [Balkh Court House], almost all clients addressed the supreme judge with *jenāb-e āli*.

More specifically, these two honorific terms were restricted to adult male play characters and were mainly exchanged between power equals, and by upper-class characters. As such, one can claim that their use is constrained by age, gender, and social class.

The least common of the honorific terms extracted were *sarkār* 'your honor' and *qorbān* 'your honor'. The former was usually associated with the speech of working-class and uneducated people. For example, in *Gowharshād* the ordinary people in the street would address police guards as *sarkār* 'officer'. The second form, i.e., *qorbān* 'your honor' was used in both military and nonmilitary settings mainly to address superiors. For example, in *Majhul-ol-hoviyye* [Unknown identity], the doctor addresses both the head of the hospital and the inspector as *qorbān* 'your honor'.

4.2.2 Generic nouns (GN + 0)

GNs were used in situations where the speaker did not know the actual name of the addressee. A few examples from this category include *javān* 'young man'; *pesar* 'boy'; *pir-e zan* 'old woman' (an insulting term); *martike* 'little man' (an insulting term); *xāhar* 'sister', etc. The first GN of this set, *javān* 'young man', was typically used in settings with superior/inferior relationships. In such settings, the speakers, besides being older than the addressees, were superior in status. One would not come across this term being used by those inferior in

rank or status to address superiors. Therefore, it seems that the semantic load of this term emphasizes superiority of the speaker in terms of status and age, with age acting as the indispensable element.

In spite of the current arguments on the importance of context in determining impoliteness of a word/expression, some words/expressions can inherently be "impolite" (Culpeper 2010). Two such impolite and offensive terms of address in this category are *mart-ike* 'little man' and *zan-ike* 'little woman'. They are formed by the affixation of the diminutive marker of *-ike* (derived from *če* 'small') to such generic nouns as *mard* 'man' and *zan* 'woman', based on the rule of "diminutive-construction" in Persian. Such terms of address were abundantly found in the speech of some play characters when addressing the interlocutors whom they considered as guilty of doing unfavorable actions in both power-laden and equal settings.

4.2.3 First name (FN + 0)

The third category of the simple forms of address in the data was the FN category, which included full first names (e.g., Ali, Minā), familiar abbreviations, e.g., *ebrām* from *ebrāhim* and diminutive forms, e.g., *feri* from *farhād* or *fereydun*.

In addition to nominal terms of address in the data, the second person singular pronoun, i.e., /to/ 'familiar you', was considered as an instance of FN in the data. In almost all dyads where the interlocutors used a FN term, one would also find instances of *to* in their exchange. No instances of *to* were detected where the interlocutors did not use FN, insofar as the marked (deviant) uses of address terms are excluded.

With regard to the semantic factors associated with the usage of FN + 0, it was noticed that FN was common not only between intimate characters in settings ranked with equality and solidarity, but it was also used frequently by superiors to known inferiors. No instances of FN were found to be exchanged in non-solidarity settings, no matter whether they were power-laden or equal. Also, FN was almost never used by inferiors to address their superiors. The only case spotted was in *Āheste bā gol-e sorx* [Slowly with the red flower] where the rich younger teenage addressed his elderly poor cousin with FN + 0, which may indicate the importance of status over age in some situations. In this respect, Brown and Ford (1961) mention numerous instances in plays where status gains priority over age. They, for instance, mention settings with such interlocutors as an adolescent girl and her family's middle-aged cook, a young navy ensign and a middle-aged enlisted man, a young executive and an elderly janitor. In all such cases, address was found to be in accordance with occupa-

tional status, so they concluded that there would be a normative rule of priority for the two criteria – that occupation would prevail over age in the determination of deference.

4.2.4 General titles (GT + 0)

The category of general titles in this study includes such titles as *āqā* 'Mr.', *xānom* 'Mrs.', and *dušize* 'Miss'. The term *āqā* is a common male title in the Persian language, which was exchanged by play characters in almost all settings marked differently with power and solidarity variables. But the majority of the instances of the use of this title were extracted from settings where inferiors were addressing their superiors.

4.2.5 Professional titles (PT + 0)

Professional titles (PTs) are another category of simple, most frequently used forms of address. It is worth mentioning that a distinction is made between such titles and other similar terms of address like official titles (OFTs) and occupational titles (OTs).

Following Braun's (1988) categorization, we have made a distinction between two groups of titles. Professional titles are defined as the titles of those who are specifically trained for that purpose. Examples are *doctor* 'doctor'; *mohandes* 'engineer'; *ostād* 'instructor'; *qāzi* 'judge', etc. The second group, i.e., official titles, includes certain types of titles for the achievement of which no institutional training is necessarily required. These positions can be inherited or awarded. Examples include *vazir* 'minister', *ra'is* 'boss/head', etc., which were placed in this category.

As shown in our data, PTs were not only used by the inferior-in-rank interlocutors; they were also exchanged between power equals, particularly when a third party was present, regardless of the status, age, and gender of this third person. It is likely that the exchange of professional titles in the presence of others by power equals who would otherwise use FN or GT + LN terms signifies the importance of attending to social privilege and higher social status in Iranian culture.

4.3 Most common compound forms of address

Compound forms of address are address terms that co-occur with other terms. One of the most frequent compound cases in the data was GT + LN. In this form of address, one of the general titles, i.e., *āqā* 'Mr.', *xānom* 'Mrs.', and

dušize 'Miss', precedes the addressee's last name. *āqā* 'Mr.', *xānom* 'Mrs.', in such cases, may change to *āqā-ye* and *xānom-e* according to the Ezafe Construction in Persian, like *āqā-ye* Ahmadi 'Mr. Ahmadi'. "Ezafe" is a grammatical particle that connects a noun to another noun, pronoun, or adjective, and its function is almost the same as that of the English preposition "of". This way of addressing is typical of settings where the interlocutors know each other. Otherwise, the speaker would use other possible forms of address.

With regard to power semantics associated with such an address term, one would find that such forms of address were received by inferiors, superiors, and power equals, provided that the solidarity variable was present. Table 3 shows the frequency of GT + LN in different settings.

Table 3: Frequency of GT + LN in different settings

Address form	Superior solidary	Superior not solidary	Equal solidary	Inferior solidary	Inferior not solidary
GT + LN	13	2	20	13	0

In general, these address forms (i.e., GT + LN) were used like PTs, either reciprocally between power equals or nonreciprocally by a person of lower status to address an interlocutor with a higher status. The address form GT + LN seems to be more intimate and less deferential than PTs, as the semantic analysis of such terms of address showed.

4.4 Other variant forms of address

The second group of address forms include several other forms that had a high level of occurrence but not as high as the ones already presented above. For the sake of clarity, they are divided into two groups of simple and compound address forms.

4.4.1 Simple forms of address

4.4.1.1 Kinship terms (KT + 0)

Kinship terms of address were used for blood relations and for affines. This category included such terms as *bābā*, *pedar*, and *āqā* 'father'; *māmān* and *mādar* 'mother'; *nane* 'grandmother'; *pedar/bābā bozorg* 'grandfather'; *dādāš* and *barādar* 'brother'; *xāhar* and *ābji* 'sister' *amu* 'paternal uncle'; *dāyi* 'maternal uncle'; *amme* 'paternal aunt'; and *xāle* 'maternal aunt'. These terms were

employed in a familial context by younger people to address those who were older. In situations where the age differences between the interlocutors was not considerable (i.e., about ten years) they exchanged FNs.

4.4.1.2 Inverted kinship terms (IKT + 0)

Inverted kinship terms appeared in certain situations where a KT was used in such a way that it did not express the addressee's but the speaker's role in the dyad, e.g., a father addressing a child as *bābā* 'father'.

4.4.1.3 Fictive kinship terms (FKT + 0)

Fictive kinship terms were forms of address used for addressing someone who was not related to the speaker in one way or another. For example, in *Kasb-e bi- hāsel* [Earn the ungainly], among other plays, the servant called the old woman who had come to visit his master as *mādar* 'mother'. It should, however, be taken into consideration that there seems to be a restriction in the fictive use of KTs. For instance, no fictive instance of *māmān* 'mother', though having the same meaning, was observed.

4.4.1.4 Last name (LN + 0)

Last naming appeared to be common between enlisted men until they became acquainted. Also, in such military contexts, the commanders and other high-ranking officers addressed enlisted men by their last names. It should be pointed out that commanders are those serving in the regular army, not those in militia forces, who, as the data show, usually employed the address term *barādar* 'brother' when addressing less intimate subordinates, and FN for the intimate ones.

   Another context in which last naming occurred quite frequently was in educational institutions. In such situations, teachers regularly addressed their pupils by LN. Classmates also practiced last naming. This usually occurred where the FN was polysyllabic and had no familiar abbreviation whereas the last name contained less syllables than the first name and was thus easier to pronounce, e.g., *Mansure* (FN) Ātaš (LN).

4.4.1.5 Religious titles (RT + 0)

Religious titles in this study included titles such as *hāji* 'Mr. pilgrim [to Mecca]', *hājiye* 'Mrs. Pilgrim [to Mecca]', *seyyed* 'Mr.' [descendant of Prophet Mohammed], *seyyede* 'Mrs.' [descendant of Prophet Mohammad], *mašhadi* 'pilgrim [to Mashhad]', i.e., a religious holy city in Iran respected by Shias, *karbalāyi*

'pilgrim [to Karbala, i.e., a religious holy city in Iraq respected by Shias]', and *šeyx* 'learned man'.

As observed in the data, *hājiye* served as a mode of address. Among acquaintances, this term was used if the addressee was known to have made a pilgrimage to Mecca. Among strangers, however, it was employed as an address term for elderly people in general. There appeared to be an age restriction in the usage of such terms of address because no instance of this title was found to be used for young addressees.

mašhadi and *karbalāyi* appeared less frequently in the data and were normally found in combination with other terms of address, especially FNs. As far as pronunciation is concerned, these terms of address get shortened when combined with FNs (e.g., *Mašt Ali*). Similarly, *seyyed*(*-e*)/ and *šeyx* also did not co-occur with LNs.

The existence of a wide variety and number of religious nominal variants in Persian address system could be indicative of the status of religion in Iran.

4.4.2 Compound forms of address

4.4.2.1 First name combinations

There are certain co-occurrence rules of FNs with other forms of address that can occupy the frames in which such a term occurs. FN typically occupies the second position in compound address instances consisting of two elements. Some examples are KT + FN, e.g., *amu ali* 'Uncle Ali'; GT + FN, e.g., *āqā reza* 'Mr. Reza'; HT + FN, e.g., *jenāb-e arsalān* 'your honor Arsalan'; and RT + FN, e.g., *hājhasan* 'Mr. Pilgrim Hasan'.

In addition to such combinations, FN can also occur as the first element in some other compound terms of address. For example, FN + GT, e.g., *ali āqā* 'Mr. Ali' or *minā xānom* 'Mrs. Mina'' FN + TE (terms of endearment), e.g., *ahmad jān* 'Ahmad dear'; and FN + CT (conventional titles), e.g., *karim xān* 'Karim Chief'.

From among these combinations, KT + FN, IKT + FN, and FN + TE exclusively occurred in familial contexts where one family member addressed the other. More specifically, KT + FN was mostly used by inferiors, whereas IKT + FN and FN + TE were used by superiors to address inferiors (e.g., *amu jān* 'dear uncle').

GT + FN and FN + GT were found to be often exchanged between solidary age mates both in familial contexts and among friends. There seems to be phonological restrictions as to whether a certain FN should follow or precede the GT. For example, as for the male GT, i.e., *āqā* 'Mr.', in the data not one

instance of FN ending with *ā* was found to precede *āqā*, and hence *rezā āqā* 'Mr. Reza' is not a possible combination.

As for combinations such as RT + FN, e.g., *hāj hasan* 'Mr. pilgrim Hasan', it should be pointed out that as the data showed, they were used when there was a higher degree of superiority in terms of age and reverence. Also, the use of this address term was mainly restricted to male characters of the plays and not exchanged as an address term between female interlocutors.

4.4.2.2 Terms of endearment combinations

Terms of endearment are another group of simple terms of address that frequently co-occurred with other address forms. These terms were mostly employed in addressing kids or interlocutors to whom the speaker felt close. Examples are KT/FKT/IKT + TE, e.g., *amu jān* 'dear uncle', and GN + TE, e.g., *farzand-e delband-am* 'my charming baby'.

One of the characteristics of TEs in the data was that they appeared only in the end positions when combined with other forms of address. Moreover, they normally co-occurred with address terms that were used in situations marked with intimacy and familiarity. That is, they usually occurred in familial contexts. In the data in this study, however, there were few instances of TEs co-occurring with LNs. A review of the context in the data revealed that the speaker using such a combination was in a state of anger. Therefore, the semantic of TEs in such instances is not one of intimacy but sarcasm. For example, in *Monji dar Sobh-e Namnāk* [Messiah in the moist morning], one of Mr. Shayan's friends addresses him with LN + TE *šāyān-e aziz* 'dear Shayan" when they have a serious argument; whereas in their ordinary conversations, they exchange GT + LN *āqā-ye šāyān* 'Mr. Shayan', or just FN.

4.5 Marked forms of address

In some cases, the choice of a certain term of address clearly violates a group norm and the customary practice of the speaker. In such cases, the speakers used a marked form of address to express their attitude or emotion or other types of stance toward their interlocutor.

Considering that the focus of this study is on the effects of two semantic dimensions of power and solidarity, these marked address forms, which typically express an emotional or ideological stance, are divided into those that break the norms of power and the ones that break the norms of solidarity.

In deviating from the norms of power in the use of address forms, the speakers may deliberately employ less respectful terms of address to reveal their agency in choosing to show that the addressee is not considered as having a higher power, or in some cases to show solidarity. Norms of solidarity are violated by considering the addressee as an outsider (GT + LN) or as an intimate when the opposite is actually true (FN). For example, in *Monji dar sobh-e Namnāk* [Messiah in the moist morning], Zanganeh who is on GT + LN terms with Shayan, the famous writer, addresses the writer as *vahši* 'wild one' when Shayan is reacting aggressively in their argument. Also, in *Majles-e Divān-e Balx* [Balkh Court House], a character when agreeing with the judge's decision at court addresses him as HT + PT, *hazrat-e qāzi* 'Your honor judge'; however, when the judge's decision is turned against him, he addresses him by FN *abolqāsem*.

As evident in these examples, by employing less respectful terms of address, speakers intentionally break the norms of power and imply that they regard their addressees as inferior or of equal status.

T and V pronouns can also be used to express ideological, attitudinal, and emotional stance like other forms of address. In cases where interlocutors typically use T, using V may show admiration or sarcasm or serve as a distance-building device. Using T where V is socially expected can signal disrespect, anger, or contempt.

In addition to the kinds of expressive usage of address terms discussed so far, some marked forms of address were observed in power-equal settings where the interlocutors were well acquainted. Unlike the former cases of expressive address usage in which the speakers tried to upgrade or degrade the addressees, the implementation of such terms of address by solidary power-equals could be humerous or ironic. For example, in *Āheste bā gol-e sorx* [Slowly with the red flower], Sina addresses his cousin Sanaz with such address terms as *mādmāzel* 'mademoislle', *hazrat-e oljā* 'your excellency [fem.]' and in return, receives address terms such as *profosor* 'professor', *ālijenāb* 'your excellency [male]', and other respectful or very formal titles. Insulting address terms used by equals were observed as well. For example, in *Makhmal* [Velvet], the stepbrother addressed his sister with such address terms as *pir-e zan* 'old woman', *ney-e qelyun* 'skinny', and in return received *čupun* 'shepherd', *aqrab* 'scorpion', and other disrespectful, humiliating forms of address.

4.5.1 Major patterns governing forms of address

In order to find the major patterns of address, Brown and Ford's (1961: 235) "controlled induction" approach was adopted.

On the basis of the effect of social variables of solidarity and power, in each play, four types of settings were considered:

1. Power-laden and solidary
2. Power-laden and not solidary
3. Equal and solidary
4. Equal and not solidary

Then in each setting regarding the characteristic features of superiority, equality, and inferiority, six groups of addressees were specified:

1. Superior and solidary
2. Superior and not solidary
3. Inferior and solidary
4. Inferior and not solidary
5. Equal and solidary
6. Equal and not solidary

It should be mentioned that in the set of plays analyzed in this study, no addressee in the "equal and not solidary" group was found.

Following Brown and Ford's (1961) approach, the next step was to find the norms of address in each group and then match the obtained address norms in opposing groups so as to infer the possible dyadic patterns governing such norms of address. As suggested by Braun (1988), for a certain address variant to be considered normative, the variant should be used with a frequency beyond 75 percent. None of the five groups of address forms in our study reached this frequency.

Each group of address forms had several potential variants that were used by the characters in the plays and therefore the multitudes of variants and the small degree of uniformity in the data led to the lack of a norm of address in each of the five groups. Tables 5–9 in the appendix show the frequency and percentage of each address type for the above groups.

To explore the possibility of arriving at a norm, we combined some of the subcategories into a major category. For example, such address types as PTs (professional titles), RTs (religious titles), OFTs (official titles), HFs (honorific titles), OTs (occupational titles), and NNOTs (occupational terms and their derivations), GTs (general titles), and CTs (conventional titles) were all considered as titles and put in one category (Title). In a similar attempt, address terms such as NNFNs (derivations of first names), NNs (nicknames), and GNs (generic names) were counted as instances of FN (first name). Finally, address terms such as FKTs (fictive kinship terms) and IKTs (inverted kinship terms) were counted as instances of KT in general. After combining subcategories to

a more inclusive category, the address terms, used for the five groups of addressees, were identified (tables 10–14 in the appendix).

Even after combining subcategories of address forms into more inclusive categories, none of the variants, addressed to the five groups of addressees, reached sufficient frequency (75 percent) to be considered a norm. However, it was possible to find predominance of a given variant. Table 4 shows such predominant variants for the five groups of addressees.

Table 4: Predominant address forms

No.	Addressee's characteristics	Predominant terms of address	Percentages
1	Superior and solidary	T + 0	46.62
2	Superior and not solidary	T + 0	57.14
3	Equal and solidary	FN + 0	46.44
4	Inferior and solidary	FN + 0	42.51
5	Inferior and not solidary	FN + 0	52.38

In fact, such predominant forms of address used for each group of addressees only give an idea of what types of forms play a more predominant role in the Persian address system, and what factors possibly affect the choice of a variant. However, as can be seen from our data and the analysis in previous sections, a socially homogeneous group of play characters used a variety of address terms, in which establishing a norm with high frequency of use was not possible. In other words, additional play characters, with the same characteristics, in terms of power and solidarity relations, were potential sources of additional variants and differentiations.

5 Conclusion and Discussion

The results of data analysis showed that Persian address system is marked with variation. The most common of these forms of address were HTs, GNs, PTs, FNs, GTs, and GT + LN (general title + last name). Data analysis also revealed that none of the variant address forms, exchanged between interlocutors with different solidarity and power relations, even after combining subcategories into more general categories, reached the sufficient frequency (75 percent) to be considered a norm. Therefore, we attempted to match the predominant address instances employed by interlocutors in counterpart groups to come up with the most probable address form use. This match produced the two exchanges of (1) reciprocal FN, and (2) nonreciprocal T and FN. Of course, the notion of

reciprocity/nonreciprocity here refers to the high possibility of the exchange of these variant forms among other available terms of address.

One major difference between the Persian and the Western languages address systems discussed by Brown and Gilman (1960) and Brown and Ford (1961) is related to the number of variants in use in Persian language. For example, English address forms, according to Brown and Ford (1961), like other Western languages, is described by a single binary contrast: FN or TLN. That is, the principle option of address in English is the choice between use of the first name and use of a title with the last name. In contrast, Persian speakers have quite a number of possible address variants to choose from when they address a certain addressee. Even after these variants were summarized into basic category types, there still remained several nominal variants for the speaker to use for one and the same addressee.

As explained by Braun (1988), the multitude of variants is evident as one turns away from European or Western societies. In non-Western communities, due to their greater social diversity, one will find variation in the use of terms of address according to factors like regional dialect, class, education, age, gender, ideology, religion, etc. Conversely, in Western communities, due to the spread of standardized varieties by mass media and other institutions and a more egalitarian ideology, the inventory of address variants is restricted, not allowing much variation or modification.

In the light of the above discussion, one notices that the rules Brown and Ford (1961) propose are more applicable to the address system of Western languages, which include normative forms of address, and do not apply as much to the Persian language. In Persian several factors govern the selection of a certain form of address, which decreases the possibility of a norm because one variant will not easily outnumber the others.

5.1 Notion of (non)reciprocity

Brown and Ford (1961) view reciprocity of address in a dyadic system, i.e., the exchange of the same variants as a signal of equality, real or pretended. Nonreciprocity has been interpreted as an expression of difference in age or professional status. Brown and Ford's (1961) interpretation of reciprocity, however, cannot be applied to a language like Persian, which is characterized by the "inversion" phenomenon. As previously explained, inversion refers to the use of a nominal variant which, in its lexical content, implies features related to the speaker rather than the addressee. Inversion consists of the reciprocation of a term that mostly expresses seniority over superiority. In such cases, the inferior speaker of the dyad, e.g., one who is inferior in age, status, or both,

uses address variants that express the other's superior position, and the superior responds with the same or similar term, inappropriate as it may seem. For example, once, in Persian, the reciprocation of a term like *amu* 'uncle' is observed between uncles and nephews or nieces, one can by no means, conclude, as Brown and Ford (1961) do, the equality of the interlocutors.

The reciprocal usage of this kind, as Braun (1988) maintains, obviously is not covered by the evaluation of reciprocity suggested by Brown and Ford (1961), and it has to be accepted that in address inversion, the reciprocation of address terms can occur as an expression of inequality.

5.2 Semantic factors involved

As for the semantic factors affecting the interlocutors' address usage, Brown and Ford (1961) make the generalization that there are two factors that determine the (non)reciprocation of terms of address: age and occupational status. They maintain that differences in the two or either dimension alone are able to generate the nonreciprocal pattern. Also, they state that there is a normative rule of priority of occupation over age in the determination of deference.

The data in this study suggest that Brown and Ford's (1961) interpretation of (non)reciprocity cannot be used to explain all instances of reciprocal and nonreciprocal exchanges of terms of address in Persian. There are instances in the data in this study, which reveal the fact that the factors governing Persian address behavior are highly varied that it is hard to fit them into a general theoretical frame. That is, not all of them can easily be traced back to the more abstract notions of superiority/inferiority, distance/intimacy, etc. The variety of existing forms of address in Persian and their dynamic and strategic use by the language users prevent a simple and straightforward mapping with well-known general parameters such as age, status, or both.

Our findings indicate that there are instances of address forms in the Persian address system that do not support Brown and Ford's (1961) universalistic views about address system in different languages. The Persian address system, as discovered within the scope of this study, is characterized by diversity, rarity of norm, and larger number of address forms compared to the English address system. In fact it can be said that the rules governing the Persian address system, unlike those of Western languages, are more complicated and function within a more extensively diversified domain. Therefore, it seems that while dealing with a language like Persian, variability is more widespread in the address form system than consistency. The choice of address term among the extensive range of options seems to be affected by variables such as age, gender, personality, social status, social distance, religious and political orientations, family relationships, and formality.

References

Archibald Afful, J. B. 2010. Address forms among university students in Ghana: A case of gendered identities? *Journal of Multilingual and Multicultural Development* 31 (5). 443–456.

Beeman, W. O. 1986. *Language, status, and power in Iran*. Bloomington: Indiana State University.

Braun, F. 1988. *Terms of address*. Berlin: Mouton de Gruyter.

Brown, R. & M. Ford. 1961. Address in American English. *Journal of Abnormal and Social Psychology* 62 (2). 375–385.

Brown, R. & A. Gilman. 1960. The pronoun of power and solidarity. In T. A. Sebeok (ed.), *Style in language*, 253–276. Cambridge: MIT Press.

Clyne, M., H. Kretzenbacher, C. Norrby & J. Warren. 2003. Address in Some Western European Languages. *Proceedings of the 2003 Conference of the Australian Linguistic Society*. 1–10

Culpeper, J. 2010. Conventionalized impoliteness formulae. *Journal of Pragmatics* 42 (12). 3232–3245.

Dittrich, W. H., T. Johansen, & E. Kulinskaya. 2011. Norms and situational rules of address in English and Norwegian speakers. *Journal of Pragmatics* 43 (15). 3807–3821.

Eslami, Z. R. & A. Eslami-Rasekh. 2007. Use of address forms by Iranian Moslem women and the effect of religion. In S. Joseph (ed.), *The encyclopedia of women and Islamic cultures*, vol. 5: 234–235). New York: Brill Koninklijke Brill NV.

Eslami, Z. R. & A. Mirzaei. 2012. Assessment of second language pragmatics. In C. Coombe, P. Davidson, S. Stoynoff & B. O-Sullivan (eds.), *The Cambridge guide to language assessment*, 198–209. New York: Cambridge University Press.

Eslami, Z. R. & A. Mirzaei. 2014. Speech act data collection measures in a non-Western context: Oral and written DCTs in the Persian language. *Iranian Journal of Language Testing* 4 (1), 137–154.

Hymes, D. 1972. Models of the interaction of language and social life. In J. J. Gumperz & D. Hymes (eds.), *Directions in sociolinguistics: The ethnography of communication*, 35–71. New York: Holt, Rinehart and Winston.

Kendall, M. B. 1981. Toward a semantic approach to terms of address: A critique of deterministic models in sociolinguistics. *Language and Communication* 1: 237–254.

Keshavarz, M. H. 1988. Forms of address in post-revolutionary Iranian Persian: A sociolinguistic analysis. *Language in Society* 17 (4). 565–575.

Keshavarz, M. H. 2001. The role of social context, intimacy, and distance in the choice of forms of address. *Language and Society* 17 (2). 6–18.

Koven, M. 2009. Managing relationships and identities through forms of address: What French-Portuguese bilinguals call their parents in each language. *Language and Communication* 29 (4). 343–365.

Labov, W. 1966. *The social stratification of English in New York*. Washington, D.C.: Center for Applied Linguistics.

Liu, Y. 2009. Determinants of stall-holders' address forms to customers in Beijing's low-status clothing markets. *Journal of Pragmatics* 41 (3). 638–648.

Mahnmoodan, A. 2013. *The semantics and pragmatics of address forms in Persian*. Trondheim: Norwegian University of Science and Technology MA thesis in linguistics.

Moreno, M. C. 2002. The address system in the Spanish of the golden age. *Journal of Pragmatics* 34 (1). 15–47.

Moyna, M. I. & B. Ceballos. 2008. Representaciones dramáticas de una variable lingüística: Tuteo y voseo en obras de teatro del Río de la Plata (1886–1911) [Dramatic Representation of a Linguistic Variable: tuteo and plays voseo in Rio de la Plata]. *Spanish in Context* 5 (1). 64–88.

Nanbakhsh, G. 2011. *Persian address pronouns and politeness in interaction*. Edinburgh: University of Edinburgh dissertation of linguistics and English language.

Philipsen, G. & M. Huspek. 1985. A bibliography of sociolinguistic studies of personal address. *Anthropological Linguistics* 27 (1). 94–101.

Sharifian, F. 2009. On collective cognition and language. In Hanna Pishwa (ed.), *Language and social cognition – Expression of social cognition*, 163–183. Berlin & New York: Mouton de Gruyter.

Zhang, X. 2011. A comparative study of the Sino-American address forms from an intercultural communication perspective. *International Journal of English Linguistics* 1 (1). 4–61.

Appendix

Table 5: Frequency/percentage of terms of address used for superior and solidary addressees

No.	Address terms	Frequency	Percentage
1	PT + 0	18	10.11
2	RT + 0	2	1.12
3	OFT + 0	6	3.37
4	HT + 0	41	23.03
5	OT + 0	2	1.12
6	GT + 0	12	6.74
7	GN + 0	3	1.68
8	FN + 0	2	1.12
9	TA + 0	1	0.56
10	NN + 0	1	0.56
11	KT + 0	12	6.74
12	FKT + 0	2	1.12
13	NNOT + 0	2	1.12
14	GT + TE	4	2.24
15	GT + LN	13	7.30
16	GT + PT	8	4.49
17	GT + RT	1	0.56
18	GT + FN	1	0.56
19	HT + OT	3	1.68
20	HT + OFT	7	3.92
21	HT+ PT	3	1.68
22	HT + LN	3	1.68
23	RT + FN	1	0.56
24	RT + GT	4	2.24

25	PT + LN	1	0.56
26	CT + KT	1	0.56
27	FN + CT	2	1.12
28	FN + TE	2	1.12
29	FN + GT	2	1.12
30	KT + FN	4	2.24
31	KT + TE	2	1.12
32	FKT + FN	5	2.80
33	FKT + RT	1	0.56
34	FKT + TE	4	2.24
35	TA + TE	1	0.56
36	GT + KT + TE	1	0.56
Total		178	100

Table 6: Frequency/percentage of terms of address used for superior and not solidary addressees

No.	Address terms	Frequency	Percentage
1	PT + 0	2	4.08
2	HT + 0	16	32.65
3	OT + 0	3	6.12
4	GT + 0	7	14.28
5	GN + 0	1	2.04
6	FKT + 0	9	18.36
7	GT + LN	2	4.08
8	GT + PT	1	2.04
9	HT + PT	3	6.10
10	RT + GT	1	2.04
11	FN + CT	1	2.04
12	FKT + TE	3	6.10
Total		49	100

Table 7: Frequency/percentage of terms of address used for equal and solidary addressees

No.	Address terms	Frequency	Percentage
1	PT + 0	4	3.20
2	OT + 0	2	1.60
3	GT + 0	4	3.20
4	GN + 0	15	12
5	FN + 0	30	23
6	LN + 0	4	3.20
7	NN + 0	13	10.4
8	TE + 0	5	4

9	FKT + 0	8	6.40
10	NNOT + 0	2	1.60
11	GT + TE	2	1.60
12	GT + LN	20	16
13	GT + PT	2	1.60
14	GT + OFT	2	1.60
15	GT + FN	1	0.80
16	GT + OT	1	0.80
17	GT + GA	1	0.80
18	HT + GT	2	1.60
19	HT + FN	1	0.80
20	RT + GT	1	0.80
21	FN + CT	1	0.80
22	FN + TE	2	1.60
23	GN + TE	1	0.80
24	HT + GN + TE	1	0.80
Total		125	100

Table 8: Frequency/percentage of terms of address used for inferior and solidary addressees

No.	Address terms	Frequency	Percentage
1	PT + 0	6	4.72
2	RT + 0	1	0.78
3	HT + 0	2	1.57
4	OT + 0	3	2.36
5	GT + 0	4	3.14
6	GN + 0	19	14.96
7	FN + 0	20	15.70
8	LN + 0	5	3.93
9	NN + 0	7	5.51
10	KT + 0	1	0.78
11	TE + 0	1	0.78
12	FKT + 0	4	3.14
13	NNFN + 0	8	6.28
14	GT + TE	1	0.78
15	GT + LN	13	10.23
16	GT + PT	2	1.57
17	GT + FN	2	1.57
18	HT + OFT	1	0.78
19	HT + PT	1	0.78
20	RT + FN	1	0.78
21	FN + CT	5	3.93
22	FN + TE	3	2.36
23	KT + TE	2	1.57

24	FKT + TE	1	0.78
25	IKT + FN	1	0.78
26	IKT + TE	4	3.14
27	GN + TE	5	3.93
28	LN + TE	1	0.78
29	TR + TE	1	0.78
30	HT + FN + CT	1	0.78
31	GT + FN + CT + LN	1	0.78
Total		127	100

Table 9: Frequency/percentage of terms of address used for inferior and not solidary addressees

No.	Address terms	Frequency	Percentage
1	GT + 0	2	9.52
2	Gn + 0	11	52.38
3	FKT + 0	3	14.24
4	IKT + 0	1	4.76
5	GT + GN	1	4.76
6	RT + GT	1	4.76
7	GN + TE	2	9.52
Total		21	100

Table 10: Frequency/percentage of terms of address used for superior and solidary addressees

No.	Address terms	Frequency	Percentage
1	T + 0	83	46.62
2	FN + 0	6	3.37
3	TA + 0	1	0.56
4	KT + 0	14	7.86
5	T + TE	4	2.24
6	T + LN	17	9.54
7	T + T	26	14.60
8	T + FN	2	1.12
9	T + KT	1	0.56
10	FN + T	4	2.24
11	FN + TE	2	1.12
12	KT + FN	9	5.05
13	KT + TE	6	3.37
14	KT + T	1	0.56
15	TA + TE	1	0.56
16	T + KT + TE	1	0.56
Total		178	100

Table 11: Frequency/percentage of terms of address used for superior and not solidary addressees

No.	Address terms	Frequency	Percentage
1	T + 0	28	57.14
2	FN + 0	1	2.04
3	KT + 0	9	18.36
4	T + LN	2	4.08
5	T + T	5	10.20
6	FN + T	1	2.04
7	KT + TE	3	6.10
Total		49	100

Hooman Saeli and Corey Miller

9 Some linguistic indicators of sociocultural formality in Persian

Abstract: Similar to many other languages, speakers of Persian utilize their language to reflect societal norms, such as opposite-gender dynamics (Mahdavi 2007), in conversations. Given the relative dearth of evidence, this study was our attempt to investigate how speakers of Persian incorporate linguistic elements to varying degrees (e.g., plural versus singular pronouns and morphemes and colloquial versus standard pronunciation and morphology) to converge to/diverge from societal norms. In particular, we hoped to unveil how speakers of Persian use their pragmalinguistic repertoire to enact societal norms in intergender exchanges. Our data comprised discourse completion task (DCT) responses from 27 Persian-speaking participants. Analysis revealed that the participants utilized a variety of linguistic variables to achieve a number of important communicative purposes in different contexts. For instance, the participants generally used singular pronouns/inflectional morphemes to minimize societal, intergender gaps. Our findings help shed light on the intricate interplay between language use and societal norms, and how language is used as a vehicle to disseminate social and personal values.

Keywords: Persian, pragmatics, politeness, colloquial, diglossia, phonetics

1 Introduction

Among the principal constituents of language is the ability of its speakers to utilize their linguistic repertoire in accordance with the sociocultural norms of a given language community. Widdowson (1998) illustrates that authentic language use occurs in natural communication in discourse communities. A number of crucial aspects of authentic language use are speakers' knowledge of what to say, how to say, and when to say. Koike (1989: 279) identifies pragmatic competence as an underlying component of competent language use and defines it as "the speaker's knowledge and use of rules of appropriateness and politeness which dictate the way the speaker will understand and formulate speech acts".

Hooman Saeli, University of Tennessee Knoxville
Corey Miller, MITRE Corporation

Koike's (1989) definition of pragmatic competence highlights the role of politeness in language use. Fraser (1990) touches upon the elusive nature of politeness and states that the type of language one uses (e.g., formal) and her or his degree of politeness are interrelated. Culpeper, Bousfield, and Wichmann (2003), in a similar vein, state that politeness is indeed difficult to define. In particular, they conclude that politeness is a phenomenon to be studied with a context-specific approach.

Languages provide different tools by which speakers control the degree of (im)politeness in their speech. Whereas in some languages a number of pragmatic strategies, such as indirectness, have been reported to make one's speech more polite (e.g., Blum-Kulka 1987), other languages wield a wide variety of linguistic tools through which one's speech can become more polite. Such tools may be lexical (e.g., colloquial vs. standard vocabulary), syntactic (e.g., word order, pro-drop), or morphological (e.g., use of inflectional morphemes). The focus of this article is to examine how these categories may be manipulated in Persian in order to affect the relative politeness of utterances.

1.1 Politeness norms in Persian academic contexts

As Fraser (1990) concludes, the concept of politeness is construed in a variety of ways based upon the context in which it occurs. Iran, as Koutlaki (2010) points out, is a country in which people of opposite genders are segregated in a number of contexts (e.g., personal and academic). Specifically, within the Iranian school context (i.e., pre-college), women and men attend different schools; therefore, in academic settings, the first "formal" encounter of women and men is at the college level. This, as Mahdavi (2007) also attests, contributes to the formality of the women-men dynamics in an Iranian context. It is, thus, safe to conclude that women and men formally (and legally) come together for the first time after entering college. As Agha (2003) and Coupland and Jaworski (2004) suggest, for instance, social and cultural ideologies are likely to surface in their language. Consequently, sociocultural norms and expectations, such as social distance between women and men, are also likely to materialize in language. Therefore, we can hypothesize that the Persian sociocultural norms that govern the women-men dynamics also surface in language exchanges.

1.2 Favor-asking as a face-threatening act

Brown and Levinson (1987) postulate that, within social interactions, interlocutors strive to save face by not appearing imposing. Goldschmidt (1998) states

that the speech act of favor-asking is by nature a face-threatening act. In particular, she shows that asking for favors in American English places an out-of-ordinary burden on hearers. Interlocutors would, therefore, attempt to minimize the face-threatening nature of asking for favors by using a variety of communicative strategies. Goldschmidt (1998) and Saeli (2016) report that speakers usually incorporate pre-favor strategies (e.g., problem statement and solution proposition) to ease such psychological burdens. This brief discussion qualifies favor-asking as an appropriate tool by which politeness norms in a Persian context can be examined.

1.3 Persian second person reference and inflectional morphemes

In Persian, second person interlocutors can be addressed by names, kinship terms, familiarizers, pronouns, imposters, or nothing. Baumgardner (1982) presents a table, here simplified, based on Obolensky et al. (1973) that summarizes the use of first names (FN) and last names (LN) in increasing order of deference:

Table 1: Use of names in increasing order of deference

Male	Female
FN	FN
FN + جان /ʤɒn, ʤun (Q[1])/ 'dear'	FN + جان /ʤɒn, ʤun/ 'dear'
FN + آقا /ɒɣɒ/ 'Mr.', آقا /ɒɣɒ/ 'Mr.' + FN	FN + خانم /xɒnom, xɒnum (Q)/ 'Ms.'
آقا /ɒɣɒ/ 'Mr.' + LN	خانم /xɒnom, xɒnum (Q)/ 'Ms.' + LN

Persian, like many other languages, allows the use of kinship terms, such as داداش /dɒdɒʃ/ 'brother' as a familiar form of address. Other vocatives, such as آقا /ɒɣɒ/ 'Mr.' without following or preceding name, may be also be used like English *man* as 'familiarizers' (Luckmann de Lopez 2013).

As a substitute for the above, the Persian language has two second person deictic references: تو /to/ and شما /ʃomɒ/, along with their corresponding enclitics ت- /æt, et (Q)/ and تان- /-etɒn, -etun (Q)/, which may be used to indicate possession or direct objects. While the former pair (تو, ت-) are singular (T in the terminology of Brown and Gilman 1960), the latter pair (شما, تان-) are plural (V in the terminology of Brown and Gilman 1960). As in many languages, the plural or V form in Persian may be used to refer to a single person, and when so

---

1 Colloquial forms marked with Q, see below.

used, connotes a higher degree of formality (Keshavarz 1988). These pronouns have verbal inflectional counterparts in order to satisfy Persian's subject-verb agreement; however, V pronominals may be combined with T verb forms (Baumgardner 1982 "mixed forms"; Nanbakhsh 2012 "mismatch construction"). In this article, we take a pragmatic and sociocultural approach to analyzing the use of different pronominals and verbal forms in a number of pragmatically laden contexts. In the following, we provide a list of sociocultural interpretations of pronominals and verb forms:

- Singular pronominal and singular verb: minimal perceived social distance; indication of intimacy (e.g., between friends)
- Plural pronominal and singular verb: some perceived social distance; indication of acquaintance, but not intimacy (e.g., between coworkers). Baumgardner (1982) notes that forms with a singular pronominal and plural verb are used only jokingly or sarcastically.
- Plural pronominal and plural verb: vast perceived social distance (e.g., between strangers, or professors and students)

Another aspect of Persian pronoun use we will explore is known as pro-drop, the phenomenon by which certain languages may optionally eschew the use of pronouns in certain situations where they would be required in other languages. Contrast Persian میروم (من)/(mæn) mirævæm/ '(I) go' with English 'I go'; whereas the Persian pronoun من /mæn/ 'I' is optional, the English pronoun "I" is obligatory. While some pro-drop languages, such as Spanish, benefit from a wide literature quantitatively and sociolinguistically exploring the use of different pronouns in different situations (e.g., Pešková 2013), Persian has only a small number of studies treating this phenomenon in any detail. Azizi (2011) examined variable interpretations by speakers of a number of sentences lacking an explicit pronoun, while Dalili (2009) examined obligatory and optional use of pronouns from a theoretical syntactic viewpoint.

The relevance of pro-drop to a study of formality and politeness is twofold. On the one hand, one may ask whether omission of pronouns renders sentences more casual; on the other hand, one may ask whether absence of a pronoun is a strategy by which speakers avoid the choice between a formal and familiar pronoun. Stemming from this is the phenomenon of imposters (Collins and Postal 2012), or Super-V (Levinson 1977) – third person non-pronominals such as "my lord", which "hide" a first or second person core. In Persian, such imposters may be honorific, such as جناب عالی /dʒenɒbe ʊli/ 'exalted sir' (for second person) or humbling such as چاکر /tʃɒker/ 'servant' (for first person), both of which connote higher deference (e.g., than use of either T or V forms) to one's interlocutor (Baumgardner 1982).

1.4 Standard vs. colloquial language

As is well known, there is a wide gulf between spoken and written Persian. While some have viewed this situation as an example of diglossia (Jeremiás 1984), others have found the differences between the two registers not as extreme as in languages like Arabic (Perry 2003). In the current age where internet communication blurs the line between spoken and written varieties (Crystal 2006), we find it more useful to characterize the L (Low) forms "colloquial" and the H (High) forms "standard". While colloquial forms are used by all speakers of Persian, regardless of social class, in certain situations, it is reasonable to explore their relative use in similar situations by different speakers. Thus, in this study, we will examine the use of variables having colloquial and standard counterparts, such as the following:

– colloquial 2PL verbal inflectional suffix /-in/ vs. standard /-id/
– colloquial /vɒse/ vs. standard /bærɒje/ 'for'
– colloquial verb stems such as /ɒr/ vs. /ɒvær/ 'bring'
– colloquial /un/ vs. standard /ɒn/, e.g., in 'that'
– colloquial use of pronominal enclitics as objects, e.g., دیدمت /didæmet/ vs. ترا دیدم /torɒ didæm/ 'I saw you'
– use of suffixed definite article /-ˈ(h)e/, e.g., /mæɣɒleˈhe/ 'the article'

2 The present study

In this study, we shed light on the use of second person reference, morphological, pronominal, and register variation that could, purportedly, manipulate the extent to which favors were asked "politely" within a Persian academic context. The following research questions are addressed:

– Do second person reference, morphological, pronominal, and register modifications contribute to generating more or less "polite" favors?
– If so, what are some ways in which norms of "politeness" are (dis)obeyed? Why?

2.1 Methodology

In this section, we provide relevant information on the participants, study context, data collection, and data analysis methods.

2.2 Participants

Twenty-seven native speakers of Persian living in both the United States and Iran were invited to participate in the study. The participants were university students in a variety of academic majors in both countries. We need to note that the participants were all graduate students because we hypothesized that female and male students came into contact with each other in such contexts. The participants willingly took part in the study, and their ages ranged from twenty-one to thirty-two. Also, care was taken to select a relatively even gender balance – thirteen women and fourteen men – to facilitate a better understanding of inter-gender social and linguistic interactions of women and men.

2.3 Instrument

The data used in this study were part of a larger project conducted in order to investigate the impact of gender and academic status differences on favors asked within formal and informal contexts. We hypothesized that favor-asking was a speech act that was imposing in nature, and therefore would require interlocutors to use politeness strategies to save face. As a result, recorded oral discourse completion tasks (DCTs) were employed to elicit favors in a typical academic context.

In order to make the final interpretations more feasible, the variable of gender was the only manipulated variable, and the DCT scenario revolved around one principal theme, as elaborated below. We attempted to utilize a scenario that was very likely to happen in academic contexts in order to increase the authenticity of the data: prompting the participants to ask for an article from a peer (i.e., university classmate). Before collecting the data, we stipulated that the participants imagine they had known the imaginary classmate for only *one* academic semester (i.e., four and a half months). We also added that the hypothetical addressee was a classmate, *not* a friend.

Also, in order to examine the role of gender differences and their effect on the elicited favors, the respondents were asked to complete two DCTs for the scenario. That is, each respondent provided two responses: one response for a same-gender classmate and one for an opposite-gender classmate. The following is the scenario used:

فرض کنید شما به یک مقاله مهم دسترسی ندارید، ولی به آن برای انجام یک پروژه دانشگاهی نیاز دارید. همکلاسیِ شما در دانشگاه این امکان را دارد که آن مقاله را بیابد (بطور مثال از طریق دانلودِ آن از یک وبسایت). لطفا از او بخواهید مقاله را برای شما بیابد.

لطفا توجه کنید که شما باید ۲ مرتبه این درخواست را ارائه دهید: مرتبه اول از یک همکلاسی از جنس موافق، و مرتبه دوم از یک همکلاسی از جنس مخالف.

با تشکر!

[Imagine you do not have access to an important article, but need it for a university project. A classmate of yours has the ability to find the article for you (e.g., through downloading it from a website). Please ask her/him to find the article for you.

Please note that you need to ask this favor twice: once from a same-gender classmate, and once from an opposite-gender classmate.

Thanks!]

2.4 Procedure

The participants were invited to a quiet public coffee shop in order to read the scenarios and record their responses. The initial briefing took three to five minutes. The Oklahoma State University's Institutional Review Board (IRB) documents were also provided to the participants. After it was confirmed that the participants had a clear idea as to the purpose of the study, they were instructed on recording their responses on a personal digital device. The recordings took ten to fifteen minutes. The participants were left alone so they would have maximum mental comfort in responding to the scenarios. After the recordings were completed, they were listened to once more to ensure that no technical failures occurred and that the recordings were of sufficient quality.

2.5 Data analysis

A total of fifty-four favors were elicited from the participants: twenty-seven same-gender favors and twenty-seven opposite-gender ones. As part of the larger project, the favors underwent content analysis (e.g., open- and axial-coding of data), and the main constituents of the favors were identified. For the purpose of this study, however, we report only on the interrelated nature of second person address, morphology, pronominal use, and register with (im)politeness of the favors. In doing so, our analyses were targeted at the following features:

–   the use of names, kinship terms, familiarizers, imposters, and pro-drop in same-gender vs. opposite-gender favors
–   the use of singular vs. plural pronouns and enclitics in same-gender vs. opposite-gender favors
–   the use of singular vs. inflectional morpheme on verbs in same-gender vs. opposite-gender favors
–   the use of colloquial forms in same-gender vs. opposite gender favors

The above features were particularly scrutinized to shed light on the differences associated with gender.

2.6 Findings

In the following, we provide excerpts from the DCT responses, along with their English translations and interlinear glosses. Each morpheme or word deemed markedly colloquial is indicated with a "Q". Where other abbreviations go beyond the material covered in the Leipzig Glossing Rules (Comrie et al. 2015), we provide an explanation at the first instance. Following each excerpt, we tally the following:
–   Form of address: Name type from Table 1, familiarizer, kinship term, T pronoun, V pronoun, honorific/humbling imposter, pro-drop (in main clause)
–   Second person verb: singular, plural /-id/, colloquial plural /-in/ (number of each type occurring in excerpt follows type in parentheses)
–   Mixed forms: Presence (1) or absence (0) of mixed forms/mismatch construction (V pronoun + singular verb)
–   Number of colloquial morphemes

We first present the findings obtained from the analysis of the favors asked by the female participants (i.e., the ones asked from women and the ones asked from men). We then move on to the favors elicited from the male participants (i.e., the ones asked from women and the ones asked from men).

2.6.1 Women's favors

2.6.1.1 Same-gender favors
Analysis of the favors women asked from women brought to light that, in all cases, singular deictic references and inflectional morphemes had been utilized. This finding was not surprising; considering the patterns of sociocultural

interactions in Iran, it appears that women are likely to feel more comfortable with same-gender interlocutors. This, therefore, materialized in the asked favors. Also, we looked at instances of pro-drop and the use of honorific and humbling imposters. In the following examples, we delve into the variables that contributed to the (in)formality of the favors.

(1) سلام. چطوری؟

salɒm   ʧetowr-i
hello   how-COP.2SG
'Hello, how are you?'

Form of address	Verb	Mixed forms	Colloquial forms
pro-drop	singular (1)	0	0

(2) میگم این مقاله رو داری تو؟

mi-g-æm   in   mæɣɒle-ro   dɒr.i   to
DUR-say.Q-1SG   this   article-OM².Q   have-2SG   you.2SG
'(I say) do you have this article?'

Form of address	Verb	Mixed forms	Colloquial forms
T, non-canonical word order with respect to verb, used contrastively (e.g., 'I don't have it, do you?')	singular (1)	0	2

(3) اگه داری ممکنه واسم پرینت بگیری فردا بیاری؟

æge   dɒr-i   momken-e   vɒse-m   perint   be-gir-i
if.Q   have-2SG   possible-COP.3SG.Q   for.Q-me   print   SBJV-take-2SG
færdɒ   bi-ɒr-i
tomorrow   SBJV-bring.Q-2SG
'If you have it, is it possible to print it and bring it for me tomorrow?'

Form of address	Verb	Mixed forms	Colloquial forms
pro-drop	singular (3)	0	4

---

2 OM = Object Marker

(4)   امکانش هست واسم دانلودش کنی؟

    emkɒn-eʃ   *hæst*   *vɒse-m*   *dɒnlod-eʃ*   *kon-i*

    possible-3SG.Q   is   for.Q-1SG   download-OM.Q   do-2SG

    'Is it possible for you to download it for me?'

Form of address	Verb	Mixed forms	Colloquial forms
pro-drop	singular (1)	0	3

As shown in examples (3) and (4), the women participants tended to ask more indirect favors even though they used singular pronouns. The use of singular pronouns indicates that they felt more socioculturally comfortable around their same-gender peers; nonetheless, we can conclude that the incorporation of indirect favors is a token of their attempt to save face.

2.6.1.2 Opposite-gender favors

The opposite-gender favors the women participants asked differed greatly from those asked of same-gender people. In particular, the instances of singular deictic references and singular inflectional morphemes were only a few. In addition, while one case of singular pronoun and inflectional morpheme use was identified, two favors included a plural pronoun, but a singular inflectional morpheme. In the following excerpts, we show how the gender of the addressee can lead to the higher formality of the favors.

(5)   سلام آقای فلانی.

    sælɒm   *ɒɣɒ-je*   *felɒn-i*

    hello   Mr.-EZ   someone-INDF

    'Hello, Mr. So-and-so'.

Form of address	Verb	Mixed forms	Colloquial forms
آقا + FN/LN	NA	0	0

(6)   احوال شما؟

    æhvɒl-e   *ʃomɒ*

    conditions-EZ[3]   you.2PL

    'How are you?'

Form of address	Verb	Mixed forms	Colloquial forms
V	NA	0	0

---

3 Ezafe (اضافه)

(7)   ببخشید، شنیدم این مقاله رو شما می‌تونین پیدا کنین.

be-bæʃ-id   *ʃenid-æm* *in*   *mæɣɒle-ro*   *ʃomɒ*   *mi-tun-in*

SBJV-excuse-2PL heard-I   this article-OM.Q you.2PL DUR-can-2PL.Q

pejdɒ   *kon-in*

find   do-2PL.Q

'Pardon, I heard that you can find this article'.

Form of address	Verb	Mixed forms	Colloquial forms
V	plural (1), colloquial plural (2)	0	3

(8)   می شه لطف کنین و واسم پرینتش کنین هرموقع تونستین؟ مرسی.

mi-sh-e   *lotf*   *kon-in*   *o*   *vɒse-m*   *perint-eʃ*

DUR-become-3SG.Q please do-2PL.Q and for.Q-me print-3SG.Q

kon-in   *har*   *mowɣe* *tun-est-in*   *mersi*

do-2PL.Q any time   can-PST-2PL.Q thanks

'Is it possible you could do me a favor and print this for me any time you could?'

Form of address	Verb	Mixed forms	Colloquial forms
pro-drop	colloquial plural (3)	0	5

(9)   اگه بتونین برام گیرش بیارین ممنون میشم.

æge *be-tun-in*   *bærɒ-m* *gir-eʃ*   *be-ɒr-in*

if.Q SBJV-can-2PL.Q for-1SG find.Q-OM.Q SBJV-bring.Q-2PL.Q

mæmnun   *mi-ʃ-æm*

thanks   DUR-become.Q-1SG

'If you can find it for me, I would appreciate it'.

Form of address	Verb	Mixed forms	Colloquial forms
pro-drop	colloquial plural (2)	0	4

As we showed in the above excerpts, the women participants revealed a great deal of formality when interacting with their opposite-gender peers. This is apparent in their use of plural pronouns and plural verbs.

(10)   مریم میگفت شما مقاله بیو ژنتیک رو داری.

mærjæm *mi-goft*   *ʃomɒ*   *mæɣɒle-je* *bioʒenetik-ro.Q* *dɒr-i*

Maryam DUR-say.PST you.2PL article-EZ biogenetic-OM.Q have-2SG

'Maryam told me that you have this biogenetic article'.

Form of address	Verb	Mixed forms	Colloquial forms
V	singular (1)	1	1

In example (10), we found that a plural pronoun is used in conjunction with a singular verb, what has been called a "mixed form" or "mismatch construction". Whereas the use of the singular verb signals higher sociocultural comfort, the incorporation of the plural pronoun still functions as a creator of social distance. Such a distance, however, might not be as large as the ones depicted in examples (7) to (9), which employ plural verbs.

(11) لطف میکنی واسم پیداش کنی؟

lotf	*mi-kon-i*	*vɒse-m*	*pejdɒ-ʃ*	*kon-i*
please	DUR-do-2PL	for.Q-1SG	find-3SG.Q	do-2SG

'Can you please find it for me?'

Form of address	Verb	Mixed forms	Colloquial forms
pro-drop	singular (2)	0	2

(12) چطوری حسین؟

tʃetowr-i	*hosejn*
how-2SG	Hossein

'What's up, Hossein?'

Form of address	Verb	Mixed forms	Colloquial forms
FN	singular (1)	0	0

(13) ببین، من این مقاله رو باید پیداش کنم.

be-bin	*mæn*	*in*	*mæɣɒle-ro*	*bɒjæd*	*pejdɒ-ʃ*	*kon-æm*
SBJV-see.2SG	I	this	article-OM.Q	must	find-3SG.Q	do-1SG

'Look, I have to find this article'.

Form of address	Verb	Mixed forms	Colloquial forms
pro-drop	singular (1)	0	2

(14) میتونی کمکم کنی؟ ممنونم.

mi-tun-i	*komæk-æm*	*kon-i*	*mæmnun*
DUR-can-2SG	help-1SG.Q	do-2SG	thanks

'Can you help me? Thanks'.

Form of address	Verb	Mixed forms	Colloquial forms
pro-drop	singular (2)	0	0

Examples (11) to (14) indicate a higher degree of informality the women participants reported with opposite-gender peers, given the use of singular verbs. We can nevertheless conclude that the social barriers and norms, to some extent,

govern the interactions between people of different genders. Through the afore-mentioned favors, we can see how the use of more formal linguistic properties helps reflect the societal norms in an Iranian context.

2.6.2 Men's favors

2.6.2.1 Same-gender favors

As expected, all of the same-gender favors asked by the men participants included singular deictic references and singular inflectional morphemes. This was an indication of fewer social barriers, if any, between two male classmates. As we show in the following examples, the use of singular pronouns, singular verbs, and humbling pronoun replacements illustrate the great deal of socio-cultural comfort among men in an Iranian academic context.

(15) هومن ردیفی؟

humæn   rædif-i
Hooman   row-COP.2SG
'How you doing, Hooman? (Are you in order?)'

Form of address	Verb	Mixed forms	Colloquial forms
FN	singular	0	0

(16) آقا میگم این مقاله هه که استاد میگفت و میتونی گیر بیاری واس من؟

ɒyɒ   mi-g-æm   in   mæɣɒle-he   ke   ostɒd   mi-goft
Mr.   DUR-say-1SG   this   article-ART.DEF.Q   that   professor   DUR-say.PST.3SG

o   mi-tun-i   gir   bi-ɒr-i   ʋɒs-e   mæn
and   DUR-can-2SG   find.Q   SBJV-bring.Q-2SG   for-EZ   me

'Dude, I say this article that the professor was talking about, can you find it for me?'

Form of address	Verb	Mixed forms	Colloquial forms
آقا (familiarizer)	singular	0	2

(17) دمت گرم آقا!

dæm-et   gærm   ɒyɒ
breath-2SG.Q   warm   man
'May your breath be warm!' (Thank you)

Form of address	Verb	Mixed forms	Colloquial forms
آقا (familiarizer)	NA	0	1

(18) خوبی داداش؟

xub-i   dɒdɒʃ
good-COP.2SG   brother
'How you doing, bro?'

Form of address	Verb	Mixed forms	Colloquial forms
داداش (kinship)	singular	0	0

(19) میگم یادتِ استادِ کلاس سخت افزار می گُف یه مقاله هس؟

mi-g-æm   jɒd-et-e   ostɒd-e   kelɒs   sæxtæfzɒr
DUR-say-1SG   remember-2SG.Q-COP.3SG.Q   professor-EZ class   hardware

mi-gof   je   mæɣɒle   hæs
DUR-say.PST.3SG.Q   one.Q   article   there is.Q

'I say, do you remember the professor of the hardware class was saying
there is an article?'

Form of address	Verb	Mixed forms	Colloquial forms
T (enclitic)	NA	0	5

(20) میتونی اونو برام دانلود کنی بیاری؟

mi-tun-i   un-o   bærɒ-m dɒnlod   kon-i   bi-ɒr-i
DUR-can.Q-2SG   that.Q-OM.Q for-1SG download do-2SG SBJV-bring.Q-2SG
'Can you download it and bring it for me?'

Form of address	Verb	Mixed forms	Colloquial forms
pro-drop	singular (3)	0	3

As we showed in the above examples, the sociocultural distance between the
interlocutors of the same gender (i.e., men) was minimal. This was apparent in
the favors the male participants asked from male addressees. The use of singular
verbs was indicative of such sociocultural intimacy. Also, replacing pronouns
with familiarizing nouns (e.g., آقا /ɒɣɒ/ 'man' and داداش /dɒdɒʃ/ 'brother') was
another means through which the informants revealed higher degrees of comfort
with same-gender peers.

2.6.2.2 Opposite-gender favors
Like women when talking to men, men showed increased formality when talk-
ing to women. Men used the T pronoun more with women than women did
with men, but their use of singular and plural verbs in the cross-gender context
was similar.

(21) سلام خانوم باستانی.

sælɒm xɒnum bɒstɒni

hello   Ms.Q   Bastani

'Hello, Ms. Bastani'.

Form of address	Verb	Mixed forms	Colloquial forms
خانم + LN	NA	0	1

(22) علی میگفت تو یه مقاله داری به درد کار من میخوره.

æli mi-goft   to   je   mæɣɒle dɒr-i   be

Ali   DUR-say.PST.3SG you.2SG one.Q article have-2SG to

dærd-e   kɒr-e   mæn   mi-xor-e

pain-EZ   work-EZ   I   DUR-eat-3SG.Q

'Ali was saying that you have an article that comes in handy for me'.

Form of address	Verb	Mixed forms	Colloquial forms
T	singular (1)	0	2

(23) می شه لطف کنی واسم بیاریش هفته دیگه؟ مرسی.

mi-ʃ-e   lotf   kon-i   vɒse-m   bi-ɒr-i-ʃ

DUR-become-3SG please   DO-2SG   for.Q-1PS SBJV-bring-2SG-3SG.OBJ.Q

hæfte-je   dige   mersi

week-EZ   other.Q   thanks

'Is it possible you could please bring it for me next week? Thanks'.

Form of address	Verb	Mixed forms	Colloquial forms
pro-drop	singular (2)	0	3

(24) مریم چه خوب شد دیدمت.

mærjæm ʧe   xub   ʃod   did-æm-et

Maryam   what good became saw-1SG-OBJ.2SG.Q

'Maryam! Good that I saw you'.

Form of address	Verb	Mixed forms	Colloquial forms
FN, T (enclitic)	NA	0	1

(25) اون مقاله هه که بهت گفتم؟

un   mæɣɒle-he   ke   be-het   goft-æm

that.Q article-DEF.Q that to-2SG.Q say.PST-1SG

'That article I told you about'.

Form of address	Verb	Mixed forms	Colloquial forms
T (enclitic)	NA	0	3

(26) یادته؟ اونو میتونی فردا واسم بیاریش؟

jɒd-et-e   *un-o*   *mi-tun-i*   *færdɒ*
remember-2SG.Q-COP.3SG.Q that.Q-OM.Q DUR-can.Q-2SG tomorrow

vɒse-m   *bi-ɒr-i-ʃ*
for.Q-1SG   SBJV-bring-2SG-OBJ.3SG.Q

'Do you remember? Can you bring it for me tomorrow?'

Form of address	Verb	Mixed forms	Colloquial forms
T (enclitic)	singular (2)	0	6

(27) میگم شما مقاله زبان شناسی و داری که استاد فلانی میگفت؟

mi-g-æm   *ʃomɒ*   *mæɣɒle-je*   *zæbɒnʃenɒsi-o*   *dɒr-i*   *ke*
DUR-say.Q-1SG you.2PL article.EZ linguistics-OM.Q have-2SG that

ostɒd-e   *folɒn-i*   *mi-goft*
professor-EZ so-and-so.INDF DUR-say.PST.3SG

'I say, do you have the linguistics article that professor so-and-so was talking about?'

Form of address	Verb	Mixed forms	Colloquial forms
V	singular (1)	1	2

(28) اگه می شه، واسم پرینت می گیریش؟ مرسی.

age   *mi-ʃ-e*   *vɒse-m*   *perint* *mi-gir-i-ʃ*   *mersi*
if.Q DUR-become-3SG.Q for.Q-1SG print DUR-take-2SG-OBJ.3SG.Q thanks

'If it's possible, will you print it for me? Thanks'.

Form of address	Verb	Mixed forms	Colloquial forms
pro-drop	singular (1)	0	4

(29) احوال شما خانوم اکبری؟

æhvɒl-e   *ʃomɒ*   *xɒnum*   *ækbæri*
condition.PL-EZ you.2PL Ms. Q Akbari

'How are you, Ms. Akbari?'

Form of address	Verb	Mixed forms	Colloquial forms
V, خانم + LN	NA	0	1

(30) شما مقاله نانومتریال و دارین؟

ʃomɒ   *mæɣɒle-je*   *nɒnomæteriɒl-o*   *dɒr-in*
you.2PL article-EZ nanomaterial-OM.Q have-2PL.Q

'Do you have the nanomaterial article?'

Form of address	Verb	Mixed forms	Colloquial forms
V	colloquial plural	0	2

(31)   اگه وقت کردید، میتونید واسم ایمیلش کنید؟

 age væɣt kærd-id mi-tun-id   vɒse-m   imejl-eʃ   kon-id

 if.Q time did-2PL DUR-can.Q-2PL for.Q-1SG email-OBJ.3SG.Q do-2PL

 'If you have time, can you email it to me?'

Form of address	Verb	Mixed forms	Colloquial forms
pro-drop	plural (3)	0	4

(32)   یه لحظه وقت دارین؟

 je   læhze-je   væɣt dɒr-in

 one.Q moment-EZ time have-2PL.Q

 'Do you have a moment of time?'

Form of address	Verb	Mixed forms	Colloquial forms
pro-drop	colloquial plural (1)	0	2

(33)   اگه وقت کردین، واسم دانلودش می کنین؟

 æge væɣt kærd-in   vɒse-m   dɒnlod-eʃ   mi-kon-in

 if.Q time did-2PL.Q for.Q-1SG download-OBJ.3SG.Q DUR-do-2PL.Q

 'If you had time, you can download it for me?'

Form of address	Verb	Mixed forms	Colloquial forms
pro-drop	plural (1), colloquial plural (1)	0	5

3 Discussion

We tabulate the presence of the various linguistic features we have been examining in Table 2.

Based on our analysis, the following tendencies emerged:

1. There is more pro-drop in opposite gender interactions.
2. The T pronoun is used most in the opposite-gender interactions of men.
3. The V pronoun is used only in opposite-gender interactions.
4. Mr./Ms. + FN/LN is used only in opposite-gender interactions.
5. Kinship terms/familiarizers are used only by men in same-gender interactions.
6. Plural verbs are used only in opposite-gender interactions.
7. Women appear to prefer the colloquial second person verb ending.
8. Colloquial forms appear more in opposite-gender interactions.

Table 2: Summary of linguistic features in all conditions

	Female		Male	
Feature	**same gender**	**opposite gender**	**same gender**	**opposite gender**
Excerpts	4	10	6	13
Pro-drop	3	5	1	5
T	1	0	1	4
V	0	2	0	3
FN	0	1	1	1
Mr./Ms. + FN/LN	0	1	0	2
Kinship/familiarizer	0	0	3	0
SG verb	6	5	4	5
PL verb	0	1	0	4
PL.Q verb	0	7	0	3
Mixed forms	0	1	0	1
Q	9	17	11	36

Some of these tendencies appear natural, while others are at first glance surprising. Among the expected tendencies are the use of the V pronoun, Mr./Ms. + FN/LN and plural verbs exclusively in opposite-gender interactions by both men and women. Also, the use of kinship terms and familiarizers exclusively in same-gender interactions seems intuitive; however, their use exclusively by men, while not especially surprising, warrants further investigation as to its generality. The preference for pro-drop in opposite-gender interaction is interesting: it leads us to think that pro-drop may be employed as a strategy to avoid choosing between T and V pronouns. While clearly pro-drop is employed in Persian for many other reasons, including a contrastive use where pronouns may be more likely to occur in utterances featuring a combination of persons (e.g., first and second, or second and third), its particular use as a pronoun avoidance strategy merits further study.

We believe our explicit counting of colloquial forms in Persian is novel. It is certainly problematic, and we welcome critique as a means for the community to improve its quantification of colloquial speech. We have indicated as colloquial only those forms that are clearly marked. Perhaps the majority of linguistic forms used in both colloquial and standard speech are neutral with respect to the colloquial/standard distinction. So, the presence of marked colloquial forms has a random element in that it is related to semantic content of an utterance; that is, not all expressions have a marked colloquial option. Despite complications with respect to quantifying colloquial forms, in our data, for example, there does seem to be a preference on the part of women in contrast to men in

the employment of the second person colloquial verbal inflection. Such a finding warrants further exploration. Also the preponderance of colloquial forms in opposite-gender interactions merits further investigation. Since we have already shown that opposite-gender interactions feature more formal properties, such as V pronouns and plural verbs, we can speculate that the robust presence of colloquial forms serves to mitigate the formality, similar to the way in which a singular verb can mitigate a V pronoun in mixed forms/mismatch constructions.

Politeness and formality have both sociopragmatic and pragmalinguistic aspects. This paper has focused primarily on the pragmalinguistic aspects. We hope to have expanded the realm of inquiry in this domain into some new areas, including examination of pro-drop and colloquial forms. It is clear that speakers employ a wide range of nuances in formality setting, and we hope that further research can illuminate some of the nascent patterns we have uncovered here.

References

Agha, Asif. 2003. The social life of cultural value. *Language & Communication* 23 (3). 231–273.

Azizi, Masoud. 2011. Overt Pronoun Constraint (OPC) in Persian: Counter evidence for its universality. In Azadeh Shafaei (ed.), *Frontiers of language teaching*, 94–103. Boca Raton: BrownWalker Press.

Baumgardner, Robert. 1982. *Sociolinguistic aspects of Persian pronouns of address: A macro/micro analysis*. Los Angeles: University of Southern California dissertation.

Blum-Kulka, S. 1987. Indirectness and politeness in requests: Same or different? *Journal of Pragmatics* 11 (2). 131–146.

Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Brown, Robert & Albert Gilman. 1960. The pronouns of power and solidarity. In Thomas A. Sebeok (ed.), *Style in language*, 253–276. Cambridge: MIT Press.

Collins, Chris & Paul M. Postal. 2012. *Imposters*. Cambridge: MIT Press.

Comrie, Bernard, Martin Haspelmath & Balthasar Bickel. 2015. The Leipzig glossing rules: Conventions for interlinear morpheme-by-morpheme glosses. https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf (accessed 30 November 2016).

Coupland, Nikolas & Adam Jaworski. 2004. Sociolinguistic perspective on metalanguage: Reflexivity, evaluation, and ideology. In Adam Jaworski, Nikolas Coupland & Dariusz Galasinski (eds.), *Metalanguage: Social and ideological perspectives*, 15–51. New York: Walter de Gruyter.

Crystal, David. 2006. *Language and the internet*. 2nd edn. Cambridge: Cambridge University Press.

Culpeper, Jonathan, Derek Bousfield & Anne Wichmann. 2003. Impoliteness revisited: With special reference to dynamic and prosodic aspects. *Journal of Pragmatics* 35 (10–11). 1545–1579.

Dalili, Mehdi Vaez. 2009. Agreement (AGR) and the pro-drop/non-pro-drop variation: A meta-analysis of GB and MP accounts. *Philologie im Netz* 49. 84–105.

Fraser, Bruce. 1990. Perspectives on politeness. *Journal of Pragmatics* 14 (2). 219–236.

Goldschmidt, Myra M. 1998. Do me a favor: A descriptive analysis of favor asking sequences in American English. *Journal of Pragmatics* 29 (2). 129–153.

Keshavarz, Mohammad Hossein. 1988. Forms of address in post-revolutionary Iranian Persian: A sociolinguistic analysis. *Language in Society* 17 (4). 565–575.

Koike, Dale April. 1989. Pragmatic competence and adult L2 acquisition: Speech acts in interlanguage. *The Modern Language Journal* 73 (3). 279–289.

Koutlaki, Sofia A. 2010. *Among the Iranians: A guide to Iran's culture and customs*. Boston: Intercultural Press.

Levinson, Stephen C. 1977. *Social deixis in a Tamil village*. Berkeley: University of California dissertation.

Luckmann de Lopez, Kathrin. 2013. Clause-final *man* in Tyneside English. In Gisle Andersen & Kristin Bech (eds.), *English corpus linguistics: Variation in time, space and genre*, 139–162. Amsterdam: Editions Rodopi.

Mahdavi, Pardis. 2007. Passionate uprisings: Young people, sexuality and politics in post-revolutionary Iran. *Culture, Health & Sexuality* 9 (5). 445–457.

Nanbakhsh, Golnaz. 2012. Moving beyond T/V pronouns of power and solidarity in interaction: Persian agreement mismatch construction. *Linguistica* 52 (1). 253–266.

Obolensky, Serge, Kambiz Yazdan Pahan & Fereidoun Khaje Nouri. 1973. *Spoken Persian*. Ithaca: Spoken Language Services.

Perry, John R. 2003. Persian as a homoglossic language. *Cahiers de Studia Iranica* 27. 11–28.

Pešková, Andrea. 2013. Experimenting with pro-drop in Spanish. *SKY Journal of Linguistics* 26. 117–149.

Saeli, Hooman. 2016. Persian favor asking in formal and informal academic contexts: The impact of gender and academic status. *Pragmatics* 26 (2). 315–344.

Widdowson, H. G. 1998. Context, community, and authentic language. *TESOL Quarterly* 32 (4). 705–716.

Behrooz Mahmoodi-Bakhtiari

10 Spoken vs. written Persian: Is Persian diglossic?

Abstract: Since 1954, when Charles Ferguson introduced the issue of "diglossia" in his seminal article and noted some languages have separate spoken and written forms for different purposes, some languages have been identified as diglossic, and among them Arabic may be one of the best known. In this article, I show that Persian is also a diglossic language, although to a minor extent, and introduce the unpredictable differences between the spoken and written forms of Persian in all four major areas of phonology, morphology, syntax, and semantics. The major contribution that this article will hopefully make is expanding our insight into spoken Persian not as a dialect or variation of Persian, but as a level of usage in terms of the social functions of this language. Given the fact that all the major international radio or TV media with Persian services broadcast their less official programs in spoken Persian, and all the native speakers use it in their casual communications, we have to admit that spoken Persian is not merely a dialect but a medium used to convey the ideas of Persian-speaking people when they talk.

All the examples used in this article are derived from the data collected from daily conversations, although I have also consulted my own intuition as a native speaker of Persian, too.

Keywords: sociolinguistics, diglossia, spoken Persian, written Persian

1 Introduction

The differences between spoken Persian and written Persian have long been a focus of study and covered by many publications since the early twentieth century (such as Boyle 1952; Hodge 1957; Vahidiyân 1343/1964; Samareh 1977; Frommer 1981; Jahangiri 1980; Deyhim 1368/1989; and Ghobadi 1996). Different names given to spoken Persian such as "colloquial Persian", "Tehrani Persian", "Persian slang", and recently "Farsi"[1] almost unanimously share the idea that

---

[1] Perry (2003: 12) believes that the replacement of "Persian" by "Farsi" in recent Anglo-American usage "implies a distinct contemporary vernacular and written standard divorced from any linguistic and cultural antecedents".

Behrooz Mahmoodi-Bakhtiari, University of Tehran

spoken Persian is a less important form of Persian, and may be treated merely as a dialect of Persian (see Perry 2003).

However, the drastic expansion of literacy, modernization, and growth of mass media in Iran during the recent decades has changed the status of spoken Persian as merely a dialect of Persian spoken by the inhabitants of Tehran.

Considering the attempts made to raise the spoken form of Persian to literary status by people such as Âxondzâde, Dehxodâ, Jamâlzâde, Hedâyat, and Čubak in the twentieth century, and the fact that there are so many published materials written in spoken Persian (such as almost all the dramatic texts and pop music lyrics, together with major literary fictions and poems) show that spoken Persian has gained a status of its own, to be studied as the major medium of speech among the Persian speakers of Iran.

In this article, I try to prove Jeremiás's (1984) assumption, which regards Persian as a diglossic language. This idea, challenged later by Perry (2003), is still acceptable to me, so I have come up with several examples and use the classical definitions of the notion of diglossia. First, I will review the notion of diglossia. Then I will consider the status of Persian according to the discussions of Jeremiás (1984) and Perry (2003).

2 Diglossia

There are languages in the world that clearly make a distinction between the use of their written forms and the spoken forms. For example, if we consider the following examples of Arabic:

Table 1: A comparison between the data from classical and Egyptian spoken Arabic (Kaye 2009: 570)

Classical Arabic	Spoken Egyptian Arabic	Gloss
raʔā	šāf	'he saw'
ħiðāʔun	gazma	'shoe'
ʔanfun	manaxīr	'nose'
ðahaba	rāħ	'he went'
mā	ʔē(h)	'what'
ʔalʔāna	dilwaʔti	'now'

We notice that what the speakers of Arabic say is not necessarily the same as what they write. This issue, which is found in some other languages, is now

referred to as "diglossia", a term coined by Charles Ferguson in his (1959) article. In that source, he defines this term as follows: "Diglossia is a relatively *stable language situation*, in which, in addition to the primary dialects of the language, there is a very divergent, *highly codified superposed variety* ... which is *learned largely by formal education* and is *used for most written and formal spoken purposes*" (Ferguson 1959: 88; emphases are mine).

Therefore, a language such as Arabic, with a classical literary tradition modeled on the language of the Koran, is a good example of a diglossic language when compared with its regional vernaculars. Modern diglossic situations postulated by Ferguson are that of Greek, "High" German vs. Swiss German, and French vs. Haitian Creole. Perry (2003: 12–13) adds the language situation in Ottoman Turkey to this list as well, considering a highly "Persianized" Turkish of divan literature against everyday spoken Turkish. The written and formal form of the language is referred to as the High variety, and the spoken, colloquial form is known as the Low variety of the diglossic language.

However, this definition lacked something important; the degrees of the structural differences that result in a diglossia were not discussed in this article. In his thirty-five-year retrospective (Ferguson 1993: 223–232), Ferguson admitted that specifying how far apart the two Low and High sociolects must be for a language to qualify as diglossic was missing in the definition, and it seems that no attempt has yet been made to measure what Perry (2003: 13) refers to as "degrees of sociolectal polarity" in a language.

Now we may ask if English is also a diglossic language, considering words such as *wanna* 'want to', *gonna* 'going to', and *ain't* 'to be not'. Here we may refer to Ferguson's terms, which admit a broad definition and a narrow one for diglossia. According to the broad definition, there are systematic stylistic differences between "High" and "Low" language varieties.[2] According to Perry (2003: 13), this is true of most, if not all, natural languages, including English and Persian, and it is hard to imagine a language for which this statement in its broadest sense would not be valid. However, we act according to the narrow definition of diglossia (see Coulmas 2013: 149), and do not technically consider English as diglossic; since the abovementioned items are, firstly, not widely used, they are, secondly, quite predictable and, thirdly, they do not belong to (or form) a separate type of English.

---

2 Here we do not consider Fishman's (1967) extended diglossia, which applies to situations where historically unrelated languages were used together, a prestige language for the High variety, a colloquial one for the Low variety, such as Hebrew and Yiddish in eastern Europe, English and Spanish in the United States, or German and other languages in the Austro-Hungarian Empire.

2.1 Features of diglossia

As it was said, diglossic languages have one variety that is used for "High" (formal, literacy) purposes: the High variety has the most prestige. This contrasts with the Low variety, which is different phonologically, grammatically, lexically, and syntactically. The Low variety is used for informal, mostly spoken purposes. The High variety dominates certain domains: literature, religion, public speaking, while the Low variety dominates in "lower" domains: jokes, intimacy, street use, and is the first language acquired by the native children of that language.

Generally speaking, there are nine distinct features between the High variety and the Low variety. They are as follows (see Schiffman [1998], for further details):

Function: The High variety is used to make sermons, speeches in the parliament, writing the news in different media, some university lectures, serious academic lectures, newspaper editorials, and the like.

Prestige: The speakers regard the High variety as superior to the Low variety in a number of respects.

High is considered more educated, more beautiful, more logical, better able to express important thoughts, etc., especially when something is supposed to be written or published.

Literary heritage: There is a sizable body of written literature in the High variety that is held in high esteem by the speech community.

Acquisition: Children acquire the Low variety, adults use the Low variety in speaking to children, and children use the Low variety in speaking to one another.

Standardization: There are grammatical studies or books, dictionaries, treatises on pronunciation, styles, and so on regarding the High variety. In other words, there are established norms for pronunciation, grammar, and vocabulary that allow variation only within certain limits. National academies are also entitled to take care of the H variety, and place their monetary and academic investments in it.

Stability: Diglossia typically persists at least several centuries, and evidence in some cases seems to show that it can last over a very long period.

Grammar: The High variety has some grammatical categories that are not in the Low, and has an inflectional system of nouns and verbs that is much reduced or totally absent in the Low variety.

Lexicon: The High variety includes in its total lexicon some technical terms and learned expressions that have no regular Low equivalents since those subjects involved are rarely discussed in the Low variety. On the other hand, the Low variety includes in its total lexicon some popular expressions and the names of very homely objects or objects of much localized distribution that have no regular High variety equivalents since the subjects involved are rarely discussed in pure High.

Phonology: The sound systems of High and Low constitute a single phonological structure of which the Low phonology is the basic system. A notable feature might be the unpredictable and irregular phonological differences between Low and High, which are not necessarily the result of fast speech.

2.2 Three misunderstandings

According to my own experience, once we study the Low variety of a language, we may confront three misunderstandings. First, we may consider it a form that lacks prestige, perhaps even "vulgar" or "ungrammatical"; spoken only by children, lesser beings, or uneducated people. As a matter of fact, it is not so. The Low variety of language is not the same as slang, and it may be seen in languages with profound literatures too, such as Persian. As an example, let's consider the word *biyâr* 'bring!', which is the spoken form of *biyâvar*. This word, which is not noticed in any written text or formal speech of Persian, is in the poetry of Rudaki and Hâfez, two of the greatest figures in Persian literature:

(1) *ey sabâ, nekhati az xâke rahe yâr <u>biyâr</u>,*
     bebar anduhe del o moždeye deldâr, <u>biyâr</u>.

     [O Breeze! Bring a fragrance from the place of my beloved to me,
     Take the sorrow of the heart away, and bring the enunciation of the beloved**]**

     Or the word *dahan* (written Persian *dahân* 'mouth'), seen in the poems of Hâfez and Sa'di:

     sâqiye sim-sâqe man, gar hame dord midahad
     kist ke tan co jâme mey, jomle <u>dahan</u> nemikonad?

     [If my cup-bearer the silver legged gives only dregs to me,
     Who is there who does not turn its body all mouth, like the wine cup?]

     garat az dast barayad, <u>dahani</u> širin kon,
     mardi ân nist ke mošti bezani bar <u>dahani</u>

     [In case you can, sweeten a mouth,
     It is not courage to hit your fist against another man's mouth**]**

Even the verb *vâ kardan* (WP *bâz kardan* 'to open') in the poem of Hâfez:

(2) *neqâbe gol kešid o zolfe sonbol*
     gereh-bande qabâye qonče <u>vâ-kard</u>

     [(The breeze) drew aside the veil of the rose, and the tress of the hyacinth,
     And loosened the knot of the coat of the rose-bud**]**

The second case we may keep in mind is that High does not necessarily mean "the polite form of Language", hence Low does not necessarily denote impoliteness. For example, let's consider the sentences below, one in written Persian and the other in spoken Persian, and both quite polite:

(3) High: *âyâ âqâye ra'is tašrif dârand?*
     Low: *âqâye ra'is tašif dâran?*
     'Is Mr Boss there?'

And the third case deals with the relationship between diglossia and bilingualism. These two have basically nothing to do with each other, and the speaker of a diglossic language is not regarded as bilingual because his or her language has two distinct varieties.

3 Diglossia and Persian

Although the clear differences between the two forms of Persian (written Persian and spoken Persian) have long been studied, it seems that the term *diglossia* for Persian was first used by Henderson (1975), who tried to show the phonological differences of written Persian and that of Kaboli spoken Persian. But the major publication that considered diglossia for the entire being of Persian is Jeremiás (1984) article. In "Diglossia in Persian", Jeremiás notes that the differences between the formal and informal varieties of Persian are large enough to warrant the term *diglossia*, and claims that contemporary Persian offers "a striking example of diglossia" (Jeremiás 1984: 286). She considers five features of written Persian and spoken Persian, and concludes that "there are fundamental differences between formal and informal Persian", which are "as marked as those between two independent languages" (Jeremiás 1984: 287). Jeremiás's five distinguishing features of Persian "diglossia" are as follows:

1. the innovation of progressive constructions of the type *dâram kâr mi-konam*, *dâštam kâr mi-kardam* 'I am/was working';
2. the stressed suffix marking a defined noun, as *pesar-è jim šod* 'the boy made himself scarce';
3. expanded and versatile use of pronominal enclitics as compared with the formal language;
4. omission of the preposition, e.g., *Tehrun raft* 'he went to Tehran'; and
5. violations of the standard verb-final word order in Low variety utterances of the type *raft bazâr* 'he went to market' and *na-didam bâbâ-t-o* 'I didn't see your dad' (Jeremiás 1984: 281–88).

On the other hand, Perry (2003) rejects the idea, and believes that Jeremiás has not measured the pertinent synchronic differences, and has contented herself with five disparate features in Persian. He notes that Jeremiás's determination to foist diglossia on Persian may lead to some exaggerated criticism of pedagogic approaches. For example, in her discussion of the progressive aspect of the verb in present and past tenses, using *dâštan* as an auxiliary, Perry notes that this is a feature of the Informal Standard Persian, and as opposed to Jeremiás (who holds that present-day grammars do not usually include such forms in the paradigms of verbs, but only mention them as a colloquialism), he counts some of the most well-known modern grammars that have included "adequate descriptions of the forms, even as they categorize them as colloquial in usage" (Perry 2003: 25).

In order to prove that Persian is not a diglossic language, Perry makes use of the model proposed by Schmidt (1974), in which he has provided a list of 900 words, and has found a maximum of 560 diglossically distinct pairs (Perry 2003: 23). This is not something to be found in the Persian lexicon. He then concludes that the greatest contrast between Arabic and Persian lexical "diglossia" is seen in the everyday non-cognate vocabulary, where Arabic shows a six- or sevenfold superiority in the other two sub-categories. Among the Functional-Frequency listings he counts some common verbs showing stem alterations, such as *mi-š-am* 'I become', *mi-g-am* 'I say', *mi-d-am* 'I give', *mi-xây-in* 'you want', *mi-tun-am* 'I can', *be-zâr* 'let', and *na-zâšt* 'he didn't let' (Perry 2003: 24). He finally shows that, on the basis of Schmidt's mechanical quantification, Arabic would appear to be about three and a half times as diglossic as Persian. This overwhelming number of diglossic morphological features in Arabic and the negligible number in Persian convince Perry to conclude that Persian is not nearly as dichotomous as Arabic in terms of a broad definition of diglossia, and claims that Arabic is, and Persian is not, diglossic in terms of the narrow definition.

4 The approach of this article

In this article, I would like to make use of the term *diglossia*, and by dealing with the features attributed to it and providing more examples than Jeremiás, show that Persian is diglossic, although not to the extent that Arabic is. In other words, Persian may not act like Arabic in terms of lexical differences in its High and Low varieties, but may show some active, unpredicted differences in terms of other linguistic features between its spoken and written forms.

If we consider the nine features of diglossic languages, we may see that almost all of them are present in Persian. Written Persian has the function of delivering the serious lectures and sermons, news reading in the media, newspaper editorials, and the like. We may even consider the notable change of the language when, for example, a broadcaster stops reading the news, and starts talking to a reporter outside the studio.

In terms of prestige, all speakers of Persian are aware that what they write should not be in spoken Persian. That is why we have no written text of spoken Persian dating back more than a century ago. As a matter of fact, the only surviving texts in spoken Persian are the plays written during the Qâjâr era in the nineteenth century, and the only one entirely written in spoken Persian is the play *Ostâd Noruz-e Pineduz* ('Ostâd Noruz, the Shoesmith') by Ahmad Kamâl ol-Vezâre-ye Mahmudi (see Mahmudi-Baxtiyâri 1388/2009). This may also account for the other feature of diglossic languages, which holds that the literary heritage of a language is basically written in the High variety.

As mentioned earlier, Iranian Persian-speaking children acquire spoken Persian before they go to school, and during their official education that they study written Persian seriously. The media program for children is always broadcast in spoken Persian. In terms of standardization, again written Persian as the High variety has been the main focus in educational institutions and, to the best of my knowledge, the Iranian academy has taken no serious action to work on or publish a lexicon or a grammar of spoken Persian. The fact that some lexical and syntactic features of spoken Persian may also be found in classical Persian (and not in modern written Persian) may also assure us that such differences have long existed, although there are few written documents to prove it. The undeniable grammatical differences between the High and Low varieties (as will be shown below) may also account for our claim about the diglossic differences between written Persian and spoken Persian. Although many phonological differences are predictable based on phonological rules or diachronic change, the large number of unpredictable phonological differences between written Persian and spoken Persian may also be added to this list. Perhaps the only

feature that may not encompass Persian fully is the case of lexicon in diglossia, about which Perry (2003) seems to be totally right.

Now, in order to show the different linguistic differences and provide examples in addition to those of Jeremiás, let us start with the phonological alternations and differences between written Persian and spoken Persian. Apart from my intuition as a native speaker, all the data used below have been extracted from a corpus of forty-eight hours of casual speech that I collected from 2009 to 2011. This corpus contains recorded sport program and roundtables broadcasted from the Iranian national TV channels, as well as some reports and commentaries downloadable from several Persian-speaking media. Since almost all the data belong to the official media programs, sociological variants were not considered in data collection as the differences between several classes of society in using spoken Persian require separate research.

4.1 Phonological alternations

4.2 Vowel shifts

Some words have different pronunciations when changed from written Persian to spoken Persian, and such changes are neither easily justified nor explainable in terms of productivity, for example, the pronunciation of the written Persian words *agar* and *magar* 'if' as *age* and *mage*, or the unjustified shift of *a > o* in the pronunciation of the written Persian word *âvardan* into *âvordan* 'to bring'.

This inconsistency exists in some other cases, which seem systematic at first glance. For example, it is generally believed that written Persian clusters *-âm* and *-ân* change into *-um* and *-un* in spoken Persian, as in the words *bâdâm > bâdum* 'almond', *gerân > gerun* 'expensive', *nân > nun* 'bread', *xâne > xune* 'house', and *âsân > âsun* 'easy'. However, cases such as *ehsân* 'kindness', *tâbân* 'shining', *peymân* 'treaty', *dâstân* 'story', *bimârestân* 'hospital', *emtahân* 'exam', *tufân* 'storm', *qahremân* 'champion', *zâyemân* 'giving birth', *zamân* 'time', *saratân* 'cancer', *kârgardân* 'director', *etminân* 'certainty', *pâdegan* 'barracks', *soltân* 'king', and many other words do not follow such a rule, although they are more or less widely used in current Persian.[3] On the other hand, there

---

3 Needless to say, "bookish" words do not undergo such phonological changes, such as *ehtešâm* 'dignity', *maqâm* 'post', *elhâm* 'inspiration', *garibân* 'collar', *bâmdâdân* 'morning', *armaqân* 'souvenir', *râyegân* 'free', and so on. Also, the plural marker *-ân* may also survive such a change, as in *jânevar-ân* 'animals', *pestândâr-ân* 'mammals', *niyâ-kân* 'ancestors', as well as the loanwords such as *âpârtemân* 'apartment', *šârlâtân* 'charlatan', *kâpitân* 'captain', and the like. The fact is that there are many other "common" words (as mentioned earlier) that are used to a large extent and do not change their ultimate vowels.

are instances in which words with such endings change their vowels, but not as we would expect, such as *tumân > toman* [*tumun*] '10 rials', *hendevâne > hendune* [*hendevune*] 'watermelon', *dahân > dahan* [*dahun*] 'mouth', and *mi-ta-vân-am > mi-tun-am* [*mi-tavun-am*] 'I can'. Also, most of the words ending in the place marker *-estân/-setân*, do not follow this change except in some rare cases such as *golestân > golestun* 'rose garden', and *qabrestân > qaberestun* 'grave yard'. It should also be noted that such a change is not merely seen in the final vowels, and in a case such as *âmadan > umadan* 'to come', the change of *â > u* is seen, but the same change may not appear in verbs with very similar phonotactics (but less usage) such as *âmuxtan > *umuxtan* 'to learn' and *âmixtan > *umixtan* 'to mix'.

On the other hand, the seemingly regular change of *e > i* takes place in spoken Persian, as in *šekam > šikam* 'belly', *jegar > jigar* 'liver', *negâh > nigâ(h)* 'sight', *šekar > šikar* 'sugar', *tekke > tikke* 'piece', and *gonješk > gonjišk* 'sparrow'. But it may be remembered that the same change does not happen in words that are very similar to the abovementioned, such as *pedar > *pidar* 'father', *sepâh > *sipâh* 'army', and *sekke > *sikke* 'coin'.

4.3 Vowel harmony

The harmony of the penultimate vowel with the ultimate one in (mostly) two-syllable words is another feature seen in the phonology of spoken Persian, which is common but not regular. Changes such as *sebil > sibil* 'moustache' and *kelid > kilid* 'key' exist along with cases such as *sefid > *sifid* 'white'. It seems that the only regular vowel harmony is the increasing of *o > u* in the words with the syllable structure CV-CVC, as in *doruq > duruq* 'lie', *šoluq > šuluq* 'crowded', *dorud > durud* 'salute', *soqut > suqut* 'fall', *qorur > qurur* 'pride', *lozum > luzum* 'necessity', *forud > furud* 'landing', and many other similar cases. The *o > u* shift also happens in very few cases, which have nothing to do with the vowel harmony noted above, and no specific phonological rule can account for it, as in *xânom > xânum* 'lady', *nâxon > nâxun* 'fingernail', and *xord > xurd* 'scattered'.

5 Deletions

5.1 Vowel deletions

There are several instances of spoken Persian in which vowels of the written Persian are deleted. For example, the deletion of /o/ in the following verbs:

mi-foruš-am > mi-fruš-am 'I sell', *mi-gozar-ad > mi-gzar-e* 'it passes', *mi-gozar-ân-am > mi-gzar-un-am* 'I spend [the time]', or in the noun *alominiyom > âlminiyom* 'aluminum'. Such a deletion leads to the verbal roots with the initial consonant clusters, which are not possible according to the phonology of standard Persian (also see below).

The other example deals with the deletion of /*e*/ in the following verbs: *mi-šenav-am > mi-šnav-am* 'I hear', *mi-šekan-am > mi-škan-am* 'I break', *mi-šenâs-am > mi-šnâs-am* 'I recognize', *mi-neš-ân-am > mi-nš-un-am* 'I settle', *mi-sepâr-am > mi-spâ-ram* 'I trust', *mi-šemo-ram > mi-šmor-am* 'I count', and *mi-ferest-am > mi-frest-am* 'I send'; along with the noun *fâyede > fâyde* 'use'. But this is not again something absolute as there are verbs such as *mi-neš-ân-am > mi-šun-am* 'I settle'. They are the same, and *mi-nsh-un-am* is one step further in terms of change than *mi-shun-am* and *mi-keš-ân-am > mi-keš-un-am* 'I pull [somebody]', which are almost identical in terms of phonotactics, but behave differently when used in spoken Persian. An example for the deletion of /*a*/ in a verb is *andâxtan* 'to drop': *mi-andâz-am > mi-ndâz-am* 'I drop'.

The vowels /*a*/ and /*e*/ are deleted as the initial vowels of the possessive enclitics when adjacent to the final vowel of the base noun, such as *sandali-y=am > sandali=m* 'my chair', *pâ-hâ-y=ešân > pâ-hâ=šun* 'their feet', and *zânu-y=at > zânu=t* 'your knee', in which both the hiatus *-y-* and the initial /*a*/ of the clitic are removed. A similar process takes place for the preposition *barâye* 'for', with the enclitics: *barây=am > barâ=m* 'for me', *barây=at > barâ=t* 'for you', *barây=aš > barâ=š* 'for him/ her', etc.

5.2 Consonant deletions

Among the several consonant deletions, some are not regular, such as the deletion of /*t*/ in *dust=at dâr-am > dus=et dâr-am* 'I love you', or *xâstegâr > xâs(s)egâr* 'suitor'. We may refer to some major ones, which seem to be more common, such as the deletion of the final consonants in the terminating consonant clusters: for example, *čand > čan* 'how many', *past > pas* 'wicked', *češm > češ* 'eye', *mošt > moš* 'fist', *fekr > fek* 'thought', and *dast > das* 'hand'.[4] This deletion is quite common, and the final consonant is heard only when a vowel follows it, such as *dast-e rezâ* 'Reza's hand'.

---

4 Persian does not allow consonant clusters with more than two consonants. Therefore, the pronunciation of words with final clusters attached to another consonant will result in the deletion of the final consonant of the cluster, as in *dastkeš > daskeš* 'gloves', *dastmâl > dasmâl* 'handkerchief', *asbsavâri > assavâri* 'horse riding', *past-fetrat > pasfetrat* 'wicked'.

But in terms of the specific consonant deletions, we may refer to /h/ deletion in several places. The glottal /h/ is normally deleted after long vowels, such as *siyâh > siyâ* 'black', *kolâh > kolâ* 'hat', *masih > masi* 'Christ', and *sotuh > sotu* 'restlessness'. Naturally, they are pronounced in case a vowel follows /h/, such as *siyâh-i* 'darkness', and *kolâh-â* 'hats'. Apart from *sobh > sob* 'morning', in which the final /h/ is deleted after the consonant, we may see some other deletions of /h/ in the middle of the word, which might take place with or without the compensatory lengthening of the adjacent vowel. In words such as *mašhad > mašad* 'Mashhad City', and *tashih > tasi* 'correction', we do not recognize a compensatory lengthening, while in the pronunciation of words such as *qahr > qa:r* 'anger', *behzâd > be:zâd* 'Behzad', *e'teqâd > e:teqâd* 'belief', and *šahr > ša:r* 'city', the deletion of the glottal stop /'/ or glottal fricative /h/ results in the lengthening of the previous vowel. Some words also represent the deletion of a consonant with compensatory lengthening of the non-adjacent vowel, such as *qat' > qa:t* 'cut', *šam' > ša:m* 'candle', *tanhâ > ta:nâ* 'alone', and *jam' > ja:m* 'group'. An interesting case of /h/ deletion takes place in *šâhzâde > šâzde* 'prince', in which the three-syllable written Persian word is reduced to a two-syllable one, in which the deletion of the final /h/ of the first syllable has resulted in the deletion of the vowel of the second syllable too.

Not only /h/ may be deleted in words. The glottal stop /'/ is also deleted in words such as *daf'e > dafe* 'turn', *na'lbeki > nalbeki* 'saucer', *ta'ârof > târof* 'offer', and *al'ân > alân* 'now'. A similar deletion may also be seen for /r/ in *tašrif > tašif* 'presence' and /l/ in *aslan > asan* 'altogether'.

5.3 Syllable deletions

In terms of deleting the syllables in spoken Persian, cases are not regular at all, and normally take place in verbs. The scattered cases that we know include the deletion of /ne/ in words such as *mi-nešin-am > mi-šin-am* 'I sit', and *na-nešast-am > na-šest-am* 'I did not sit', along with the deletion of /go/ in *mi-go-zâr-am > mi-zâr-am* 'I put'. The words *xodâ hâfez* and *âqâ mirzâ*, which turn to *xodâfez* 'goodbye' and *âmirzâ* 'Mr Mirzâ' in casual speech, are examples of syllable deletions in non-verbal words of spoken Persian. I mean nouns or adjectives, etc.

5.4 Non-syllabic cluster deletions

There are also cases in which an adjacent vowel and consonant are deleted without having formed a syllable, such as in *čahâr > čâr* 'four', *čehel > čel* 'forty', *pirâhan > piran* 'dress', *mo'allaq > mallaq* 'suspended', *golule > gulle* 'bullet', and

the verbs *mi-âvar-am > mi-yâr-am* 'I bring', and *dar-bi-y-âvar > dar-bi-y-âr*, or *dar-âr* 'take out!' The two verbs *biyâvar* 'bring!' and *biyandâz* 'drop!' turn to *biyâr* and *bendâz* respectively, and exhibit the irregular phonological changes further.

6 Insertion

Since spoken Persian basically focuses on reducing the linguistic items, insertion of phonemes are not so common in it. However, there are several examples of consonant and vowel insertions, such as the vowel insertion in the words *mehrbân > mehrabun* 'kind', *kârgar > kâregar* 'worker', and *pâsbân > pâsebân* 'police'. Examples of consonant insertion basically deal with the gemination of specific consonants, such as *be-par > be-ppar* 'jump!', *be-pâ > be-ppâ* 'beware!', *do-tâ > do-ttâ* 'two items', *tipâ > tippâ* '(angry) kick', *be-kan > be-kkan* 'remove!'.

7 Morphological alternations

7.1 Free morphemes

Spoken Persian has some specific morphemes that seem to be uniquely its own. These morphemes are not used in written Persian, and act differently in terms of being lexical or functional. For example, the copula *ast* 'is' shows itself as *=e* in *un medâd=e* 'that is a pen', and *-s(t)* in *in rezâ-s(t)* 'this is Reza'. Also, verbs such as *vâysâdan* 'to stop, to stand up' (cf. written Persian *istâdan*), *pâ šodan* 'stand up' (cf. written Persian *boland šodan*), or words such as *var* 'direction', *bâše* 'all right, OK', *yavâš* 'slowly, softly', *gošne* 'hungry', *vâse* 'for', *âre* 'yes'; are all specifically found within the lexicon of spoken Persian. Among the items mentioned, the verb *vâysâdan* (or *vâstâdan*) stands out as its verbal root is *-st-*, which is basically impossible in written Persian morphology. We keep in mind that this verb may not be treated identically with spoken Persian verbs such as *mi-škan-am* 'I break', *mi-šnav-am* 'I hear', or *mi-šnâs-am* 'I recognize' as these verbs are the reduced forms of written Persian verbs, while *vâstâdan* is used on its own and has no written Persian equivalent. The conjugation of this verb is also interesting and unique:

Table 2: Conjugation of the verb *istâdan* in spoken Persian

Simple present	*vây-mi-st-am, vây-mi-st-l, vây-miste, . . .*
Negative present	*vây-ne-mi-st-am, vây-ne-mi-st-i, vây-ne-mi-st-e, . . .*
Simple past	*vây-st-âd-am, vây-st-âd-i, vây-st-âd-Ø, . . .*
Negative past	*vây-na-st-âd-am, vây-na-st-âd-i, vây-na-st-âd-Ø, . . .*
Present subjunctive	*vây-s(t)-am, vây-s(t)-i, vây-s(t)-e, . . .*
Present durative	*dâr-am vây-mi-st-am, dâr-i vây-mi-st-i, dâr-e vây-mi-st-e, . . .*
Imperative (positive and negative)	*vâ(y)-st-â, vây-na-st-â/ *vâ(y)-st-Ø, *vâ(y)-na-st-Ø*

Another important feature of spoken Persian morphology is the addition of personal enclitics to prepositions such as *az* 'from', *be* 'to', and *bâ* 'with'. This addition does not change the initial vowel of the clitic when the preposition ends up in a consonant: *az=am* 'of me', but the vowel harmony process takes place with respect to *be* and *bâ*, which terminate in vowels, and results in the vowel change of the clitic: *be-h=emun* 'to us', *bâ-h=âšun* 'with them', which never happens in written Persian:

Table 3: Personal enclitics with the prefixes

az	*az=am, az=at, az=aš, az=amun, az=atun, az=ašun*
be	*be-h=em, be-h=et, be-h=eš, be-h=emun, be-h=etun, be-h=ešun*
bâ	*bâ-h=âm, bâ-h=ât, bâ-h=âš, bâ-h=âmun, bâ-h=âtun, bâ-h=âšun*

Among the free functional morphemes, we may count *tu* 'in' (cf. written Persian *dar, dâxel*); *dam* 'near, by' (cf. written Persian *kenâr*), *dar* 'in front of, before' (cf. written Persian *jelo*) and *vâse* 'for' (cf. written Persian *barâye*), along with the question words *ki* 'who' (cf. written Persian *ke*) and *či* 'what' (cf. written Persian *če*) and their compound nouns: *harki* 'every one', *hičči* 'nothing', *hiški* 'no one'. The question words *ki* and *či* have their own distributional behavior, too, as they can be pluralized and accept the plural marker *-â* (written Persian *-hâ*), while *ke* and *če* do not:

(4)   **ke-hâ bâ mâ mi-ây-and? > ki-y-â ba mâ mi-y-ân?*
      'Which people will come with us?'

(5)   **če-hâ râ mi-šav-ad dar havâpeymâ bord? > či-y-â ro mi-še bord tu havâpeymâ?*
      'What items may be taken on board?'

Another question word that may be pluralized in spoken Persian is *key* 'when', in the form of *keyâ*:

(6) **key hâ âzâd-tar-i?* > *key-â âzâd-tar-i?*
     'Which times are you more free'?

7.2 Bound morphemes

The verbal endings of spoken Persian (SP) (especially in the present tenses) are different from those of written Persian (WP):

Table 4: Personal verb endings in WP and SP

Person	1SG	2SG	3SG	1PL	2PL	3PL
WP	*-am*	*-i*	*-ad*	*-im*	*-id*	*-and*
SP	*-am*	*-i*	*-e*	*-im*	*-in*	*-an*
SP (for *xâstan* and *âmadan*)	*-m*	*-y*	*-d*	*-ym*	*-yn*	*-n*

As Table 4 shows, the conjugations of the verbs *xâstan* 'to want' and *âmadan* 'to come' are exclusively their own as their spoken Persian form ends in a vowel. Table 5 shows the conjugation of these verbs in written Persian and spoken Persian:

Table 5: Conjugation of the verb *xâstan* in WP and SP

Person	1st	2nd	3rd
WP (SG)	*mixâham/ miâyam*	*mixâhi/ miâyi*	*mixâhad/ miâyad*
SP (SG)	*mixâm/ miyâm*	*mixây/ miyây*	*mixâd/ miyâd*
WP (PL)	*mixâhim/ miâyim*	*mixâhid/ miâyid*	*mixâhand/ miâyand*
SP (PL)	*mixâym/ miyâym*	*mixâyn/ miyâyn*	*mixân/ miyân*

8 Affixes

Among the derivational affixes, the suffix *-aki* seems to be specific to spoken Persian, and is used in making adverbs such as *zur-aki* 'forcefully', *piš-aki* 'in advance', *yavâš-aki* 'silently', *râst-aki* 'really', *moft-aki* 'freely', *dozd-aki* 'sneaky, in a sneaky manner', *doruq-aki* 'fakefully', *šâns-aki* 'by chance', *xar-aki* 'foolishly', *zirzir-aki* 'covertly', and *holhol-aki* 'hurriedly'. It also attaches to some infinitives to make adverbs such as *xâbidan-aki* 'in a lying manner' and

vâstâdan-aki 'in a standing manner'. Another suffix is *-u*, which forms from nouns or verb stems an adjective denoting a physical or moral characteristic: *sibil-u* 'moustache wearing' (*sebil/ sibil* 'moustache'), *tars-u* 'cowardly' (*tarsidan* 'to be afraid'), *šekam-u* 'owner of a big appetite' (*šekam* 'belley').

The subjunctive prefix *be-* which is attached to the present roots, may be deleted in some spoken Persian sentences, such as *age gom (be)-š-e či?* 'What if it gets lost?'; *bâyad kâr (be)-kon-e* 'he should work'. However, in the sentences with compound verbs, when the nominal part of the verb is definite, it should necessarily appear:

(7) *mitarsam eštebâh bokonam/ konam.*
     'I fear to make a mistake'.

     *mitarsam hamin eštebâh ro bokonam/ *konam.*
     'I fear to make the same mistake'.

     mitunin kar bokonin/ konin.
     'You can work'.

     *Mitunin az in kârâ bokonin/ *konin.*
     'You can do such works'.

     bâyad fekr bokonin/ konin.
     'You must think'.

     *bâyad ye fekri bokonin/ *konin.*
     'You must make a thought'.

9 Clitics

The personal enclitics that denote both possession and the direct object case are slightly different in written Persian and spoken Persian in terms of form:

Table 6: Possessive and direct object personal enclitics

Person	1SG	2SG	3SG	1PL	2PL	3PL
WP	*-am*	*-at*	*-aš*	*-emân*	*-etân*	*-ešân*
SP	*-am*	*-et*	*-eš*	*-emun*	*-etun*	*-ešun*

However, in terms of use, they act very differently. Perry (2003: 22) counts as three separate features for the use of the enclitic in spoken Persian: as a subject marker (*raft-eš* 'he went'), in combination with a preposition (*be-h-eš goft* 'he said to him'), and as attached to a sentence constituent other than these or a

verb (*če-t-e* 'what's wrong with you?'). It should be noted, however, that these usages are all to be found in earlier styles of Persian, notably classical poetry, as well as in modem informal colloquial (such as *be-irân-at bord* 'took you to Iran').

For another example, direct objects in the perfective verbs cannot be denoted by clitics in written Persian, while they can in spoken Persian. The question *če kasi to râ zad-e ast* 'Who has beaten you?' cannot be rewritten as **če kasi zade ast=at?*, but *ki zad-at=et?* is acceptable in spoken Persian. The same holds true for the question *kasi ke to râ na-did-e ast? > kasi ke na-did-at=et?* 'Nobody saw you, did he?'

Another difference lies in the spoken Persian clitics adjoining the distributive determiners such as *hame* 'all', *har* 'every', and *har-do* 'both', while such a thing does not take place in written Persian, as seen in the following examples:

(8) *har-do-ye šomâ > *har-do-y=etân > har do=tun*   'both of you'[5]

   hame-ye mâ > *hame-ye=mân > ham=amun   'all of us'

These clitics may also be used as secondary predicates in spoken Persian, but not in written Persian:

(9) *man bad-axlâq=et ro ham dus dâram.*
   'I love you (even when you are) bad-tempered'.

   qeyr-e rasmi=t ham qabule.
   'You are accepted (although) not official'.

10 Compounding

An interesting issue in the study of spoken Persian is how reduplication adds to the lexicon of this linguistic type. Compounding, which is shown mostly in reduplication here, makes adverbs or secondary predicates. In written Persian we also have sentences such as *u râ zende-zende suz-ân-d-and* 'They burnt him alive'. But there are also some other reduplicative adverbs that do not seem to be used in written Persian, and are perhaps among the properties of spoken Persian:

---

5 For more than two, the quantifier *tâ* is also added to the number: *har se-tâ=tun* 'all three of you'.

(10) *teflaki javun-javun mord.*
'Poor fellow, he died young'.

čerâ čâyi=t ro dâq-dâq mi-xor-i?
'Why do you drink your tea so hot?'

dust o rafiqâ-ye bad šuxi-šuxi 'amali=š kard-an.
'Bad friends and pals gleefully made him an addict'.

ba'zi-yâ alaki-alaki be hame jâ res-id-an.
'Some people have reached everywhere so so easily'.

âdam košt-e, râst-râst ham râh mi-r-e.
'He has done a crime, and is walking so safely too'.

10.1 A specific adjective structure: Noun/adjective + *bešo*

When Persian infinitive receives the adjective making suffix *-i*, the produced adjective denotes the potential of the content of the verb, such as *xord-an-i* 'edible' (*xord-an*) 'to eat', *did-an-i* 'sight to see' (*did-an*) 'to see'. Sometimes it denotes the character is willing to do something, as in *mehmân-hâ-ye mâ raftan-i nistand, mând-an-i hastand* 'Our guests do not seem to be willing to go, they would stay' (they are not going-like, they are staying-like). In spoken Persian, the adjective *šodan-i* 'capable of becoming' is denoted by the use of the imperative verb *bešo* 'become!', together with the noun, mostly in negative sentences:

(11) *in bačče âdam-bešo nist.* (*âdam* 'human')
'This child is not treatable'. (lit. 'not capable of becoming a human')

in mâšin az avval ham dorost-bešo nabud. (*dorost* 'correct')
'This car was not repairable from the beginning'.

in virune dige xune-bešo nemiše. (*xune* 'house')
'This wreckage will not be a house again'. (lit. 'is not capable of becoming a house')

11 Syntactic alternations

There are many cases of syntactic differences between written Persian and spoken Persian. Apart from the case of the progressive constructions proposed by Jeremiás (which I approve in terms of their specific use in spoken Persian),

perhaps the first difference shows itself in the case of agreement and definiteness. Written Persian does not allow adjectives to be pluralized, and the adjectival phrases are pluralized by adding *-hâ* to the head noun, such as *pesar-e xub*, *pesar hâ-ye xub* 'good boy, good boys'. But the spoken Persian data show that the adjectives may also be pluralized, provided that the *Ezafe* marker (*-e*) is not pronounced: *doxtar-hâye xošgel > doxtar-xošgel=â* 'beautiful girls' (*xošgel* 'beautiful'), *bačče hâye porru > bačče porru=â* 'rude kids' (*porru* 'rude'). A similar case holds true with the definite marker *-e*, which can be added to the adjectives in spoken Persian and not in written Persian: *doxtar-xošgel-e* 'the beautiful girl', *piremard-badaxlâq-e* 'the bad-tempered old man'. In terms of definiteness, in spoken Persian there is a definite marker, stressed *-é* (*-yé* after *-*; *-hé* after other vowels), which may be appended to a singular: *pesar-é umad* 'the boy (in question) came', *tule-hé mord* 'the puppy died' (see Perry 2007: 981).

However, the greatest syntactic differences between written Persian and spoken Persian concern the differences in word order and movements. Often a preposition is deleted from a sentence, and its modifier is moved as a result. For example:

(12) *mixâham <u>be</u> sinamâ beravam. > mixâm beram Ø sinamâ.*
'I want to go to cinema'.

u râ <u>be</u> esfahân ferestâdim. > ferestâdim=eš Ø esfahân.
'We sent him to Esfahan'.

<u>be</u> baqal=am biyâ. > biyâ Ø baqal=am.
'Come into my arms'.

hamin al'ân <u>be</u> xâne residam. > hamin al'ân residam Ø xune.
'I arrived home right now'.

The movement of the modifier as a result of preposition deletion may not take place when some other identical sentences follow the main clause with conjunctions:

(13) *diruz be sinamâ raftim, be bazâr ham raftim, be resturân ham raftim. > diruz raftim Ø sinamâ, Ø bazâr-am ratim, Ø resturân-am raftim.*

'Yesterday, we went to cinema, we went to the market too, and we went to restaurant too'.

However, the complex (or reciprocating) conjunctions '*ham … ham*' 'both … and' keep both the movements and the deletions:

(14) *diruz ham raftim Ø sinamâ, ham raftim Ø bâzâr, ham raftim Ø resturân.*
'We went both to cinema, and the market, and the restaurant'.

Although such a change seems to take place mostly with the dative preposition *be*, the destination is not always a concrete noun (15), and the verbs are not necessarily movement verbs (16). Examples for (15) are:

(15) *be donbâl=at miâyam. > miyâm Ø donbâl=et.*
'I will come to take you'. (lit. 'come after you')

be sorâq=aš raftam. > raftam Ø sorâq=eš.
'I went to him'. (lit. 'went to his situation')

Examples for (16):

(16) with the verb *andâxtan* 'to drop':
tofang=at râ ruye zamin biyandâz > tofang=et=o bendâz Ø zamin.[6]
'Drop down your gun'. (lit. 'drop your gun on the ground')

čerâ hame čiz râ be garden=e man miandâzid? > čerâ hame čizo mindâzin Ø garden=e man?
'Why do you put all the blame on me? (lit. 'Why do you drop all the blame on my neck?')

âxar to râ be zendân mi-andâzand. > âxâr mi-ndâzan=et Ø zendun.
'They will finally put (lit. 'throw') you in prison'.

with the verb *gozâštan* 'to put':

kolâh=aš râ bar sar=aš gozâšt. > kola=š=o gozâšt Ø sar=eš.
'He put his hat on his head'.

yek lahze xodat ra be jâye man be-gozâr. > ye lahze xodeto be-zâr Ø jâye man.
'Put yourself in my place for a minute'.

gâhi mive'i dar dahân=am migozâram. > gâhi ye mive'i mizâram Ø dahan=am.
'I sometimes put a fruit in my mouth'.

with the verb *xordan* 'to eat, to hit':

---

6 Of course for the preposition *ruye* 'on', it is deleted if the item is dropped on the ground (or, metaphorically, down). That is why there is no sentence such as **tofang=et ro bendâz Ø miz* 'drop your gun on the table'.

be zamin xordam. > xordam Ø zamin.
'I hit the ground'.

with the verb *gereftan* 'to hold':

čerâ tofang=at ra be tarafe man gerefte'i? > čerâ tofangeto gerefti Ø tarafe man?
'Why have you pointed your gun to me?'

However, we should keep in mind that the verbs *residan* 'reach' and *raftan* 'go' cannot pave the way for such a change in case they are used in their idiomatic or metaphoric meanings:

(17) *bel'axare be ârezu=yam residam. > *belaxare residam Ø ârezu=m.*
'I finally reached my desire'.

*sa'y kon xodat ra be kelâs beresâni. > *sa'y kon xodet ro beresuni Ø kelâs.*
'Try to reach (the level of) the class'.

*dâr o nadâr=aš be yaqmâ raft. > *dâr o nadâr=eš raft Ø yaqmâ.*
'All his belongings were plundered'. (lit. 'went to nowhere')

*har če badi karde bud be sar=aš âmad. > *har či badi karde bud umad Ø sar=eš.*
'He suffered all the wrongdoings he had done'. (lit. 'all his wrongdoings came to his head')

Directional adverbs can also lead to such movements and deletions:

(18) *be bâlâ raftam/ be pâyin âmadam. > raftam Ø bâlâ/ umadam Ø pâyin.*
'I went up/ I came down'.

be in taraf raftam/ be ân taraf raftam. > raftam Ø in var, raftam Ø un var.
'I went to this direction/I went to that direction'.

be samte čap boro/ be samte râst biyâ. > boro Ø samte čap/ biyâ Ø samte râst.
'Go to the left side/come to the right side'.

The deletion of a preposition may also take place without any change in the word order, especially when the verbs are static:

(19) *če dar dastat ast? > či Ø dastete?*
'What are you holding?' (lit. 'What do you have in your hand?')

barâye šâm či dârim? > Ø šâm či dârim?
'What do we have for dinner?'

čerâ kafš _be_ pâyat nist? > čerâ kafš Ø pât nist?
'Why aren't you wearing shoes'? (lit. 'Why are shoes not on your feet?')

Apart from this, the use of clitics is an interesting syntactic issue of spoken Persian. Until recently, many of the Persian endings now known to be clitics were treated as affixes. But now, in the light of the new linguistic findings, we know that Persian and most of the contemporary Western Iranian languages make use of enclitic pronouns to mark objects for their verbs or possessors for their nouns. The major clitics of Persian comprise the verbal endings denoting possession: _ketâb=am_ (book=first person possessive), and objects: _zad-am=aš_ 'I hit him' (past of 'hit'-first person singular=third person singular). The _ezafe_ marker -_e_ and the definite marker -_i_ indefinite are the other enclitics of standard Persian or written Persian. But there are some other clitics that are the contracted forms of the free functional morphemes and are used solely in spoken Persian: -_â_ (as the contracted form of the plural suffix -_hâ_), -_am_ for _ham_ 'also', as well as -_o_, which is an allomorph both for _râ_ (the direct and specific object marker) and the conjunction _va_ 'and'. _Râ_ also has a contracted form, _r_, which acts as a clitic.

In spoken Persian syntax, when a pronoun ends in a consonant, the clitic for _râ_ is deleted. Compare:

(20) _man ham miâyam > man=am miyâm._   'I come too'.

  man râ ham bebarid > man=am bebarin.   'Take me too'.

  ânhâ râ ham mibarim > unâ=r=am mibarim.   'We will take them too'.

Another important issue in terms of the syntax of clitics in spoken Persian is the fact that they may also be attached to some adjectives, something very unlikely in written Persian:

(21) _nemixâham qiyâfeye šomâ xâ'en-hâ râ bebinam. > nemixâm qiyâfeye xâ'en=etun ro bebinam._
  'I do not want to see the faces of you traitors'.

  hâlam az didane qiyâfeye toye bišaraf be ham mixorad. > hâlam az didane qiyâfeye bišaraf=et be ham mixore.
  'Seeing the face of you wicked makes me sick'.

In the case of the emphatic constructions with *xod* 'self', personal clitics may even get attached to the adjectives:

(22) *begozârid xod-e ân bišaraf ham bedânad. > bezârin xod-e bišaraf=eš ham bedune.*
'Let that wicked personally know it, too'.

xod-e ân kesâfat ast. > xod-e kesâfat=eš-e.
'It is the same him, the filthy'. (lit. 'It's exactly his filthy self')

Other examples of specific spoken Persian syntactic constructions refer to the special uses of the particle *ke.* When the particle *ke* follows the first constituent of an utterance, it acts as an indignant asseverative: *to ke man=o košti* 'you (almost) killed me', *mâ ke bâ ham da'vâ nadârim* 'We (really) have nothing to fight over'. In the case of compound verbs, when *ke* is placed between the noun and the light verb, it indicates the time of the action and means 'when': *vaqti ke bozorg šodi, mifahmi > borozg ke šodi, mifahmi* 'You will know it when you grow up'; *vaqti ke tamâm šod, nešânat midaham > tamum ke šod, nešunet midam* 'I will show it to you when it is finished'. It also means 'of course' when placed after the first constituent as well: *pul ke dâštam, vali naxâstam xarj=eš konam* 'Of course I had money, but I didn't want to spend it'; *istgâhe âxar ke nist, vali bâyad piyâde šim* 'This is not, of course, the last station, but we have to get off' (see also Bâteni 1355/1976; Perry 2007: 985; Mahmoodi-Bakhtiari and Tâjâbâdi 1392/2013).

Sadat-Tehrani (2003) refers to another use of this particle as "indifference *Ke*-construction". Emerging from Jackendoff's theory of "parallel architecture", indifference *Ke*-construction has the following structure and denotes that the content of the verb has not been of importance to the speaker:

Verb$_i$ - *Ke*- Verb$_i$

For this structure (which is not seen in written Persian), we may cite the following examples:

(23) *nayumad ke nayumad, xodemun mirim.*
'It is not important that he didn't come. We go by ourselves'. (lit. 'He didn't come-that-he didn't come')

This construction may also denote that the action is totally terminated, and nothing can be done about it:

(24)   *raft ke raft.*
       '(I don't care) that he went'. Or 'He went (and did not return)'.
       (lit. 'He went that he went') (See also Mahmudi-Baxtiyâri and
       Tâjâbâdi 2013.)

In comparison to Sadat Tehrani's indifference *Ke*-construction, I would like to
propose an Indifferent *ham* Construction construction, in which the coordinator
ham acts like *ke*, with a difference in changing the stress pattern of the verbs.
That is to say, in this construction, the stress of the first verb is placed on its
last syllable, while it lands on the first syllable of the second one. The verbs are
mostly in simple past tense as they are supposed to denote a type of conditional
structures:

(25)   *raftì ham ràfti.*
       '(It is not important) if you go'. (lit. 'You went, you went too'.)

       xordìm ham xòrdim.
       '(There is no problem) even if we eat'. (lit. 'We ate, we ate too'.)

To the list of such symmetrical structures, we may add another one, with a
similar function with *na tanhâ . . . balke. . .* 'not only. . . but also. . .'. This spoken
Persian construction is as follows:

(26)   S₁-*hičči 'nothing'*-S₂ (with *ham* 'too')

       *na tanhâ komak nemikonad, balke aziyyat ham mikonad. > komak ke
       nemikone hičči, aziyyat ham mikone.*
       'He not only doesn't help, but also bothers'.

Finally, let's review some verbs that have additional meanings in spoken
Persian, and naturally, different valences in terms of their transitivity. Some
transitive verbs in written Persian can also be used as intransitive ones (with
metaphorically different meanings). Some examples are:

(27)   <u>*boridan*</u> 'to give up, to get tired' (lit. 'to cut'):
       inqadr kâr kardam ke dige boridam.
       'I worked so much, that I am so tired'.

       <u>*tamâm kardan*</u> 'to die' (lit. 'to finish')
       teflaki dišab tamum kard.
       'Poor fellow died last night'.

mâlidan 'to finish' (lit. 'to rub')
refâqate mâ dige mâlide.
'Our friendship is totally over'.

kešidan 'to endure, to tolerate' (lit. 'to pull')
dige nemikešam. az injâ miram.
'I do not stand it any longer. I will leave here'.

sâxtan 'to get along with, to tolerate' (lit. 'to make')
čâre'i nist, bâyad besâzi.
'There is no choice. You have to tolerate it'.

kam âvardan 'to run out of patience' (lit. 'to be little of something')
xeyli talâš kardam, ammâ dige kam âvordam.
'I tried so much, but I ran out of patience/energy'.

qâti kardan 'to (suddenly) get so angry' (lit. 'to mix')
ye vaqt qâti mikonam o mizanam tu un dahanet hâ!
'I may suddenly get angry, and hit you on your mouth, you!'

javâb dâdan 'to be of use' (lit. 'to answer')
sedâqat dige tu in zamune javâb nemide.
'Honesty does not work out these days'.

These verbs (in their spoken Persian use) are all intransitive verbs, while we know that they are transitive verbs in their common use of the language, with different meaning. We terminate our syntactic examples with these items, which are samples of the interface of syntax and semantics in modern Persian.

12 Semantic Alternations

Semantic alternations are not among those issues proposed by Ferguson (1959) in his theory of diglossia. However, providing some examples of several semantic readings of words and phrases in written Persian and spoken Persian can clarify the different uses of these two types even better.

Persian words show considerable semantic differences when they are used in spoken Persian. It is not possible to count all of them here. However, the examples below may provide a general view of such differences:

(28)   a.   The word *tâze* 'new' is also used as a conjunction in spoken Persian,
             meaning 'as a matter of fact, also': *in pârk kolli vasileye bâzi dare. Tâze
             estaxr ham dâre.* 'This park has many play tools. As a matter of fact, it
             has a pool, too'.

       b.   The adjective *âxar* 'final' (spoken Persian *âxe*) may also be used in
            spoken Persian as an exclamative marker to show complaint or
            surprise over something: *âxe čerâ râsteš=o nemigi?* 'Why (on earth)
            don't you tell the truth?' It also means 'because' in spoken Persian:
            xunašun nemiram, âxe bâhâš qahram 'I do not go to their house
            because I am not on speaking terms with him'.

       c.   The word *digar* (SP *dige*) 'else', has several meanings in spoken
            Persian other than its common written Persian use: *dige či goft?* 'What
            else did he say?'; and the indignant asseverative marker, as in *to dige
            harf nazan* 'you don't speak *at all*'; *dige nemitunam tahammol konam*
            'I cannot tolerate it *any longer*'; *qol midam dige dars bexunam* 'I promise,
            from now on I will study'; *dige az man komak naxây-â!* 'Don't you *ever*
            ask me for help!'; *boro dige!* 'go *then*!'; *mohandes mixâstan, man ham
            mohandesi xunde budam dige* 'They needed an engineer; *as you know*,
            I had studied engineering'; *in dige češe?!* 'What *the hell* is wrong with
            this?!'

       d.   The word *kolli* (general), with a transition of stress to the first syllable
            as *kólli*, means 'so much, so many' in spoken Persian: *kólli âdam jam'
            šodan* 'So many people have gathered'.

       e.   The structure *ham ke šode*, means 'even if' in spoken Persian: *piyâde
            ham ke šode, bâyad xodemun ro beresunim unjâ* 'We have to reach
            there, *even if* (we go) on foot'.

       f.   The conjunction *ham* 'also' means 'even if' in spoken Persian too:
            pesare ra'is ham bâši, nemituni beri tu 'even if you are the boss's son,
            you may not enter there'; *man be barâdaram ham komak nemikonam*
            'I do not even help my brother'.

       g.   The adverb *aslan* (spoken Persian *asan*) 'not at all' means 'altogether,
            at all' in spoken Persian: *qabl az nâhâr miyâd, albatte age asan biyâd*
            'He will come before lunch, if he comes at all'.

       h.   The adverb *hâlâ* 'now' can act as a discourse marker in spoken Persian
            in order to change the topic: *hâlâ man ye čizi goftam, to čerâ be del
            gerefti?* 'Well, I said something. Why are you so insulted'? (lit. 'Why do
            you take it to your heart?').

i. *Nâsalâmati* (lit. 'without health') means 'It seems that you have seemingly forgotten that': *nâsalâmati to mard-e in xune hasti!* 'Don't forget that you are the man of this house!'

j. The verb *oftâdan* 'to fall' means 'to invite oneself' in *dišab xuneye rezâ oftâde budim* 'We had invited ourselves to Reza's house'; and 'go' in *biyoft jolo, râh=o nešun bede* 'go forward, and show the way'.

k. The adjective *xâli* 'empty' paradoxically means 'full' in spoken Persian when it follows a word in *Ezafe* construction: *kabâb kubide čarbi-ye xâliye* 'Minced kebab is full of fat' (lit. 'only fat').

l. The word *šekl* 'picture' means 'similar' in spoken Persian: *čeqadr šomâ šekl-e hamin* 'How similar you two are'; *čerâ šekle gedâ hâ lebâs mipuši?* 'Why do you dress like beggars?'

m. The lexicalized verb *nagu* 'don't say' means 'while' in spoken Persian: *fekr kardam sâ'atam ro dozdidan, nagu tu daftaram jâ gozâšte budam* 'I thought my watch was stolen, while I had left it in my office'.

n. In case the phrase *yèk daf'e* 'once' is lexicalized, the word *yekdaf'è* (spoken Persian *yedafè*) means 'suddenly': *dâštim bâ ham harf mizadim ke yedafè zad zire gerye* 'We were talking when she suddenly burst into tears'.

o. The word *qadr* 'value' with the *Ezafe* marker (*qadde*) means 'as big as' or 'as much as': *tu xiyâbun ye muš didam, qadde ye gorbe* 'I saw a rat on the street, as big as a cat'.

p. In spoken Persian, two phrases *mes(l)e inke* 'like, as' and *ma'lum-e* 'It is clear that' are used to denote the meaning of 'apparently', or 'It seems that': *mese inke delet kotak mixâd* 'You are apparently looking for getting beaten', *ma'lume xodet badet nemiyâd beri unjâ* 'You do not seem to be reluctant to go there'.

q. The adverb *ettefâqan* 'accidently' may also mean 'as a matter of fact' in spoken Persian: *avval nemixâstam beram, ba'd fahmidam ke ettefâqan unjâ jâye mane* 'I did not want to go there at first, but later I realized that as a matter of fact, there was my place to go'.

r. The verb *nemixâd* (written Persian *nemixâhad* 'he doesn't want') means 'it is not necessary' in spoken Persian too: *mâ nemixâd qâtiye in qaziyye bešim* 'we do not need to get involved with this issue'; *šomâ nemixâd kâr konin* 'You are not supposed to work'.

s. The word *hame=aš* (spoken Persian *hamaš* 'all of it') is used as an adverb of frequency in spoken Persian too, meaning 'always': *hamaš qor mizane* 'he always nags'.

t. The phrase *ân-ham* (spoken Persian *un-am* 'that one, too') means 'especially' in spoken Persian: *in kâr=â zešte, unam az yek ostâd-e dânešgâh* 'These deeds are wrong, especially from a university professor'.

u. The phrase *haminjur* 'this way' also means 'consecutively, repeatedly': *haminjur harf mizad, saremun ro bord* 'he kept talking, he talked our heads off'.

v. The phrase *doròst-e* means 'that's right', but the word *dorostè* is an adverb meaning 'in one piece': *un mitune to ro dorostè qurt bede* 'he can swallow you in one gulp'.

w. *Balke* 'but also' may mean 'may be' or 'in the hopes that': *mixâd bere šahr, balke kâri peydâ kone* 'he wants to go to the town, maybe he can find a job'.

x. The lexicalized phrase *umadim o* lit. 'We came, and', means 'imagine', or 'suppose': *umadim o taraf kutâh nayumad, čekâr konim?* 'Imagine that the guy does not compromise, what should we do?'

13 Conclusion

I hope I have shown that the differences between spoken Persian and written Persian are not limited to some deletions in syntax or some sound shifts in phonology. I agree with Perry that Persian does not provide us with as many unique spoken Persian lexical items as Arabic, but keep Jeremiás's discussions in mind, too, that the unavoidable differences between the two varieties of Persian cannot be allocated within the study of the "dialects of Persian", especially now, when the population of the speakers of spoken Persian has grown, and the spoken Persian that was once the dialect of the people of Tehran has now become the spoken variety of Persian used by almost all Iranian Persian speakers. I have tried to show that Persian is diglossic in comparison to many other known languages in terms of the remarkable differences in its written and spoken forms. If we consider diglossia as a continuum, Persian may have a place on that continuum, although it may not necessarily be close to classical diglossic languages.

Perry (2003: 26) concludes his article by saying that the real question to be answered is not "How diglossic is Persian?", but "How did Persian avoid diglossia?" Given the above discussion, I dare to ask "To what extent is Persian diglossic?"

References

Bâteni, M. R. 1355/1976. Este'mâl-e *ke*, *dige*, *âxe*, *hâ* dar fârsi-ye goftâri [The use of *ke-*, *dige*, *âxe*, and *hâ* in spoken Persian]. *Majalleye Dâneškadeye Adabiyyât va 'Olum-e Ensâniye Dânešgâh-e Tehrân* 93–94. 257–271.

Boyle, John A. 1952. "Notes on the colloquial language of Persia as recorded in certain recent writings. *Bulletin of the School of Oriental and African Studies, University of London* 14 (3). 451–462.

Coulmas, F. 2013. *Sociolinguistics: The study of speakers' choices*, 2nd edn. Cambridge: Cambridge University Press.

Deyhim, G. 1368/1989. Gerâyeš hâ-ye âvâyi va vâji-ye fârsi-ye goftâri-ye tehrân [The Phonetic and Phonological Trends of the Spoken Persian of Tehran]. *Majalleye Zabânšenâsi* 6 (2). 97–105.

Ferguson, Charles A. 1959. Diglossia. *Word* 15. 325–340.

Ferguson, Charles A. Diglosia revisited. *Southwest Journal of Linguistics* 10. 34.

Fishman, J. 1967. Bilingualism with and without diglossia, diglossia with and without bilingualism. *Journal of Social Issues* 23 (2). 29–38

Frommer, Paul Robert. 1981. *Post-verbal phenomena in colloquial Persian syntax*. Los Angeles: University of Southern California dissertation.

Ghobadi, Chokofeh. 1996. La langue parlée reflétée dans les écrits. *Studia Iranica* 25 (1). 135–158.

Henderson, Michael M. T. 1975. Diglossia in Kabul Persian phonology. *Journal of the American Oriental Society* 95 (4). 651–654.

Hodge, C. T. 1957. Some aspects of Persian style. *Language* 33. 355–369.

Jahangiri, Nader. 1980. *A sociolinguistic study of Tehrani Persian*. London: University of London dissertation.

Jeremiás, Éva. 1984. Diglossia in Persian. *Acta Linguistica* 34 (3–4). 271–287.

Kaye, A. 2009. Arabic. In B. Comrie (ed.), *The world's major languages*, 560–577. London: Routledge.

Mahmudi-Baxtiyâri, B. 1388/2009. Namâyešnâme-ye *Ustâd Noruz-e Pineduz*: Manba'i Mohem barâye Motâle'e-ye Fârsi-ye Goftâri-ye Dore-ye Qâjâr [The play *Ustâd Noruz, the shoesmith*: An important reference for the study of spoken Persian in the Qâjâr era]. *Majalle-ye Pažuheš-e Olum-e Ensâni* (Bu Ali Sinâ University of Hamedân) 25. 87–111.

Mahmudi-Baxtiyâdi, B. & F. Tâjâbâdi. 1392/2013. Čand kârbord-e kalâmiye digar az ke- dar fârsiye goftâri [Some other pragmatic functions of *ke-* in spoken Persian]. *Motâle'âte Zabân va Guyešhâye Qarb-e Irân* 2. 103–116.

Perry, John. 2003. Persian as a homoglossic language. *Iran Questions Et Connaissances* in *Cultures Et Sociétés Contemporaines* III. 11–28.

Perry, John. 2007. Persian morphology. In Alan S. Kaye (ed.), *Morphologies of Asia and Africa*, vol. 2, 975–1019. Winona Lake, IN: Eisenbrauns.

Sadat-Tehrani, Nima. 2003. The indifference *ke* construction in modern Persian. *Linguistica Atlantica* 21. 137–151.

Samareh, Yadollah. 1977. *A course in colloquial Persian (for foreigners)*. Tehran: University of Tehran Press.

Schiffman, H. 1998. Diglossia as a sociolinguistic situation. In F. Coulmas (ed.), *The handbook of sociolinguistics*, 205–216. Malden: Blackwell.

Schmidt, Richard Wilbur. 1974. *Sociostylistic variation in spoken Egyptian Arabic: A re-examination of the concept of diglossia*. Providence: Brown University, dissertation.

Vahidiyân, Taqi. 1343/1964. *Dastur-e Zabân-e 'Âmiyâne-ye Fârsi* [A grammar of spoken Persian]. Tehran: Amir-Kabir.

Lewis Gebhardt

11 Accounting for *yek ta* in Persian

Abstract: Persian has been described as using numeral classifiers in expressions such as *se ta ketab* 'three CL book' 'three books'. However, unlike classifiers in many languages, *ta* is not used with the numeral for 'one'. Based on a syntactic theory of feature checking in the syntax, I argue that Persian *ta* has some characteristics of being a number marker but also has properties of numeral classifiers.

Keywords: Persian, number, classifiers

1 Introduction

Persian uses the numeral classifier *ta* in expressions that enumerate count nouns (Mahootian 1997: 195; Ghomeshi 2003: 55; Lambton 1974: 43–44). Example (1a) illustrates a classifier construction in Persian, paralleling example (1b) in Mandarin, another classifier language.[1] Note that *ta* is not used with mass nouns, as in (1c).

(1) a. *do/se/dæh   ta   ketab*[2]
         two/three/ten   CL   book
         'two/three/ten books'

     b. *yi/liang   ge   xuesheng*
         one/two   CL   student
         'one/two student(s)'
         (Sonya Chen, p.c.)[3]

     c. **cænd   ta   cai*
         how.many   CL   tea

With regard to (1c), it's important to note that the expression is ungrammatical on the intended reading of quantifying a mass volume of tea. However, *cænd ta*

Lewis Gebhardt, Northeastern Illinois University

čai can be used to ask how many cups, or other units, of tea, where the intended quantification is over the number of cups. This is similar to the general ban on pluralizing mass nouns in English, where nonetheless one can ask "How many teas?" with the intended meaning of how many cups of tea.

While Persian speakers favor using *ta* in colloquial speech, it isn't obligatory for grammaticality, and in writing and more formal speech *ta* is typically omitted. Even in colloquial Persian, *ta* is sometimes not used and in some contexts, such as with very large numerals, the use of the classifier is awkward or ungrammatical, as in (2a), a fact consistent with the optional or obligatory absence of classifiers with large numerals in other languages (Aikhenvald 2000: 100). *Ta* is not used with quantifying words other than numerals, such as *xeyli* and *ziyad* in (2b), both meaning 'many, a lot of', and *hær* 'each' in (2c), though it does appear in the expression *cændta* 'how many' in (2d).

(2) a.   *6.022 × 10²³ molkul*   /   **6.022 × 10²³   ta   molkul*
6.022×10^{23} molecule 6.022×10^{23} CL   molkul
'6.022×10^{23} molecules'

   b.   **xeyli   ta   ketab   /   xeyli   ketab*
many   CL   book   many   book
'many books'

   c.   **hær   ta   ketab   /   hær   ketab*
each   CL   book   each   book
'each book

   d.   *cændta   ketab*
how.many   book
'how many books?'

In this regard, Persian contrasts with Mandarin, which allows a classifier with non-numeral quantifying elements.

(3) *hen   duo   (ge)   xuesheng*
very   many   CL   student
'many students'
(Sonya Chen, p.c.)

Persian doesn't use classifiers with demonstratives on their own, as in (4a), though it's acceptable to use a classifier with the demonstrative if a numeral

appears, as in (4b). This contrasts, again, with Mandarin, which allows co-occurrence of classifier and demonstrative, as in (4c)[4]; Cantonese also allows a classifier and demonstrative, as in (4d).

(4) a. *un   ta   pærænde
         that   CL   bird

b. un   se   ta   pærænde
    that   three   CL   bird
    'those three birds'

c. zhei   ben   shu
    this   CL   book
    'this book'
    (Cheng and Sybesma 1999: 510n17)

d. li¹/go²   bun²   syu¹   hai⁶   ngo⁵-ge³
    that/this   CL   book   be   1SG-GE
    'that/this book is mine'
    (Cheng and Sybesma 2012: 642)

One other point of contrast is that Persian *ta* never has a determiner-like function, though in Chinese languages classifiers can function like definite and/or indefinite articles (see Cheng and Sybesma [2005] for numerous examples and discussion).

Finally, there are said to be other classifiers in Persian that are less commonly used such as *nafar* for people and *jeld* for books and volumes (Lambton 1974: 43–44), although Gebhardt (2009) argues that those items are classifier modifiers rather than classifiers per se. The distinction is unimportant for the scope of this article.[5]

Summarizing, I focus on *ta* and assume for simplicity in argumentation but, at the slight risk of oversimplification, that *ta* is required for numerals greater

---

4 Cheng and Sybesma point out that *zhei* 'this' might be analyzed as a demonstrative plus classifier (Cheng and Sybesma 1999: 510 n. 17).

5 For the record, besides *do ta ketab* 'two CL book', Persian allows expressions like *do jeld ketab* 'two "CL" book', where *jeld* is used for books or volumes and apparently replacing *ta*; I put gloss *jeld* as "CL" (in scare quotes because of the argument that *jeld* is not really a classifier. Persian also allows *do ta jeld ketab* 'two CL "CL" book'. Gebhardt (2009) argues that in such cases *ta* is the closest thing to a classifier in Persian while *jeld* and some other words merely modify, by adjunction, *ta* itself, which in this case and others is not obligatory. Other languages lexicalize both the classifier function and the semantically descriptive function into a single item, as in Mandarin *san zhang zhuozi* 'three CL.Furniture table'.

than *yek* 'one' and is barred from appearing with *yek*, putting aside the expression *čændta* 'how many'. It is the narrow purpose of this article to argue that *ta* isn't exclusively a classifier and that it also has some properties of the category number, plural in particular. It will be shown that, being plural, *ta* is then barred from appearing with singular *yek*. In a broader picture, the idea that *ta*'s category is squishy between classifier and number is in contrast to theories that see classifier and number as distinct categories. However, at the same time those theories also assert or assume that classifier and number are, more or less, mutually exclusive categories in complementary distribution. Complementary distribution suggests that they might be two manifestations of a single more abstract category, and this is part of the approach here.

The article is organized as follows. Section 2 follows with background literature on classifiers. In section 3, I lay out syntactic assumptions and identify features involved in the syntactic operation of Merge regarding items in the determiner phase, whereby syntactic features of a lexical or functional item license that item to combine with other elements. In section 4, I address the specific problem of accounting for why the numeral *yek* 'one' doesn't occur with the classifier *ta*, and I briefly conclude and look forward in section 5.

2 Background

A well-known proposal to explain the use of classifiers is Chierchia's (1998) nominal mapping parameter. Theoretically, Chierchia makes the common assumption that a determiner phrase, headed by D, dominates the noun phrase, overtly at least in some languages. Empirically, Chierchia points to attested differences between classifier languages like Mandarin and non-classifier languages like English and French. Generally, classifier languages lack articles and lack general plural marking (Chierchia 1998; also see below). Since classifier languages lack articles, in contrast to English-type languages with articles, it is claimed that nouns don't require a determiner to be arguments and can appear bare in argument position, as in (5a). English-type languages, however, require an article in count singular contexts, as in (5b). Further, in classifier languages plural marking is not required in plural contexts, as in (5c), again in contrast to English-type languages where the plural must appear, as in (5d). In fact, to the degree that it exists at all, plural morphology is often highly restricted in classifier languages.

(5) a. *gorbe   xabid-e*
      cat   sleep.Past-3.SG
      'The cat is sleeping'

    b. *(the/a) cat is sleeping

    c. *se   ta   gorbe*
      three   CL   cat
      'three cats'

    d. three cat-*(s)

Chierchia notes that the nonappearance of plural and articles in classifier languages has a parallel in English mass nouns, which ordinarily don't take plural, as in (6a), and don't require an article to be a verbal argument, as in (6b).

(6) a. *We pumped airs into the tires

    b. (Some) air has been pumped into the tires

Following from earlier analyses of count and mass (Krifka [1995] and Gillon [1992], among others), Chierchia proposes that all Mandarin nouns are in some sense mass, while some English nouns like "cat" and "student" are count. Mass nouns, in Chierchia's analysis, don't need articles to be used as arguments and they can't be pluralized. Further, since Mandarin nouns are mass, they require a classifier to make countable units, so the use of a word like Persian *ta* is similar to the use of an expression like "loaves (of)" with English mass nouns as in the expression "three loaves of bread". This leads to Chierchia's nominal mapping parameter, which states that languages parametrically set their nouns as mass or count and, correspondingly, arguments or predicates. If a language's nouns are all mass, then they can appear as bare arguments, don't have regular plural, and require a classifier in contexts of enumeration. If a language's nouns are count, then they cannot appear bare in argument positions and require a plural in semantically plural contexts. In Chierchia's view, Mandarin nouns are all mass arguments, French nouns are count predicates; English-like languages have both count and mass nouns.

    Borer (2005: 87–107) presents a number of empirical and theoretical objections to Chierchia's analysis. An important point of departure for her is that, contra Chierchia, languages cannot be cleanly cloven into those that have classifiers and no plural and those that have plural and no classifiers. Armenian, for example, has both. So it can't be that languages set their nouns as mass or

count across the board. Nonetheless, Borer observes, there is a complementary distribution *within* Armenian such that while an enumeration expression may have only a classifier, as in (7a), *or* only plural, as in (7b), *or* neither, as in (7c), a classifier cannot appear with plural and classifier in the same construction, as in (7d); all mean or are intended to mean 'two umbrellas'.

(7) a. *yergu   had   hovanoc*
         two   CL   umbrella

  b. *yergu   hovanoc-ner*
       two   umbrella-PL

  c. *yergu   hovanoc*
       two   umbrella

  d. **yergu   had   hovanoc-ner*
       two   CL   umbrella-PL
       (adapted from Borer 2005: 117–118)

Still, both Borer's and Chierchia's proposals depend partly on the idea that classifiers and number serve similar or parallel functions in unitizing a noun's denotation into countable units, an approach earlier investigated by Doetjes (1997). Very briefly, Borer's proposal is that a classifier phrase can be headed by either a classifier or number. The function of the head of the classifier phrase is to individuate a root, which is unspecified for mass or count. Since a classifier and number compete for the same head, their co-occurrence is ruled out in Armenian, as Borer predicts.

While Chierchia's analysis argues for the denotation of nouns across a language and Borer's analysis focuses on particular constructions, both Chierchia and Borer assume a kind of mutual exclusivity of classifiers and number markers. However, there is ample evidence that languages don't come in pure types as Chierchia would suggest. In three studies of language samples, Greenberg (1972), Sanches and Slobin (1973), and T'sou (1976) show that, while individual languages may tend toward having classifiers or number, the complementarity is only a rough one, with many languages having both. *The World Atlas of Language Structures* (WALS) bears similar results (Dryer and Haspelmath 2013). Crossing WALS features 33A and 34A, we come up with the following distribution of languages that have classifiers and/or plural marking.

Table 1: Numbers of languages that have classifiers/nominal plural (from WALS)

	No nominal plural	Nominal plural exists
No classifiers	8 (7.0%)	80 (70.2%)
Classifiers exist	4 (3.5%)	22 (19.3%)

Thus, about a fifth of languages in the WALS sample have both classifiers and plural, a fact that at least in a superficial sense contradicts predictions from Chierchia's analysis. Borer's analysis can accommodate the facts in Table 1, since for her the important restriction is that they not appear in the same construction, although both can be available within a language. But troubling for her analysis is that languages with both can and do use them in the same construction, contrary to her prediction. Example (8) is from Paiwan, (9) from Itzaj Maya, (10) from Tariana, (11) from Akatec, and (12) from Jacaltec.

(8) **_ma_**-*telu*   *a vavaya**vavaya**n*
    CL-three   A girl.Redup
    'three girls'
    (Tang 2004: 385)

(9) *ka'=***tuul***   *im-mejen*   *paal-**oo**'-ej*
    2=CL.Animate   is.a-small   child-PL-Top
    'my two small children'
    (Hofling 2000: 228)

(10) *duha*   *inaɾu*   *kanapeɾi-pidana*
     Art.Fem   woman   give.birth-Rem.P.Rep

     *ñham-**epa***   *emi-**peni***
     two-NumCL.human   youngster-PL
     'The woman gave birth to two children'
     (Aikhenveld 2003: 94)

(11) *kaa-**(e)b**'*   *poon*
     two-NumCL   plum
     'two small plums'
     (Zavala 2000: 124)

(12) *xwil*   *ca-**waŋ***   **_heb_**'   *no'*   *winaj*
     I.saw   two-NumCL   PL   NounCL   man
     'I saw two men'
     (Craig 1977: 137)

Persian too can use both the classifier and the plural marker *-ha*, which also indicates definiteness/specificity.

(13)   *pænj ta   gorbe-ha*
       five CL   cat-PL
       'the five cats'

Without the suffix *-ha*, the example in (13) would have an indefinite interpretation, 'five cats'. Most Persian speakers I have consulted in previous research find sentences like (13) with both ta and -ha acceptable. Co-occurrence of the classifier and plural marker is also possible in other Persian languages, such as Tajik.

(14)   a.   бист   **нафар**   студент-**он**
            twenty   CL   student-PL
            'twenty students'
            (Ido 2005: 37)

       b.   дар мактаб-и   мо   122 **нафар**   пионер-он   **ва**
            in school-EZ   us   122 CL   pioneer.Masc-PL   and

            пионерка-гон ″астнд
            pioneer.Fem-PL are
            'There are 122 pioneer boys and pioneer girls in our school'
            (Perry 2005: 163)

Since the purported mutual exclusivity of classifiers and number is not absolute, either across languages or within languages, a proper treatment of their distribution vis-à-vis each other requires a more general theory. In the next section I propose a primarily syntactic account based on the syntactic checking of features that may bundle in various ways. Depending on the precise bundling of features, an item may be a "classifier", "number", or something in between.

3 Syntactic Assumptions

3.1 Merge

I assume that items in the lexicon are bundles of features and that these bundles are combined through the operation of Merge. Features may be semantically interpretable and contribute meaning or they can be uninterpretable, in which case they are purely for syntactic computation. Since an uninterpretable feature [u-F]

can't be interpreted at the interface of syntax and, especially, semantics, after Merge [u-F] acts as a probe that seeks a matching goal, [F], its interpretable partner. Following Match, [u-F] is checked by [F] and eliminated from the derivation. If uninterpretable features are not eliminated by the end of syntax, the derivation is said to crash, i.e., to result in an illicit structure. Checking is often assumed to be between a [u-F] probe and its c-commanded [F] goal. Finally, when a lexical item merges with another lexical item or with an already merged syntactic object (a phrase), the newly formed object is labeled, typically with the category or feature of the head of the merged item. (For summary overviews of Merge, see Chomsky [2008: 137–147], among many others, and Hornstein [2009: ch. 3]. See Baker [2008], among others, on the possibility of the goal [F] c-commanding the probe [u-F].)

As an example, take the English article "the" and the noun "kitty". Assume that the feature bundle for "the" comprises at least [u-N, Def], indicating that it contributes definiteness to the meaning of the merged expression and that it will probe for something with the feature [N], i.e., a noun. The item "kitty" is minimally [N, KITTY], where [N] is the interpretable feature indicating the nominal category of the item; [KITTY] is a semantically interpretable feature or set of features specifying felineness. The two items merge as in (15).

(15)

[u-N, Def]   [N, KITTY]
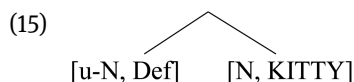
In (15), [u-N], which c-commands N, matches N, then checks with N and is eliminated from the derivation, the strikeout indicating that it has been checked, as in (16).

(16)

[~~u-N~~, Def]   [N, KITTY]
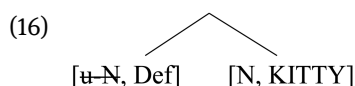
The bundle [u-N, Def] is a determiner and heads the merged phrase; it thus labels the new phrase, in traditional terms, a determiner phrase, or DP.[6] When sent to the interfaces, this constituent will get its pronunciation, [ðəkʰæt], and its semantic interpretation, something to the effect of "the particular kitty known to both speaker and hearer".

---

6 It is often assumed that lexical roots are uncategorized in the lexicon, to be later categorized with a dedicated node such as *n*, *v*, etc., as in Distributed Morphology (e.g., Harley and Noyer 1999; Embick and Marantz 2008), or are categorized in the syntax as in Borer (2005). To streamline exposition, however, however, I will simply indicate nouns as having an inherent nominal feature [N].

3.2 Some features of heads within DP

The proposed features that go into the bundles are similar to and overlap with those in a feature geometry proposed by Harley and Ritter (2002), where the features are hierarchically arranged. In (17), adapted from Harley and Ritter (2002: 486), a referring pronoun has at least a [participant] feature, which specifies first and/or second person, and an [individuation] feature, which specifies for number. The [participant] feature is further specified as [speaker] and/or [addressee] while [individuation] may be, simply put, [group], which is roughly plural, or [minimal], which is, more or less, singular. The features are arranged as in (17), and a concrete example is for the English pronoun "we" in (18).

(17)   Referring Expression (= pronoun)

    [participant]   [individuation]

[speaker] [addressee]   [group]   [minimal]

(18)   *we*

   [participant]   [individuation]

     [speaker]   [group]
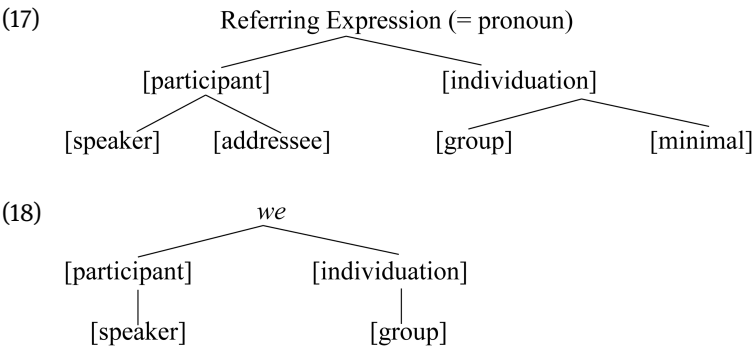
Note that the presence of lower features entails the presence of their dominating features. Also, other features are available that aren't present in (17) or (18) make reference to animacy and gender: a [feminine] feature would be present for "she", for example.

Harley and Ritter's feature geometry is supposed to capture crosslinguistic facts as well. For example, languages are more likely to have singular and plural morphemes than a dual morpheme, and if a language does have dual, it will likely have a singular/plural distinction as well. Therefore, the specification of dual must be lower in the feature tree. This particular fact they finesse with [individuation] having both the subfeatures [minimal] and [group], together indicating that the pronominal form is the minimum plurality of twoness.

Here I adapt Harley and Ritter's overall schema, though I present more features for other elements in the DP, focusing on number markers and classifiers. Again, at the head of the nominal projection, leaving out Distributed Morphology details, is a noun. Nouns have the category feature [N] and whatever semantic

features are needed for interpretation. I assume a number phrase, headed by a morpheme that indicates singular/plural (and presumably dual) (Ritter 1991, 1992). Since number merges with nouns, it has an uninterpretable feature [u-N] along with its feature-geometric feature and value of [individuation: minimal] or [individuation: group]. So a plural like English -s is specified by the feature bundle in (19). I assume a null morpheme for the English singular (20).

(19) -s [u-N, individuation: group]

(20) -Ø [u-N, individuation: minimal]

On top of number phrase are determiner phrases. Historically, before the introduction of a number phrase, Abney (1987), following from Brame (1981, 1982), Horrocks and Stavrou (1987), Szabolcsi (1981, 1984, 1987, 1994), and others, had proposed a functional projection headed by a determiner, whose complement was the NP, as in (21a). With the later incorporation of a number phrase, the DP took the form of (21b).

(21) a.

```
        DP
       /  \
          D′
         /  \
        D    NP
```

b.

```
        DP
       /  \
          D′
         /  \
        D   NumP
            /  \
           Num′
          /  \
        Num   NP
```
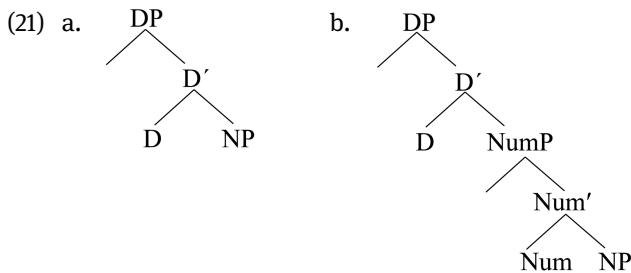
However, based on semantic facts and distributional criteria, several analyses have argued that there are at least two D levels. Going back to at least Bowers (1975), Jackendoff (1977), Milsark (1979), and Keenan (1987), it seemed that determiners were of at least two types. Among the differences between them is that it's harder to have a determiner phrase headed by a definite article in existential constructions, as in (22c),[7] although indefinites are possible in the same position (22a, 22b). Further, strong determiners can precede weak ones but not vice versa, as shown in (22d).

---

7 In (22), "there" is existential, not deictic. In fact, though restricted, the marked interpretation is possible.

(22)   a.   There are three books on the table.

       b.   There's a book on the table.

       c.   ?There are the tree books on the table.

   d.   the few kitties / *few the kitties

In short, determiners such as "the" and "each" were deemed "strong" while "few", numerals, and others were "weak". The distinction between them requires the presence of at least two determiner positions, as in Zamparelli (1995). Gebhardt (2009) called *the*-type determiners strong quantifying determiners and numeral-type determiners weak quantifying determiners. Each kind occupied the head of a separate phrase such that one version of an expanded DP is as in (23).

(23)

SQP
   the   WQP
       a, some,   NumP
       numerals   *-s*   NP
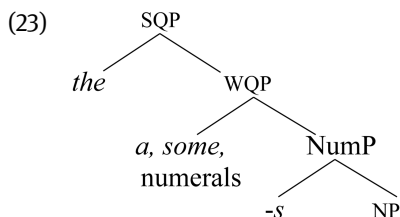
Since the focus of this article is number and classifiers, I limit the analysis to only as high as Weak Quantifier Phrase since that phrase's head is what merges with Number P. Hence, my focus is indefinites.

Since a numeral like "three" merges with a plural noun, the uninterpretable feature of "three" must therefore be [u-group], which checks with [group] on the plural marker. The numeral also provides interpretable quantification, which is indicated as [q]. So far the feature bundles we have are in (23), and an example of iterated Merge is in (24). Note that for [group] and [minimal] since these two features entail the presence of [individuation], [individuation] is for simplicity omitted from the feature bundle. In (24) only heads and complements are indicated; it's assumed that Spec positions do not project.[8]

---

8 Baker (2003) argues that, in the simple case, the presence of Spec appears for verbs and is the distinguishing syntactic feature of a verb.

(24) a. nouns: [N]

      plural: [u-N, group]

      singular: [u-N, minimal]

      numerals: [q, u-group]

b.

```
                    WQP
            _____/‾‾‾‾_____
         three                  NumP
       [q, u̶-̶g̶r̶o̶u̶p̶]          __/‾‾\__
                              -s            cat
                         [u̶-̶N̶, group]      [N]
```
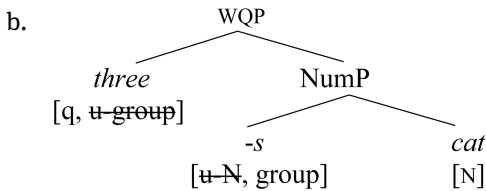
In (24b), *-s* has merged with "cat" and checked its uninterpretable [u-N] feature. Then, "three" has merged with NumP and in turn checked its uninterpretable [u-group] feature.

To the degree that number and classifiers are mutually exclusive, that mutual exclusivity suggests a commonality between the two categories. In Borer's analysis, in fact, each potentially occupies the head of a classifier phrase, although since they compete for that position, only one of them may appear. In this article, number markers and classifiers share syntactic features and thus may be more or less alike. In those languages that discourage the co-occurrence of number and a classifier, the relevant features for each must account for this fact. However, it's also in principle possible for a classifier to co-occur with a plural marker as long as there's no clash in features and that, as always, Merge results in eliminating any uninterpretable features.

The literature mentioned in section 2 also often argues that, in languages that have them, classifiers serve to somehow individuate the noun into countable units. And while classifiers may have functions other than just intermediaries between numerals and nouns, their canonical function is in numeral + noun constructions. The individuation function can be accommodated under Harley and Ritter's [individuation]. And if we assume the general case that classifiers have noun complements, they are [u-N]. As suggested in (1b), the Mandarin classifier is insensitive to whether the noun is singular or plural; it only seeks the general [individuation] feature rather than one of its potential [minimal] or [group] daughters.

(25) classifier: [u-N, individuation]

A classifier takes an NP to yield a classifier phrase (CLP). In turn, the numeral head is a weak quantifying determiner that takes a CLP complement to yield

WQP. The feature in the numeral that assures Merge with CLP is [u-individuation], which matches [individuation] in the classifier; the numeral also has the quantification feature [q]. Successive Merge is illustrated in (26b) for the Persian expression in (26a).

(26)　a.　*do*　　*ta*　　*danešju*
　　　　　two　CL　student
　　　　　'two students'

　　　b.

```
                        WQP
             ┌───────────┴───────────┐
            WQ                        CL
            do                ┌───────┴────────┐
        [u-indiv]            CL                NP
           [q]               ta                △
                           [u-N]            danešju
                          [indiv]             [N]
```
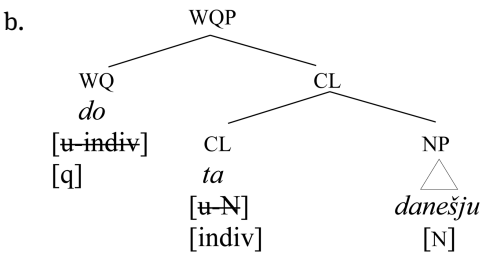
We now have in place a basic set of computational features involved in iterated merge of items within DP for numeral + CL + noun constructions. The following section provides some refined details to account for additional facts, particularly the ungrammaticality of Persian *yek* 'one' and the classifier *ta* occurring together.

4 Variant feature bundles and ruling out *yek ta*

For the data investigated so far, I have proposed that the items in the categories below comprise the feature bundles as follow.

(27)　nouns:　　　[N]
　　　plural:　　　[u-N, group]
　　　singular:　　[u-N, minimal]
　　　numerals:　　[q, u-group]
　　　classifier:　　[u-N, individuation]

First of all, languages vary in which of the items in (27) they have in their lexicons. English has nouns, numerals, plural, singular (although singular in English has no phonetic content) but it lacks a classifier. Persian has nouns, numerals, a

plural marker, and a classifier. Mandarin has nouns and classifiers, but no singular/plural distinction, at least not as in English.[9]

Like English, Persian singular has no pronunciation. Persian does have a plural marker, but it's used for definite nouns.[10]

(28)   *gorbe-ha*
       cat-PL
       'the cats' / #'cats'

Therefore, while it's correct enough to say that both Persian and English have a plural marker, the Persian plural morpheme differs in feature content from English *-s*; their feature distinction is represented as in (29).

(29)   a.   Persian *-ha*:   [u-N, group, Def]

       b.   English *-s*:   [u-N, group]

That is, English and Persian have bundled available features in different ways.

Unlike Mandarin *ge*, the Persian classifier *ta* is sensitive to cardinality, occurring with numerals other than *yek* 'one'. Therefore, *ta* must contain the feature [group]. So for Persian we revise feature makeup of the classifier in (27) as that in (30). Plural numerals must therefore be [u-group].

(30)   Persian *ta*:   [u-N, individuation, group]

       plural numerals: [q, u-group]

That is, *ta* has features of both a classifier and number. Numerals must therefore be [u-group].

I assume that, as in English, Persian has a null singular marker that nonetheless has syntactic content. It seeks nouns and has the number feature of [minimal], as above.[11]

---

9 There are Mandarin morphemes that contain plural meaning, such as *-men*, but these are restricted in semantics and are not generalized plural markers. The point is that in Mandarin plural marking is not obligatory for semantically interpreted plurality. Since the focus here is Persian, I overlook Mandarin plurals and adopt Chierchia's characterization that Mandarin doesn't have "true" plural marking (Chierchia 1998: 355).

10 More precisely, *-ha* is specific rather than definite (e.g., Karimi 1999: 8). Since the specific/ definite difference isn't important here, I will indicate the feature as [Def].

11 Persian can indicate singular indefiniteness with the numeral *yek* and/or an indefinite suffix *-i*: *yek gorbe* / *yek-i gorbe* 'a cat'. A more complete syntactic account will accommodate the availability of these markers in the DP complex, but I put those aside here.

(31)   Persian Ø singular: [u-N, minimal]

Persian numerals but not other quantifying words like *xeyli* 'much, many' require a classifier. So they contain the computational feature [u-individuation]. Summarizing for Persian is the list in (32). And since numerals want not only an individuated noun but one that's an absolute quantity (with a numeral) rather than a relative quantity ("few", "many", etc.), numerals are specified as [u-abs] seeking an [absolute] feature in *ta*. Summarizing, we have the following Persian items with their feature makeup.

(32)   nouns: [N]

       null singular   [u-N, minimal]

       -ha: [u-N, group, Def]

       ta: [u-N, individuation:absolute, group]

       yek: [q, u-minim, u-absolute]

       other numerals: [q, u-group, u-absolute]

The structure for an expression with a numeral other than *yek* (33a) is in (33b), with the probes' uninterpretable features checked.

(33)   a.   *do   ta   danešju*
            two   CL   student
            'two students'

       b.

The structure of an expression with *yek* (34a) is in (34b), again with uninterpretable features checked.

(34)   a.   *yek   danešju*
            one   student
            'one student'

b.

```
                    WQP
           ┌─────────┴──────────┐
          WQ              CL/NumP
          yek           ┌────┴─────┐
          [q]           ∅          NP
        [u-minim]     [u-N]        △
        [u-abs]      [minim]     danešju
                     [abs]        [N]
```
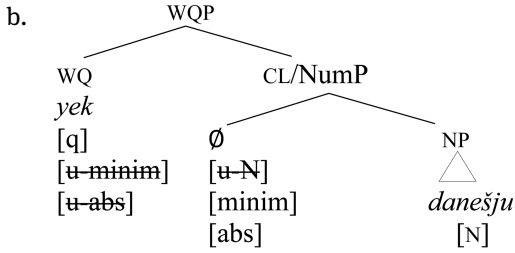
The ungrammaticality of *yek* occurring with *ta* (35a) is evident in (35b), with the feature [u-minimal] left unchecked.

(35) a. *yek   ta   danešju*
          one   CL   student

b.

```
                    *WQP
           ┌─────────┴──────────┐
          WQ              CL/NumP
          yek           ┌────┴─────┐
          [q]           ta         NP
        [u-minim]     [u-N]        △
        [u-abs]       [abs]     danešju
                     [group]
```
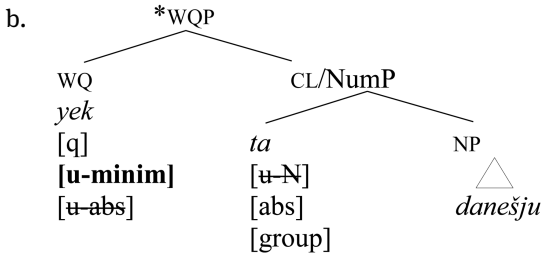
Similarly, to the degree that *ta* is required with plural numerals in colloquial speech, the ungrammaticality of (36a) is illustrated in (36b), where [u-group] is unchecked.

(36) a. *do   danešju*
          two   student

b.

```
                    *WQP
           ┌─────────┴──────────┐
          WQ              CL/NumP
          do            ┌────┴─────┐
          [q]           ∅          NP
        [u-group]     [u-N]        △
        [u-abs]       [abs]     danešju
                     [minim]
```
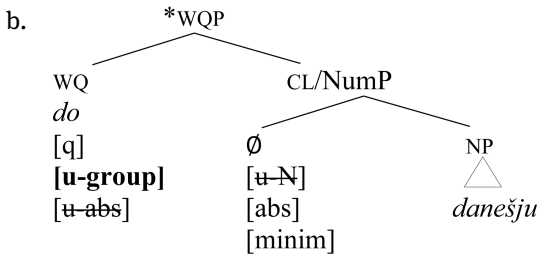
5 Conclusion

The ungrammaticality of **yek ta* is easily accounted for within a standard framework that holds items from the lexicon to be bundles of features, some of which are purely syntactic for computational purposes that must be checked and eliminated by the end of the derivation. *Yek* 'one' is incompatible with the classifier *ta* because *yek*'s uninterpretable [u-minimal] is never checked since *ta* is [group] and not [minimal]. Since *ta* is [group], it is essentially, in Persian, a plural marker and not really a classifier in the sense that Mandarin *ge*, insensitive to cardinality, is a classifier. But since *ta* occurs only in context of a numeral, neither is it a precisely plural marker like English *-s*, which is used in all plural contexts regardless of whether a numeral is present. On the one hand it seems that *ta* belongs to a mixed category, but since lexical items are frequently difficult to categorize discretely, *ta*'s bicategorial nature should be no surprise. Squishy or overlapping categories are standardly treated as subcategorization details. The approach taken in this article treats the subcategorization via the features inside feature bundles, and it should be no surprise that a language can take presumably universally available features and bundle them in different ways for the particular items in its lexicon. Abstractly, for features [a], [b], [c], and [d] relevant for items in the determiner phrase, one language may lexicalize [a, b] and [c, d], another language may lexicalize [a] and [b, c, d], while a third language may lexicalize [a, b] and [c] while leaving feature [d] unattached to any phonetically realized morpheme.

The focus in this article is the incompatibility of *yek* with *ta*. Other issues remain, however, such as how plural specific *-ha* works out in the derivation, not to mention the indefinite marker *-i*. Quantifying determiners such as *xeyli* were introduced in passing as a point of reference, but details how *xeyli* undergoes Merge must also be presented. Another issue is the differential case marker *-ra*, which only appears on specific direct objects, and what its features are and how its features check in the syntax. Finally, since the focus in this article is on Persian, I glossed over details of Merge in Mandarin, English, and other languages. However, the proposal makes clear predictions about features and how they are bundled across languages. Further research on Persian and other languages will corroborate or falsify the proposed theory.

References

Abney, Steven 1987. *The English noun phrase in its sentential aspect*. Cambridge, MA: MIT dissertation.

Aikhenvald, Alexandra. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.

Aikhenvald, Alexandra. 2003. *A grammar of Tariana*. Cambridge: Cambridge University Press.

Baker, Mark. 2003. *Lexical categories: Verbs, nouns, and adjectives*. Cambridge: Cambridge University Press.

Baker, Mark. 2008. *The syntax of agreement and concord*. Cambridge: Cambridge University Press.

Borer, Hagit. 2005. *In name only*. Oxford: Oxford University Press.

Bowers, John. 1975. Adjectives and adverbs in English. *Foundations of Language* 13. 529–562.

Brame, Michael. 1981. The general theory of binding and fusion. *Linguistic Analysis* 7 (3). 277–325.

Brame, Michael. 1982. The head-selector theory of lexical specification and the nonexistence of coarse categories. *Linguistic Analysis* 10 (4). 321–326.

Cheng, Lisa Lai-Shen & Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30 (4). 509–542.

Cheng, Lisa Lai-Shen & Rint Sybesma. 2005. Classifiers in four varieties of Chinese. In Guglielmo Cinque & Richard Kayne (eds.), *The Oxford handbook of comparative syntax*, 259–292. Oxford: Oxford University Press.

Cheng, Lisa Lai-Shen & Rint Sybesma. 2012. Classifiers and DP. *Linguistic Inquiry* 43 (4). 634–650.

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6. 339–405.

Chomsky, Noam. 2001. Derivation by phase. In M. Kenstowicz (ed.), *Ken Hale: A life in language*, 1–52. Cambridge: MA: MIT Press.

Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos P. Otero & Maria Luisa Zubizarreta (eds.), *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*, 133–166. Cambridge, MA: MIT Press.

Craig, Collette Grinevald. 1977. *The structure of Jacaltec*. Austin: University of Texas Press.

Doetjes, Jenny. 1997. *Quantifiers and selection: On the distribution of quantifying expressions in French, Dutch and English*. Leiden: Leiden University dissertation.

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info (accessed 28 August 2015).

Embick, David & Alec Marantz. 2008. Architecture and blocking. *Linguistic Inquiry* 39 (1). 1–53.

Gebhardt, Lewis. 2009. *Numeral classifiers and the structure of DP*. Evanston: Northwestern University dissertation.

Ghomeshi, Jila. 2003. Plural marking, indefiniteness, and the noun phrase. *Studia Linguistica* 57 (2). 47–74.

Gil, David. 2013. Numeral classifiers. In Matthew Dryer & Martin Haspelmath, (eds.), *The world atlas of languages structures online*. Max Planck Institute for Evolutionary Anthropology, Leipzig. http://wals.info (accessed 15 September 2014).

Gillon, B. 1992. Towards a common semantics for English count and mass nouns. Linguistics and Philosophy 15. 597–640.

Greenberg, Joseph. 1972. Numeral classifiers and substantive number: Problems in the genesis type. In Keith Denning & Suzanne Kemmer (eds.), *On language: Selected writings of Joseph H. Greenberg*, 166–198. Stanford, CA: Stanford University Press.

Harley, Heidi & Rolf Noyer. 1999. Distributed morphology. *Glot International* 4 (4). 3–9.

Harley, Heidi & Elizabeth Ritter. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78 (3). 482–526.

Haspelmath, Martin. 2013. Occurrence of nominal plurality. In Matthew Dryer & Martin Haspel-math (eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology, Leipzig. http://wals.info (accessed 15 September 2014).

Hofling, Charles. 2000. *Itzaj Maya grammar*. Salt Lake City: University of Utah Press.

Hornstein, Norbert. 2009. *A theory of syntax: Minimal operations and universal grammar*. Cambridge: Cambridge University Press.

Horrocks, Geoffrey & Melita Stavrou. 1987. Bounding theory and Greek syntax: Evidence from wh-movement in NP. *Journal of Linguistics* 23, 79–108.

Ido, Shinji. 2005. *Tajik*. Muenchen: Lincom GmbH.

Jackendoff, Ray. 1977. X'-Syntax: A Study in of Phrase Structure. Cambridge, MA: MIT Press.

Karimi, Simin. 1999. A note on parasitic gaps and specificity. *Linguistic Inquiry* 30. 704–713.

Keenan, Edward 1987. A semantic definition of "indefinite NP". In Eric J. Reuland & Alice G. B. Meulen (eds.), *The representation of (in)definiteness*, 286–317. Cambridge, MA: MIT Press.

Krifka, Manfred. 1995. Common nouns: A contrastive analysis of Chinese and English. In Greg Carlson & Jeffrey Pelletier (eds.), *The generic book*. Chicago: University of Chicago Press.

Lambton, Ann K. S. 1974. *Persian grammar*. Cambridge: Cambridge University Press.

Mahootian, Shahrzad. 1997. *Persian*. London: Routledge.

Milsark, Gary. 1979. *Existential sentences in English*. New York: Garland.

Perry, John R. 2005. *A Tajik reference grammar*. Leiden: Brill.

Ritter, Elizabeth. 1991. Two functional categories in noun phrases: Evidence from modern Hebrew. In Susan Rothstein (ed.), *Syntax and semantics 25: Perspectives on phrase structure*, 37–62. New York: Academic Press.

Ritter, Elizabeth. 1992. Cross-linguistic evidence for number phrase. *Canadian Journal of Linguistics* 37 (2). 197–218.

Sanches, Mary & Linda Slobin. 1973. Numeral classifiers and plural marking: An implicational universal, 1–22. Stanford, CA: Stanford University Press.

Simpson, Andrew. 2005. Classifiers and DP structure in Southeast Asia. In Guglielmo Cinque & Richard Kayne (eds.), *The Oxford handbook of comparative syntax*, 806–838. Oxford: Oxford University Press.

Szabolcsi, Anna. 1981. The possessive construction in Hungarian: A configurational category in a non-configurational language. *Acta Linguistica Scientiarum Academiae Hungaricaae*, 31. 261–289.

Szabolcsi, Anna. 1984. The possessor that ran away from home. *The Linguistic Review* 3. 89–102.

Szabolcsi, Anna. 1987. Functional categories in the noun phrase. In Istvan Kenesei (ed.), *Approaches to Hungarian*, vol. 2: 167–190. Szeged: Jate.

Szabolcsi, Anna. 1994. The noun phrase: The syntactic structure of Hungarian. In Ferenc Kiefer & Katalin Kiss (eds.), *Syntax and semantics* 27. 179–274. New York: Academic Press.

Tang, Chih-Chen Jane. 2004. Two types of classifier languages: A typological study of classification markets in the Paiwan noun phrase. *Language and Linguistics* 5 (2). 377–407.

T'sou, Benjamin. 1976. The structure of nominal classifier systems. In Philip Jenner, Stanley Starosta & Laurence Thompson (eds.), *Austroasiatic Studies*, vol. 2: 1215–1248. Honolulu: University of Hawaii Press.

Zamparelli, Roberto. 1995. *Layers in the determiner phrase*. Rochester, NY: University of Rochester dissertation.

Zavala, Roberto. 2000. In Gunter Senft (ed.), *Systems of nominal classification*, 114–146. Cambridge: Cambridge University Press.

Jila Ghomeshi

12 The associative plural and related constructions in Persian

Abstract: This article takes constructions in Persian that consist of a proper name followed by a third person plural pronoun in Persian, *Kian ina* 'Kian and his family', and identifies them as associative plural constructions. Their properties as associative plurals are consistent with what we know of this construction in general, but the article goes further to show that they bear a great resemblance to a general extender construction in Persian in which the third person plural *ina* can follow *any* type of constituent to mean something like "etc." The associative plural also bears some resemblance, at least in meaning, to co-compounding in the language. These resemblances are formalized within the model of the hierarchical lexicon which we find in Construction Morphology.

Keywords: Persian, associative plurals, general extenders, co-compounds, Construction Morphology, hierarchical lexicon

1 Introduction

In this article I consider the associative plural construction in Modern Persian in the context of similar constructions across other languages. I first establish that the properties of this construction are quite straightforward, given what we know of associative plurals in general. I then use data from a corpus of spoken Persian to show a close connection between the associative plural and the use of

Jila Ghomeshi, University of Manitoba

general extenders both in form and meaning. I conclude from this that general extenders are a possible source of associatives, pending an appropriate diachronic study. Moreover, I show that the associative plural also bears a resemblance to co-compounding in Persian and argue that the existence of co-compounds in the language is significant in giving rise to the possibility of a grammaticalized associative plural.[1] The discussion of these issues is framed in terms of Construction Grammar.

In Modern Standard conversational Persian, a proper name (PN) can combine with the third person plural pronoun *ina* 'they'[2] to form a compound meaning "[PN] and his or her family and close friends". The resulting phrase, [PN *ina*], can be used for enquiring about people known to the speaker and addressee, for example:[3,4]

(1)   a.   *æz*   *Babak*   **ina**   *che*   *xæbær?*
           from   Babak   3PL   what   news
           'How's Babak (and family)?'

---

1 The link between associative plurals, similative plurals, general extenders, and echo word formation is also found in Mauri (to appear). In a survey of sixty languages, Mauri identifies these among a range of linguistic strategies that are used to build *ad hoc* categories in discourse. She notes that more research is needed to better understand the "synchronic and diachronic patterns of multifunctionality" (p. 18) and is carrying out a research program on precisely this – something I was unaware of when I wrote the first draft of this article. I thank Edith Moravcsik for bringing Mauri's work to my attention. I expect a further look at the creation of *ad hoc* categories in Persian is bound to yield interesting results.

2 The form *in* is also the proximate demonstrative 'this', which contrasts with the distal demonstrative *an* 'that'. Both can bear the plural marker *-ha* (*-a* after consonants in colloquial speech):

      Proximate Distal
SG   *in* 'this'   *un* 'that' (*an* in formal pronunciation)
PL   *ina* 'these'   *una* 'those' (*anha* in formal pronunciation)

Persian third person pronouns *u* 'she/he' and *ishan* 'they', which are used only for animates, have come to acquire formal/polite and/or honorific status and so the demonstratives *in* and *ina* are now used as the neutral third person forms for animates as well as inanimates. The distal demonstratives *un* and *una* can also be used pronominally, although for some they carry a less respectful connotation than *in/ina* when used in reference to human beings.

3 All naturally occurring examples are taken from CALLFRIEND Farsi (Canavan and Zipperlen 1996), a corpus of unscripted telephone conversations between native speakers of Farsi (Persian) placed inside the continental United States and Canada. See http://catalog.ldc.upenn.edu/LDC96S50 for more information about the corpus.

4 The abbreviations used in the glosses are given at the end of the article. Where data are taken from other sources, the abbreviations and glosses have been changed to conform to the system used in this article.

b. *xæbær næ-dar-im   do   se   hæfte-iye*
    news   NEG-have-1PL   two   three   week-3SG.COP
    'We have no news, it's been two or three weeks'.
    (CALLFRIEND Farsi, Canavan, and Zipperlen 1996: FA 4117:4:00)

This type of construction is quite common cross-linguistically and is known as the associative plural (APL) construction. In their survey of 237 languages Daniel and Moravcsik (2013) show that a surprising 84 percent of them – that is, 199 languages – have APL constructions. The Persian construction bears almost all the hallmark features of associative plurals, as outlined by Moravcsik (2003) and Daniel and Moravcsik (2013) (see also Daniel [2000] and Corbett [2000]).[5] First, with respect to their form, associative plurals commonly involve the same marker as the one used for additive plurals, however, there are languages in which a distinct form such as a plural pronoun is used. The example below from Mandarin shows that it, like Persian, uses a plural pronoun (cf. also the English *John 'n them*, which has a reduced conjunction between the PN and the pronoun):

(2) *zhangsan   tamen*
     Zhangsan   they
     'Zhangsan and his group'
     (Moravcsik 2003: 470)

Second, the nominal expression to which the APL marker is added is usually restricted to human referents.[6] Moravcsik (2003:472.G-1) expresses this as a scale and notes that in any given language, if a nominal on the scale can form an associative plural all nouns to the left of it can also form associative plurals:

(3) Proper Name < Definite Kin Noun < Definite Title Noun
     < Other Definite Human Noun

Persian falls on the more permissive end of the scale in that not just proper names but kin terms and title nouns can all be following by *ina*:

---

5 More recent work that explores the syntax of associative plural marking within a formal generative framework includes Nakanishi and Ritter (2009) on Japanese, Görgülü (2011) on Turkish, and Forbes (2013) on Gitskan, a First Nations language of northwestern British Columbia.
6 As we will see in section 4, *ina* can be added to nouns with inanimate reference as well, but as a general extender, not as an associative marker. One of the points of this article is that these two uses should be kept distinct as some languages may have one but not the other.

(4) a. *færhad   ina*
   Farhad   3PL
   'Farhad and his family/close friends'

  b. *xahær-et* *ina*
   sister-2SG.CLC   3PL
   'your sister and her family/close friends'

  c. *pæri   xanom   ina*
   Pari   lady   3PL
   'Pari (formal) and her family/close friends'

  d. *aqa-ye   mohændes   ina*
   sir-EZ   engineer   3PL
   'Mr. Engineer (honorific) and his family/close friends'

In addition to showing the kinds of terms that can form an associative plural in Persian, examples (4a), (4b), (4c), and (4d) also show that the APL construction in Persian is not a *morphological* process. The expression that combines with *ina* need not be a single word, but can be complex. It can be an inflected common noun that serves as a name (4b),[7] or a title noun consisting of at least two words (4c and 4d).[8] This sets the construction apart from compounding, even though it shares with compounds a binary form. It is more aptly described as a kind of syntactic juxtaposition.

 In terms of its semantics, the Persian APL construction is also in line with what is known about such constructions in general. Associative plurals refer to a set of individuals who form a conceptually coherent group. The group is ranked such that there is a prominent member who is identified by name, the "focal referent", and unnamed associates who are typically other family members (Moravcsik 2003: 471–473).

 The set of properties exhibited by the Persian APL construction that have been described above can be summarized as follows:

(5) The Associative Plural in Persian
  The expression [X *ina*], where X is a proper name, kinship term, or title, refers to the person named and his or her family and close friends.

---

7 With kinship terms, there may be a preference for *ina* to follow possessed nominals, though whether this is true requires further investiation. Edith Moravcsik (personal communication) suggests that the same may be true for plural kinship terms in Hungarian.
8 For lack of a more accurate cover term comprising proper names, kinship terms, and titles, I will continue to characterize the Persian APL construction as targeting proper names.

In this article I consider the Persian associative plural construction within the context of the grammar of spoken Persian in general. In section 2 I discuss plural marking in Persian and show that the morphological plural is not incompatible with an associative meaning. This makes it all the more interesting why the associative plural is expressed via juxtaposition instead. In section 3 I present a brief survey of coordinate compounds in Persian in order to argue that the semantics of these types of compounds are closer in meaning to the associative than the morphological plural is. In section 4 I argue that the APL construction bears a significant resemblance to a type of general extender in Persian and hypothesize that this is the source from which it has grammaticalized. In section 5 I further discuss the ways in which the APL construction is similar to and different from the general extenders and coordinate compounds and formalize the relationships within the framework of Construction Grammar. Section 6 concludes the article.

Throughout this article I use the term *construction* in the sense of Construction Grammar, that is, as a pairing of form with meaning that involves aspects of the meaning that cannot be attributed to the component parts (see Fried [2015] for an overview of this approach). Since my aim is to present work that is primarily descriptive, I don't take a position on whether constructions are acquired via a process of categorization (Goldberg 2006) or whether constructions are made available by UG (Universal Grammar [Borer 2005]). I also set aside the exceedingly interesting diachronic questions regarding which came first, the more general or the more specific construction that I link together. Rather I intend this work to highlight the advantages of considering a particular construction in relation to similar phenomena in a given language.

2 Plural marking in Persian

Plural marking on common nouns in Persian is of interest to contemporary linguists in large part because it differs from what we expect of inflectional number marking in general, but also because Persian possesses both classifiers and number marking, making it typologically somewhat rare. When discussing plural marking in Persian, most linguists focus on the suffix *-ha* as it is the default marker, even though there are other ways of forming plurals. For instance, the suffix *-an* is used with some animate nouns (*mærd* 'man', *mærd-an* 'men') and words of Arabic origin may take their plural form according to Arabic rules (e.g., *tæræf* 'side', *ætraf* 'sides'; *šæxs* 'person', *æšxas* 'people', see Lazard (1957, 1992) for more on these points.

Returning to *-ha*, it has been noted that it is a stress-affecting suffix, unlike others in the language that are clearly inflectional and do not affect stress placement (see Kahnemuyipour [2000, 2003], who argues that plural marking is derivational). Thus in the following examples, stress falls on the second syllable of *ketab* 'book' when it appears on its own (6a) or with a pronominal clitic possessor (6b), but the stress moves onto *-ha* when it is plural (6c, this example also shows that when the stem ends in a consonant the plural suffix shows up as just *-a* in informal colloquial speech):

(6)   a.   *ketáb*   b.   *ketáb-æm*   c.   *ketab-á*
         book   book + 1SG.CLC   book + PL
         'book'   'my book'   'books'

Apart from the fact that *-ha* doesn't behave morphophonologically like a typical inflectional affix in Persian, it has unusual syntactic and semantic properties as well. First, it is not required on indefinite common nouns in order to obtain a plural reading in certain contexts, as shown in (7a). When a bare common noun is construed as definite, it is also interpreted as singular as shown in (7b).[9] Example (7c) shows a bare noun in object position where it is also number-neutral. (See Ghomeshi [2008] for more on the construal of bare nouns and number neutrality in Persian.)

(7)   a.   *ketab*   *ru*   *miz*   *hæst*
         book   on   table   be.PRS.COP.3SG
         'There are/is books/a book on the table'.

       b.   *ketab*   *ru*   *miz*   *hæst*
         book   on   table   be.PRS.COP.3SG
         'The book is on the table'.

       c.   *ketab*   *xærid-æm*
         book   buy.PST+1SG
         'I bought books/some books/a book'.

---

9 As pointed out by an anonymous reviewer, the agreement facts correlate with definiteness in (7a) and (7b) such that plural agreement on the copula (*hæst-ænd* be.PRS.COP-3PL) is not possible in (7a) but is optional with a plural definite noun in (7b):

(i) *ketab-ha ru miz hæst/   hæst-ænd*
   book-PL on table be.PRS.COP.3SG/ be.PRS.COP-3PL
   'The books are on the table'.

The optionality of plural agreement in this case has to do with the animacy (or, more precisely, the lack thereof) of the subject (see Sedighi [2010] for more on this animacy effect).

Second, it is connected to definiteness but not necessarily so as examples (8b) and (8c) show (see Ghomeshi [2003]; Gebhardt [2008, 2009]; Ghaniabadi [2010] for various analyses of this property):

(8) a. *se-ta   ketab*
     three-CL   book
     'three books'

   b. *se-ta   ketab-a*
      three-CL   book-PL
      'the three books'

   c. *ketab-a-ye   jaleb-i*
      book-PL-EZ   interesting-INDEF
      '(some) interesting books'

Third, it can appear on mass nouns as well as count nouns, but without the kind of coerced reading that plural mass nouns in English receive, (see Ghaniabadi [2012] for a formal account and Sharifian and Lotfi [2003, 2007] for a conceptual-functional one):

(9) *bærf-a   ab=shod*
    snow-PL   water=become.PST.3SG
    'The snow melted' (meaning all the snow in a given context, not types or given quantities of snow)

Fourth, it can appear on constituents other than argument nominals. In (10a) we see plural marking on an adverb and in (10b) we see it on the non-verbal element within a complex predicate (see Hincha [1961] as well as Ghomeshi [2003] and Ghaniabadi [2010] where these facts are mentioned but not formally accounted for):

(10) a. *(un) bala-ha*
        that   above-PL
        'up thereabouts'[10]

---

10 An anonymous reviewer has point out that the English translation illustrates what might be a similar use of English -*s*. The use of nominal morphology on non-nominal elements is an area that is left for future research.

b. *dærd-ha=keshid-im*
pain-PL=pull.PST-1PL
'We have suffered a lot'.

The set of properties that the Persian plural marker exhibits has led to a number of proposals on how to treat it within formal theory. It has been analyzed as an adjunct rather than a head (e.g., Ghaniabadi [2010] drawing on Wiltschko [2008]) and as an affix that bears a definiteness/specificity feature (e.g., Gebhardt [2009]; Ghaniabadi [2010, 2012]). For Cowper and Hall (2012) it is a morpheme that instantiates a meaning of "augmented assemblage", which they abbreviate as "AGGLOM". This feature differs from the one they abbreviate as ">1", which contributes the meaning more typically associated with plural, namely "more than one referent" (cf. Hincha [1961], who also proposes that the semantics of the Persian plural marker is closer to augmentation than "greater than one").

Given that the distribution and meaning of *-ha* is wider than a grammaticalized inflectional plural like *-s* in English, it is within the realm of possibility that it could be the marker for associative plurals as well. This would make Persian more like Japanese in which the same marker *-tati* is used for additive and associative plurals (Nakanishi and Ritter [2009]). Moreover, *-ha* can appear on proper names. Ghomeshi and Massam (2009) note that Persian proper names can be pluralized, as in English, to refer to groups of people with the same name:

(11) a. *qomeshi-a   æksæræn   æz   shomal-e   iran-æn*
          Ghomeshi-PL mostly   from north-EZ   Iran-COP.3PL
          'Ghomeshis are usually from the north of Iran'.

      b. *æli-a   in   vær   vais-æn   bæqiye   un   vær*
          Ali-PL   this side stand.PRS-3PL the.rest   that side
          'The Alis should stand on this side and the rest (of you) on the other side'.
          (Ghomeshi and Massam 2009: 87.40)

Significantly, plural proper names formed with *-ha* cannot refer to a family the way they can in English:

(12) a. Have you invited the Ghomeshis over recently?

      b. The Smiths are arriving at noon.

This distinction between a group of members sharing the same name and a group of members linked to each other as parts of a whole is what Moravcsik (2003: 476–477) calls type plurals vs. group plurals. Type plurals are taxonomic and membership in the set denoted by the plural term is based on similarity, i.e., set members are tokens of the same type (e.g., bread and cake, beef and pork). Group plurals are partonomic and membership in the set is based on a sense of cohesion, i.e., set members are parts of a whole (e.g., bread and butter, beef and potatoes). The examples above show that plural proper names in Persian can only be type plurals, not group plurals.[11]

Given that *-ha* on proper names yields a type, not group, reading and that the group reading is attained by adding the associative plural *ina*, the question arises as to whether *-ha* can ever have an associative or group reading. The use of plural marking on second person pronouns in Persian shows that it can.

The use of personal pronouns in Persian is complicated by the social and cultural imperatives to be polite and to confer honorific status on one's social superiors as well as on those of the same status who are not close or familiar. Thus, as in many European languages, the second person plural pronoun in Persian is often used as a polite *singular* while the second person singular pronoun is used when the addressee is familiar and of the same social status. As we see in (13a) and (13b) below, the shift to the polite singular *shoma* is often accompanied by the use of a more formal verb. In (13c) we see the use of *-ha* on *shoma* to denote a neutral second person plural:

(13) a. (*to*)   *bayæd*   *bi-yay*   *inja*
          you.SG   must   SBJ-come.PRS.2SG   here
          'You (SG) should come here'.

      b. *shoma*   *bayæd*   *tashrif=bi-yar-id*   *inja*
          you.PL   must   welcome=SBJ-bring.PRS.2PL   here
          'You_{honorific} (SG) should come_{honorific} here'.
          (plural reading for subject also possible)

      c. *shoma-ha*   *bayæd*   *bi-yayd*   *inja*
          you.PL-PL   must   SBJ-come.PRS.2PL   here
          'You (PL) should come here'.

---

11 Daniel and Moravcsik (2013) distinguish additive plurals, which are referentially homogeneous in that every member of the set is the same type, from associative plurals, which they call "referentially heterogeneous". They add that the sets referred to by associative plurals exhibit cohesion – they are not random collections. This appears to be the same distinction Moravcsik (2003) is making with type vs. group plurals. I will continue to use her terms as the term *additive* is used for a type of co-compound in section 3.

Moravcsik (2003: 492–493) argues that first and second person plural pronouns tend to be associative plurals, although they can also be non-associatively interpreted. In the case of (13c) above, if there is only a single addressee, the most likely reading is that the speaker is including known associates of the addressee, i.e., the addressee plus her or his friends or family. Clearly then, there is no obvious reason why the plural marker could not be used to mark associative plurals on proper names as well, yet the language employs what Daniel and Moravcsik (2013) call "periphrastic" means to signal the associative plural.[12]

In the next section we turn to compounding constructions, particularly of the coordinating type, to show that set extension is often expressed via compounding rather than through affixation. The availability of this strategy may explain how the APL construction in Persian has arisen.

3 Compounding, echo reduplication, and hedging your sets

We saw at the beginning of this article that an associative plural refers to a set made up of the individual named by the proper name and the associates of that individual, typically his or her family members. Moravcsik (2003) refers to this kind of set as a group plural in that the members form a cohesive whole, e.g., a family. She further categorizes it as a ranked set in that it contains one focal member, and a partially enumerated one in that the associates are not named. In this section we consider compound constructions that vary along one or more of these parameters of meaning. Compounds are relevant because of their superficial resemblance to the APL construction and also because compounding is a productive and creative area in Persian and many related and contact languages.

The types of compound we will consider here are called co-compounds or coordinating compounds (Wälchli [2005], also known by the Sanskrit term *dvandva* compounds or as copulative compounds). These compounds are so named because their meaning involves the coordination of the meaning of their parts. They are found throughout the languages of central Eurasia and express a variety of semantic relations. For example, in Lezgian (spoken in the eastern

---

12 Ackema and Neeleman (2014) claim that the associative effect of plural marking with first and second person pronouns is linked to the person system, not the number system. If this is correct, then *-ha* may only be an additive plural in Persian. I leave this issue for further research as it does not directly affect the exploration of *ina* as an associative marker.

Caucasus), the two members of the compound may appear as a pair (14a), may represent a larger class by virtue of being identifiable members of that class (14b), may be synonyms of each other (14c), or may involve one nonce-member (14d):

Lezgian (Haspelmath 1993:108)

(14) a. *buba-dide*   *buba* 'father'   *dide* 'mother'   'parents'

     b. *xeb-mal*   *xeb* 'sheep'   *mal* 'cattle'   'domestic animals'

     c. *gaf-č'al*   *gaf* 'word'   *č'al* 'word, language' 'talking'

     d. *ajal-kujal*   *ajal* 'child'   **kujal*   'child'

Wälchli (2005) provides further detail about each of the four types of co-compounds above as well as identifying many more. He calls the type in (14a) an additive co-compound and notes that these compounds refer to items that naturally occur in pairs or that together exhaustively represent a set:

Georgian Additive Co-compounds (Wälchli 2005: 137)

(15) a. *da-dzma*   'sister-brother'

     b. *xel-p'exi*   'hand-foot'

Wälchli calls the second type of compound, exemplified in (14b), a collective co-compound. In this case the compound denotes a collective of which the two listed elements are prototypical members.

Chuvash Collective Co-compounds (Wälchli 2005: 141)

(16) a. *sĕt-śu*   lit. 'milk-butter'   'dairy products'

     b. *erex-săra*   lit. 'vodka/wine-beer'   'alcoholic beverages'

     c. *xyr-čărăš*   lit. 'pine-spruce'   'conifers'

Collective co-compounds can also be viewed as hyperonymic in that they denote superordinate-level concepts. Arcodia, Grandi, and Wälchli (2010) contrast hyperonymic with hyponymic coordinating compounds such as "singer-songwriter" in English, which denote subordinate-level concepts (in this case, the intersection of the set of singers with the set of songwriters).[13]

---

13 See also Bauer (2010) for discussion of co-compounds in Germanic languages such as English, German, and Dutch. Such languages are notable for the striking lack of the kinds of co-compounds being discussed here.

The examples in (14c) and (14d) show that it is possible to have co-compounds in which one of the elements is not meaningful, either by virtue of being synonymous with the other element (see also Singh [1982] on synonym compounds in Hindi), or by not being a recognizable word at all, which Wälchli (2005) calls imitative (see also Ourn and Haiman [2000] on "servant words" in Khmer).

Examples of each of these types of co-compounds: additive, hyperonymic, synonym, and imitative, can be found in Persian:

Persian (cf. Ghaniabadi et al. 2006; Stilo 2004; Shaki 1967)

(17)	a.	*kot-šælvar*	*kot* 'coat', *šælvar* 'pants'	'suit'	**additive**
	b.	*kard-o čængal*	*kard* 'knife, *čængal* 'fork'	'cutlery'	**hyperonymic**
	c.	*dad-o færyad*	*dad* 'shout', *færyad* 'cry'	'brawl'	**synonym**
	d.	*pul-o pæle*	*pul* 'money', **pæle*	'wealth'	**imitative**

A significant point to be made about the compounds in (17) above is that most retain the coordinator -*o* 'and', which seems to be true of many types of compounds in Persian (Stilo 2004: 285–286).[14] Thus with the frequently occurring additive compound meaning 'parents', we can get the nouns *madær* 'mother' and *pedær* 'father' in either order, both with and without the coordinator: *pedær(-o) madær* and *madær(-o) pedær*. Consequently, the term *compound* here does not refer to a particular *form*. Rather, it has to do with whether the expression is lexicalized in the sense of being listed as a chunk in the lexicon with a somewhat idiomatic meaning. Moreover, just as there are expressions containing a coordinator that are indisputably lexicalized, so there are spontaneous coordinated expressions in which the coordinator is not present. Stilo (2004: 308–309) points out that when a coordinator between two noun phrases is deleted, the resulting sequence takes on the stress pattern of a compound (stress indicated by small caps):

---

14 An exception would be numeral compounds such as *do-se* 'two or three' (lit. 'two-three'), *čar-panj* 'four or five' (lit. 'four-five') and which are treated phonologically as compounds but never take an overt coordinator as noted by Stilo (2004: 286). They can also express ranges (e.g., *hæft-hæsht-dæh ta* 'seven to ten' (lit. 'seven-eight-ten' CL, or *dæh-punzdæh ta* 'ten to fifteen' [lit. 'ten-fifteen' CL], which sound better when followed by the default classifier *ta*). These are very common in everyday speech.

(18) a. *emruz-færDA*   *mi-r-e.*
          today-tomorrow   CONT-go.PRS.3SG
          'S/he's going today or tomorrow'.
          (Stilo 2004: 308.127)

     b. *pepsi-koKA*   *be-færma-id.*
          Pepsi-Coke   IMP-command.PRS.2PL
          'Have a Pepsi or a Coke'.
          (Stilo 2004: 308.130)

In the above cases, a disjunctive reading is more likely, but Stilo notes that it is not always clear whether a conjunctive or disjunctive reading is intended and in some cases the distinction is irrelevant:

(19) *xahæ̀r-bæraDÆR*   *dár-in?*
      sister-brother   have.PRS-2PL
      'Do you have brothers or sisters?'/'Do you have a brother or a sister?'
      (Stilo 2004: 308.131)

Note that in the likely event that *xahær-bæradær* is an additive compound like *madær(-o) pedær* 'mother-father > parents', the example in (19) could also be translated as 'Do you have any siblings?'

    Returning to co-compounds, Wälchli (2005: 167–168) notes that many co-compounding languages also have compounds made up of echo-words. This phenomenon, also known as echo reduplication, involves the repetition of a base X with part of the base replaced by a fixed segment (see, for example, Lidz [2001]; Keane [2001]; Inkelas and Zoll [2005]). The resulting compound has the meaning "X and the like", "X and related stuff", or "X, etc." Echo reduplication is found in Iranian (e.g., Persian), Indo-Iranian (e.g., Bengali, Hindi), Turkic (e.g., Turkish), and Dravidian (e.g., Kannada) languages and is thus considered an areal feature of South Asia:

(20) a. Bengali   *bari* 'house'   *bari ṭari* 'house, etc.'

     b. Hindi   *aam* 'mango'   *aam vaam* 'mangoes and such fruit'

     c. Kannada   *kannu* 'eye'   *kannu ginnu* 'eyes and so forth'

     d. Tamil   *maaṭu* 'cow'   *maaṭu kiiṭu* 'cattle in general'
    (Keane 2005: 240.1–4)

In Persian the fixed segment is /m-/ and occasionally /p-/ (see Ghaniabadi [2008]; Ghaniabadi et al. [2006]):

Persian

(21) a. *ketab* 'book'   *ketab-metab* 'books and stuff/and the like/, etc.'

   b. *pul* 'money'   *pul-mul* 'money and stuff/and the like/ etc.'

   c. *mive* 'fruit'   *mive-pive* 'fruit and stuff/and the like/, etc.'

There are several ways in which compounds formed by echo reduplication resemble the other types we have briefly reviewed. They share a similar form with imitative compounds in that they contain one "nonce" member, and they share a similar meaning to collective, or hyperonymic, compounds in that they extend the denotation of the base. Collective compounds evoke a superordinate category by naming two typical members (e.g., *kard-o čængal* 'knife and fork →
cutlery'), echo words evoke a larger set of items *like* the named member (e.g., *kard-mard* 'knives and stuff'). We can define both as set-extending constructions (set here meaning the denotation set) as follows:

(22) Coordinate compounding in Persian
   Given X and Y, the compound [X Y] denotes the hypernym, i.e., a super-
   ordinate set including X and Y as typical members.

(23) Echo reduplication in Persian
   Given X, the compound [X *m/p*-X] denotes a set including X and the like.

Note that the semantics of echo reduplication, which is usually characterized as "X and the like" or "X and related stuff" (see references given above), is some-
what vague. A more precise formulation is challenging, however, as the set evoked can vary from context to context. For example, in one situation, what is associated with X might be based on physical similarity (books and things that look like books) while in another, it might be the kinds of things typically scattered on a desk (books, pens, notepaper, etc.)

   Given that echo reduplication can be characterized as a kind of set exten-
sion, we can now link it back to associative constructions. Indeed there are con-
structions that could easily fit one or the other definition. For instance, Wälchli (2005) cites examples from two Uralic languages spoken in Russia, Mordvin and Udmurt, in which echo-compounds involve pronominal "echoes" rather than phonologically modified copies of the base:

Mordvin (Wälchli 2005: 169)

(24)   a.   *jam.t-mez.t'*   'soup.PL-what.PL > soup and the like'[15]

   b.   *koŕon.nek-mez.ńek*   'root.NEK-what:NEK= with roots and everything, with all its roots'

The only reason he does not consider these constructions to be associative plurals is that they bear the morphological markers of co-compounds (plural marking in [24a] and *-ńek*, a marker is specific to generalizing co-compounds in Mordvin in [24b]).[16]

Having drawn links between co-compounding, echo reduplication, and associative plurals, there are a few distinctions, particularly between the latter two, that ought to be made. First, echo-compounds are, as Wälchli (2005) notes, of a low informal register. They are unlikely to occur in written texts and have a careless or almost pejorative air. This is not true of the APL construction in Persian. It is similar to echo reduplication in that it is more characteristic of spoken than written language and it is informal, but it is neither "low" nor pejorative or dismissive. Second, echo-compounds are an instance of a type plural, where membership in the set they denote is based on similarity, and the reading is taxonomic. Associative plurals, as we have seen, are group plurals where membership in the set is based on cohesion, and the resulting reading is partonomic, according to Moravcsik (2003). Third, there is a strong tendency for echo-compounds to be based on nouns with inanimate reference while associative plurals are based on proper names, titles, and kinship terms, i.e., on terms referring to human beings.[17]

---

15 Note the resemblance here to the English "and whatnot", which can also be seen as an associative construction. This construction is vividly defined on Urban dictionary as "A more hip hop way of saying 'etc.', or a verbal way of expressing '. . .' It is said by those that have so much poppin' that they don't have the time or energy to explain what the 'what not' is." (http://www.urbandictionary.com/define.php?term=whatnot, accessed 6 November 2015). The definition itself is astonishingly similar to one given one hundred years earlier by Poutsma (1916: 914, as cited in Cheshire 2007: 165) for "and (all that)" as a form that "sometimes stands for a vague *etc.*, which the speaker is not prepared to specify in the hurry of the discourse".

16 Daniel and Moravcsik (2013) also note the similarity between echo reduplication (which they call the similative plural with reference to Telugu) and the associative plural construction in that both pick out sets that do not require their members to be of the same type. Daniel and Moravcsik refer to this as referential heterogeneity.

17 Again, this claim excludes the use of *ina* as a general extender, which will be discussed in section 4. While (25b) shows that *ina* cannot be used as an associative plural with an inanimate noun such as *mive*, it can be used in the sense of "etc." For this reason I have not marked the example *mive ina* with an asterisk, indicating outright ungrammaticality but with two question marks. While I am arguing that associative plurals and general extender constructions are distinct, in some cases the meanings are similar enough to bleed into one another.

Persian

(25) a. *mive* 'fruit' *mive-pive* 'fruit and stuff/and the like/, etc.'
        kian 'Kian' ?? *kian-mian*

   b. *mive* 'fruit' ??*mive-ina*
        kian 'Kian' *kian-ina* 'Kian and his family/friends'

We have seen arguments for associative plurals being closer to compounds than to plural marked nouns. They fall in easily with the other types of compounds in Persian, many of which are "set-extending". Rather than viewing associative plurals as simply one kind of compound construction, however, I will argue in the next section that they are connected to a more pervasive means of extension or generalization via coordination.

4 General extenders

The associative plural in Persian consists of a proper name and *ina*, which on its own is the neutral third person plural pronoun 'they/them' as well as the plural proximate demonstrative 'these' (see note 1). As expected, it is easy to find examples of associative plural and pronominal uses of *ina* in chunks of casual spoken discourse. What is unexpected, however, is the large number of occurrences of *ina* do *not* fall into these two categories.

Let us consider the first eight minutes of one telephone conversation from the CALLFRIEND corpus (Canavan and Zipperlen 1996: FA_6345). The call is between two men and, apart from general enquiries about each other and family, the conversation revolves around a government shutdown in the state where speaker A lives, and the news speaker A has just received at the doctor's office about his high cholesterol count. There are eighteen instances of *ina* in this chunk of conversation, seven produced by speaker A and eleven by speaker B. Among the eighteen occurrences of *ina* there are no cases where it functions as a pronoun meaning 'they' and exactly one instance of an associative plural:

(26) A: *xune-ye færhad **ina** hæst-i, chikar mi-kon-æn?*
        house-EZ Farhad 3PL be.COP-2SG what CONT-do.PRS-3PL
        'So you're at Farhad (and family)'s house, what are they up to?'
        (CALLFRIEND Farsi, Canavan and Zipperlen 1996: FA 6345:4:10)

The rest of the occurrences of *ina* are what I will call, following Cheshire (2007), instances of its use as a general extender, along the lines of '… and stuff' in

English.[18] I will first present a range of examples and then discuss this use further.

In the following two examples *ina* is linked to a noun via the coordinator *-o* and extends the denotation of the noun to related things:

(27) A: *væ muze-ha-o   ina hæme tæ'til-e*[19]
      and museum-PL-CONJ 3PL all   closed-be.COP.3SG
      'And places like museums are all closed' (here the relevant places are
      probably state-run as the topic of conversation is a 'shutdown')
      (CALLFRIEND Farsi, Canavan and Zipperlen 1996: FA 6345:4:10)

(28) B: *fæqæt mi-tun-i   mesinke   ab-o   ina bo-xor-i*
      only CONT-be.able.PRS-2SG seems-like water-CONJ 3PL SBJ-eat.PRS-2SG
      'You can only drink water and stuff'. (with regard to the morning on
      a day you have to go for tests for which you are required to fast,
      presumably the other "stuff" in this case is clear liquids)
      (CALLFRIEND Farsi, Canavan and Zipperlen 1996: FA 6345:3:15)

The following example similarly shows the set-extending function of *ina*. We also see that *ina* can follow a fully inflected noun, in this case a noun bearing both plural marking and a possessive clitic, and one that has been scrambled out of a subordinate clause, in this case the complement of *bebinin* (the imperative form of 'see'):

(29) A: *xærj-a-tun   ina be-bin-in   che-qædr-e*
      expense-PL-2PL.CLC 3PL IMP-see.PRS-2PL how-much-be.PRS.COP.3SG
      'See how much your expenses and stuff are'.
      (CALLFRIEND Farsi, Canavan and Zipperlen 1996: FA 4099:9:20)

In the examples so far, *ina* has extended the denotation of the noun that precedes it. However, there are examples like the one below where *ina* is linked to a preceding noun via the coordinator *-o*, but it is not clear what the '… and stuff' has scope over: is it generalizing over diets and other things like that, or over activities like putting someone on a diet and other things like *that*:

---

18 Cheshire in turn credits Overstreet (1999) for the term *general extender*, but notes that plenty of other terms have been used (see Cheshire [2007: 156] and references cited therein).
19 While not related to the topic of this article, this example also illustrates another feature of Persian whereby inanimate plural subjects do not trigger plural but rather singular agreement on the verb. See Sedighi (2010), for discussion.

(30) A: *un-æm*   *dige*   *be-zar-æt-æm*   *ru*   *dayet-o*   ***ina***
      that-as.well   well   SBJ-put.PRS-3SG.1SG.CLC   on   diet-CONJ   3PL
      '... and, you know, he'll put me on a diet and stuff'
      (with regard to an appointment he has to make to see a nutritionist)
      (CALLFRIEND Farsi, Canavan and Zipperlen 1996: FA 6345:2:10)

Apart from nouns, *ina* can also follow a verb or a verb phrase with or without a coordinator. As with nouns, the generalization may be over other actions (expressed by the verb) or other propositions (expressed by the clause). We consider three examples in turn. In the following example speaker A recounts how he first got the news of his high cholesterol count. Here the complex predicate *test=kærdæn* 'to test' (lit. 'test=do') is following by *ina* meaning that they 'ran tests and stuff':

(31) A: *injuri*   *ke*   *chiz*   *šod,*
      this.way   that   thing   happen.PST.3SG

      hæfte-ye   *piš*   *ræft-æm-o*   *test=kærd-æn*   ***ina,***
      week-EZ   last   go.PST-1SG-CONJ   test=do.PST-3PL   3PL

      bæ'd   *goft*   *divist-o-pænjah-st*
      then   say.PST.3SG   two.hundred-CONJ-fifty-be.COP.3SG
      'This way that it happened (is), last week I went and they did tests and stuff (lit. 'they tested and stuff'), then they said it's 250'.
      (with regard to his high cholesterol news)
      (CALLFRIEND Farsi, Canavan and Zipperlen 1996: FA 6345:1:46)

In the following example *ina* is linked to the verbal complex *avordæn=pa'in* 'to bring down' via the coordinator *-o* 'and'. The verb is transitive though the object is unexpressed and is construed as 'cholesterol'. The generalization is not necessarily over actions performed to cholesterol but over the whole verb phrase 'bring cholesterol down':

(32) A: *mi-g-e*   *bayæd*   *bi-yar-i=pain-o*   ***ina,***
      CONT.say.PRS-3SG   must   IMP-bring.PRS-2SG=down-CONJ   3PL

      ta   *be-bin-i*   *chi*   *piš=mi-ya-d*
      until   SBJ-see.PRS-2SG   what   near=CONT-come.PRS-3SG
      'He says you should bring it down and stuff so that you see what happens'. (talking about what his doctor recommends regarding his cholesterol)
      (CALLFRIEND Farsi, Canavan and Zipperlen 1996: FA 6345:6:10)

The third example below is similar in that the scope of *ina* is over the verb phrase "walk on a treadmill":

(33) B: *to   estres-test=kærd-i?*
          2PL   stress-test=do.PST-2SG
          'Did you do a stress test?'

    A: *are*
       yeah
       'yeah'

    B: *ke   ru   tered-mil   bayæd   ra=be-r-i*   **ina**
       that on treadmill must walk=SBJ-go.PRS-2SG 3PL
       'Where you have to walk on a treadmill and stuff …'
       (CALLFRIEND Farsi, Canavan and Zipperlen 1996: FA 6345:7:30)

Ina can also form compounds with numerals that refer to times to mean 'around' the time specified (cf. footnote 14 on numeral compounds). For example, in one conversation between an older woman (speaker A) and a male relative (speaker B), she asks if he's coming over the next day and he replies that he is. She then asks several times what time he'll be coming. He says it'll be in the afternoon but he can't be sure exactly when, it depends on what else is going on. Then he continues and says:

(34) B: … ***do   ina   mi-res-æm***   …
         … two 3PL CONT+arrive+1SG
         '… I'll arrive around two …'
         (CALLFRIEND Farsi, Canavan and Zipperlen 1996: FA 4099:0:33)

The use of *ina* is so pervasive that it can occur more than once during a single turn in conversation. Consider the following example in which speaker B is trying to explain good vs. bad cholesterol to speaker A. The first instance of *ina* follows *mæqz* 'brain' and means something like 'brain and other organs', while the second instance simply generalizes over the effects that bad cholesterol can have. In this example we also see that the coordinator is truly optional as it occurs in one case and not in the other:

(35) B: *væli kolestrol-e   bæd-et   ke   mæsælæn ru*
    but cholesterol-ez bad-3sg.clc that as.if   on

    artri-a   hæst,   ru artri-a-ye   ræg-a-i   ke
    artery-PL be.PRS.COP.3SG on artery-PL-EZ vein-PL-REL that

    *xun   mi-bær-e   tu mæqz-o   **ina** æsær=mi-zar-e   **ina***
    blood CONT-take.PRS-3SG in brain-CONJ 3PL effect=CONT-put.PRS.3SG 3PL

    'But your bad cholesterol that is in the arteries . . . in the arteries of the
    veins that carry the blood to the brain and stuff, (it) has an effect
    and stuff'.
    (CALLFRIEND Farsi, Canavan and Zipperlen 1996: FA 6345:7:10)

As mentioned at the beginning of this section, this use of *ina*, which appears com-
monly in spoken discourse, has been identified as a general extender. General
extenders are multifunctional constructions that may be put to a number of dif-
ferent uses at the same time. They have been described as implicating a category
of which the named member is an exemplar. They often evoke a category that
does not have a lexicalized name but rather has been created spontaneously.
We can note that in this use they bear more than a passing resemblance to
echo reduplication, discussed above. They also have pragmatic functions such
as marking solidarity or rapport, hedging, and/or signaling politeness (see
Cheshire [2007] and Parvaresh et al. [2012: 262–263] for a review of the literature
on these points).

    Cheshire (2007) states that general extenders like "and stuff" in English are
thought to have grammaticalized from longer expressions such as "and stuff
like that", citing Aijmer (2002) and Brinton (1996) in this regard. In her analysis
of general extenders used by adolescent speakers of British English, she con-
siders forms such as "and things like that", "and all that lot", "and all the
things like that", "and all this type of stuff" as well as disjunctive general
extenders such as "or something like that", "or whatever", "or things like that".
It is clear from her work and the work she references (see, for example, Erman
[1995]; Overstreet and Yule [1997]) that the further along the grammaticalization
path a general extender is, the shorter it is. She argues that "and that", "and
everything", and "or something" lead the way in British English in contrast
with "and stuff", which is arguably the most grammaticalized general extender
in American English.

    In a similar study, Parvaresh et al. (2012) survey the use of general extenders
by adult speakers both in their native Persian and in their non-native English
discourse. Of the general extenders they count in the native Persian corpus starting

with *væ* 'and' the most frequent by far is *væ ina* 'and these'.[20] Table 1 gives the top five of their twenty-one types of general extenders in terms of frequency. Together they represent 64.85 percent of the total number of occurrences of general extenders.

Table 1: Average frequency of general extenders in Persian (taken from Parvaresh et al. (2012.264 Table 1)

Form	Frequency	Percent
væ ina 'and stuff' (lit. 'and these')	91	37.60
væ æz in *hærf ha* 'and of such talks' (lit. 'and of these talks')	21	8.67
væ in *chiz ha* 'and such things' (lit. 'and of these things')	17	7.02
væ nemidunæm æz in *hærf ha* 'and I don't know of such talks'	16	6.61
væ hæme chiz 'and everything'	12	4.95

It is clear, then, that Persian exhibits the same phenomenon as in English: there are numerous expressions that can serve as general extenders, out of which there has emerged one concise, frequently used form that has arguably become grammaticalized. In Persian this expression is *(-o) ina* 'and these', which can be added to noun phrases, numerals, verbs, verb phrases, or clauses in order to generalize or extend the meaning of the category referred to (an object, a time, an action, a proposition).

(36) The general extender construction in Persian
      The expression [X(-*o*) *ina*], where X is any constituent in an utterance, extends the denotation of that constituent to related things.

In the next section I will argue that, despite appearances, the associative plural in Persian does not fall under the general extender use of *ina* but constitutes a construction that is more specific in ways that ties it in with the co-compounds discussed in section 3.

---

20 Parvaresh et al. (2012) also consider general extenders starting with *ya* 'or' and call them disjunctive as opposed to adjunctive general extenders that begin with *væ*. It should be noted here that their *væ* corresponds to *-o* in this article. I did not find many instances of *væ* 'and' in the spoken corpus I used and indeed, according to Stilo (2004: 288), *væ* is more common in the formal written language and is rarely used in casual speech. I can only assume that Parvaresh et al. (2012) chose to represent their data in a more formal way for clarity so as to abstract away from the messiness that is the transcription of actual speech.

5 The place of the associative plural construction within the grammar of conversational Persian

In this section I will first show that the associative plural construction has properties that distinguish it from the general extender construction in Persian. I will then go on to outline how the relationships between general extenders, the associative plural, and coordinate compounding can be represented within a Construction Grammar framework. While there is much work on Construction Grammar (see Fillmore [1988], Goldberg [1995], [2006], and Croft [2001], to name but a few references), I will be drawing primarily on Booij's (2010) *Construction Morphology*.

5.1 The distinct properties of the associative plural construction

We saw in section 4 that the most frequent kind of general extender in Persian is *-o ina*, which literally means 'and these' but functions more like "and stuff" in English. This extender can link to a variety of constituents ranging from nouns and noun phrases to verbs, verb phrases, and clauses. In contrast, the associative plural use of *ina* is restricted to proper names, kinship terms, and titles. Its meaning is also restricted in that it can only denote the family and close friends of the person named, while a general extender may evoke any type of category, depending on context and the meaning of the expressions involved. A third difference between the two constructions has to do with the number of expressions that can be combined. The associative plural involves adding *ina* to exactly one nominal, though as mentioned in the introduction to the article, this nominal can be made up of more than one word. A general extender, in contrast, may follow two or more constituents that make up a list:[21]

Persian

(37) a. *qælæm(-o)*   *medad(-o)*   *ina*
          pen(-CONJ)   pencil(-CONJ)   3PL
          'pens and pencils and stuff'

      b. **sima*   *færhad*   *ina*
          Sima   Farhad   3PL
          'Sima and Farhad and family'

---

21 According to Parvaresh et al. (2012), it has been proposed that general extenders are used to complete three-part lists and they cite Jefferson (1990) in this regard.

Fourth, while the general extender use of *ina* may involve the coordinator *-o*, the associative plural cannot. Indeed the presence of the coordinator following a proper name signals that *ina* should be interpreted as a general extender and *not* an associative. In the following example we see such a use where the speaker is listing her children or others who live with her. Importantly, the interpretation is additive rather than associative:

(38) A: *ax susæn-jun   mæn ælan umæd-æm   dær dæftær-e*
     oh Susan-dear 1SG   now come.PST-1SG in   office-EZ

     *mæxsus-æm,   væqti mi-xa-m,   bita-o   mæmmæd-o   **ina***
     special-1SG.CLC when CONT-want.PRS-1SG Bita-CONJ Mohammad-CONJ 3PL

     chiz næ-baš-æn   mi-a-m   tu dæftær-e mæxsus-æm,
     thing NEG-be.SBJ.COP-3PL CONT-come.PRS-1SG in office-EZ special-1SG.CLC

     ke   dær tualet-e   <LAUGHS>
     that in   toilet-be.COP.3SG

     'Susan dear I've come now to my special office, when I want, Bita and Mohammad and others not to be a bother [lit. 'thing'] I come to my special office, which is in the bathroom'.
     (CALLFRIEND Farsi, Canavan and Zipperlen 1996: FA 6723:1:15)

The ways in which the associative plural and the general extender use of *ina* differ from each other is summarized in the table below:

Table 2: Associative vs. general extender *ina* in Persian

	Associative *ina*	General extender *ina*
constituent *ina* attaches to	definite human nominal	any clause or nominal or verbal phrase
resulting meaning	ranked plural	related things or activities
number of conjuncts	exactly two	any number
coordinator *-o*	not permitted	optional

Table 2 shows that there are aspects of the APL construction that are not predictable from the general extender construction. Nor is the meaning of the construction determined compositionally from the meaning of its parts. For these reasons, I consider the associative plural to be a constructional idiom, an expression that contains both lexically fixed and variable positions (see Booij [2010:13] and references cited therein), and to be lexically listed as such. In the next section I turn to consider the structure and organization of the lexicon and the place of the associative plural within it.

5.2 The associative plural in the hierarchical lexicon

Construction Grammar takes the lexicon to be a hierarchically organized network of generalizations over words as well as multi-word expressions. These generalizations can be formalized as schemas or rules (see Booij [2010: 4 and ch. 2] for arguments why it is preferable to view such generalizations as schemas rather than rules). An example of a schema for compounding is given below:

(39) $[[a]_{Xk} \, [b]_{Ni}]_{Nj} \longleftrightarrow [SEM_i \text{ with relation R to } SEM_k]_j$
       (Booij 2010: 17.17)

The schema above is for right-headed nominal compounds in Germanic languages. Examples from English include words like "book shelf", "pull tab", "hard disk", and "afterthought" (Booij 2010: 17). In each case, the right-hand member of the compound, a noun, is the head and it is preceded by a noun, verb, adjective, and preposition, respectively. Booij explains the formalism in the schema in (39) as follows: the lower case letters *a* and *b* are for arbitrary sound sequences, the variable X stands for the major lexical categories (N, A, V, P), and the lower-case variables *i*, *j*, *k* stand for the indices that link the phonological, syntactic, and semantic properties of words (cf. Jackendoff's [2002] tripartite parallel architecture). The relation R remains unspecified in the semantic representation. It links the semantics (abbreviated SEM) of the head noun to the non-head in some way that is determined by the meaning of the individual words as well as the context.

Within Construction Morphology, as presented in Booij (2010), schemas for word formation processes (e.g., inflection and derivation) are not only lexically listed but are linked to one another, forming hierarchies.[22] A given schema may be dominated by a more abstract general schema (one that generalizes over a number of related constructions), from which it inherits certain pieces of information, or it may dominate a sub-schema that specifies aspects of meaning or form that are specific to a certain subclass of words that participate in the construction. To give a concrete example of the latter, the schema for compounding given in (39) above has a sub-schema in Dutch where by the word *wereld* 'world' can appear as the first constituent and function as an intensifier rather than

---

22 While I am drawing primarily on Booij (2010), the notion of an inheritance hierarchy, i.e., a taxonomic network whereby generalizations captured by higher-level constructions are inherited by subordinate ones, runs through all varieties of Construction Grammar. See Boas (2013: 244–246) and references cited therein for one discussion of such hierarchies and Fried (2015: 13–14) for hierarchical inheritance vs. partial inheritance.

contribute its literal meaning (*wereld-vrouw* 'fantastic woman', *wereld-kans* 'great chance', Booij [2010: 57]). This sub-schema can be expressed as a constructional idiom as follows:

(40) $[[wereld]_N [x]_{Ni}]_{Nj} \leftrightarrow [excellent\ SEM_i]_j$
      (Booij 2010: 57.13)

Turning to the type of compounding discussed in section 3, we can note that it is of a different type than the Germanic type represented in (39) above, expressing coordination rather than modification. Thus in Persian a compound like *kot-šælvar* (literally 'coat-pant') means 'suit', while in English a compound "coat pants" would likely be interpreted as "pants that go with a coat" or something along those lines. (Note that Persian has modificational type compounds as well.) The data presented in section 3 lend themselves to a hierarchical analysis whereby there is a general schema for coordinate compounds that dominates various types of sub-schemas (additive compounds, collective compounds, synonym compounds, etc.). This can be represented as in Figure 1 below in which I have shown two possible sub-schemas below the general schema for coordinate compounds, one for hyperonymic co-compounds and the other for echo reduplication:[23]

Schema for coordinate compounding in Persian
$[[a]_{Ni} [b]_{Nk}]_{Nj} \leftrightarrow$ [a category that includes SEM$_i$ and SEM$_k$ and possibly other members]$_j$

Sub-schema for hyperonymic compounding
$[[a]_{Ni} [b]_{Nk}]_{Nj} \leftrightarrow$ [a category that includes SEM$_i$ and SEM$_k$ as typical members]$_j$

Sub-schema for echo reduplication
$[[a]_{Ni} [m/p+a]]_{Nj} \leftrightarrow$ [a category that consists of SEM$_i$ and the like]$_j$
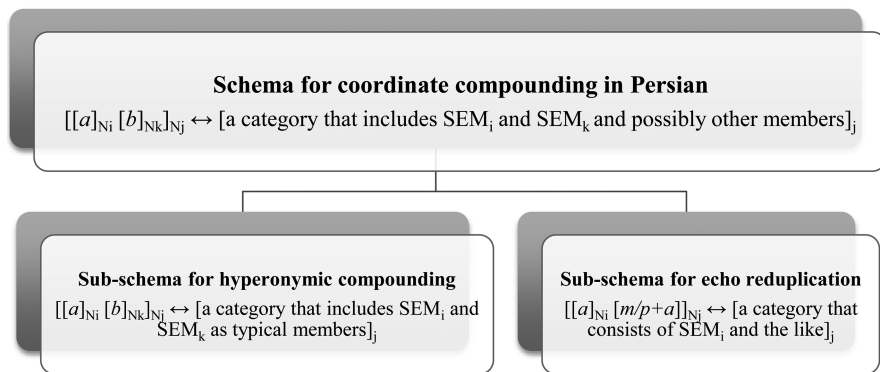
Figure 1: A partial hierarchy for coordinate compounding

Another aspect of the hierarchical lexicon that bears on the phenomena discussed here has to do with the fact that lexical entries do not only encode relationships among words and express schemas for word formation, the lexicon

---

23 As noted at the end of section 3, echo reduplication is limited to a highly informal register. It is an interesting challenge, left for future work, to theorize how such information about register and style should be represented as part of lexical entries.

also contains schemas for larger constructions such as constructional idioms. In other words, pieces of syntactic structure can be listed in the lexicon with their associated (constructional) meanings (see Jackendoff [2008], for example). Thus using the same descriptive formalism we can represent general extenders in Persian as construction idioms with the associative plural as a sub-schema:

Schema for general extenders in Persian

$[[a]_{XPi} [(\text{-}o)\ ina]]_{XPj} \leftrightarrow$ [a category that includes SEM_i and related things or activities]$_j$

Sub-schema for associative plural

$[[a]_{PNi} [\ ina]]_{NPj} \leftrightarrow [SEM_i$ and his/her family and close friends]$_j$
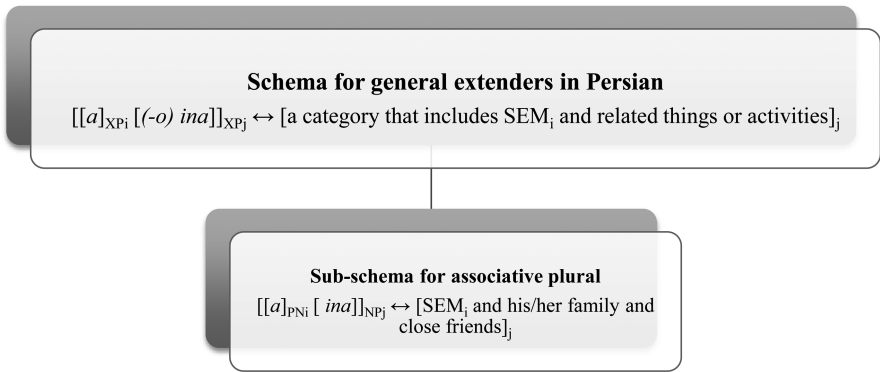
Figure 2: A partial hierarchy for general extenders

Ultimately these two partial hierarchies are linked, though the points of contact require further elaboration of both. The overarching schema is coordination as it is both a source for general extenders and is believed to be a source for co-compounds in at least some languages (Arcodia, Grandi, and Wälchli 2010: 8).

6 Conclusion

Starting with the associative plural construction, this article has explored some of the ways in which the denotation of nominal expressions can be extended in Persian. With proper names, the addition of *ina* yields the meaning of the person named, plus his or her family and close friends. With common nouns, coordinate compounding evokes categories consisting of at least one of the members of the compound and related entities. Both types of construction exhibit referential heterogeneity (Daniel and Moravscik 2013). That is, rather than denoting more than one entity of the same type, they denote sets of related entities. We have also seen the way in which general extenders also extend the denotation of a variety of constituents (nouns, verbs, verb phrases, etc.) in order to evoke a category of which the named constituent is an exemplar.

Having explored the resemblance among general extenders, co-compounds, echo reduplication, and associative plurals, we note that not all languages have

as full a range of constructions as Persian does. English, for example, has a productive general extender construction but no associative plural.[24] English also lacks a productive process of co-compounding. Arcodia, Grandi, and Wälchli (2010) claim that this type of compounding is "areally skewed" in that hyperonymic coordinate compounds are found in Eastern Eurasia (among other places) but not in Standard Average European languages, which have hyponymic coordinate compounds instead. (Hyponymic compounds have a referent that is a hyponym [subordinate] of the meaning of the parts.)

It is tempting to connect the lack of an associative plural that derives from a general extender construction in English to the lack of coordinate compounding in general. In other words, it might be the case that English lacks an associative plural because it lacks coordinate compounding. This is not to say that coordinate compounding itself gives rise to an associative plural, but that the step from general extender to associative plural is more easily made in a language that already has the means to stretch the denotation of a nominal to other related things and/or to create superordinate categories by coordination. If this hypothesis turns out to have some validity, it, in turn, supports the view of the lexicon put forward by Construction Grammarians. In a hierarchical lexicon, a densely populated network might give rise to new grammatical constructions more easily than a sparsely populated one.

The correlation between co-compounding and associative constructions remains to be explored further as does a diachronic investigation of Persian to determine which uses of *ina* predate others. The results of these studies will be revealing for those interested in associatives, coordination, and grammaticalization, and Persian has much to contribute in this regard.

Table 3: Abbreviations

1	first person	IMP	imperative
2	second person	NEG	negative
3	third person	PL	plural
CL	classifier	PN	proper name
CLC	clitic	POSS	possessive
CONJ	conjunction	PRS	present
CONT	continuous	PST	past
COP	copula	SBJ	subjunctive
EZ	ezafe	SG	singular

24 As mentioned in the introduction, some speakers can say [PN "'n them"] to mean something like an associative plural but the "'n them" does not necessarily refer to family and "'n them" is not otherwise a common general extender. This example does point to the fact that associatives may arise from coordinating constructions, however.

References

Ackema, Peter & Ad Neeleman. 2014. On the associative effect in first and second person plural pronouns. Talk presented at MGRG meeting, University of Edinburgh, 17 December 2014.

Aijmer, Karin. 2002. *English discourse particles: Evidence from a corpus*. Amsterdam: John Benjamins.

Arcodia, Giorgio F., Nicola Grandi & Bernhard Wälchli. 2010. Coordination in compounding. In Sergio Scalise & Irene Vogel (eds.), *Cross-disciplinary issues in compounding*, 177–197. Amsterdam: Benjamins.

Bauer, Laurie. 2010. Co-compounds in Germanic. *Journal of Germanic Linguistics* 22 (3). 201–219.

Boas, Hans C. 2013. Cognitive construction grammar. In Thomas Hoffman & Graeme Trousdale (eds.), *The Oxford handbook of construction grammar*, 233–254. Oxford: Oxford University Press.

Booij, Geert. 2010. *Construction morphology*. Oxford: Oxford University Press.

Borer, Hagit. 2005. *Structuring sense*. Vol. I. Oxford: Oxford University Press.

Brinton, Laurel. 1996. *Pragmatic markers in English. Grammaticalization and discourse functions*. Berlin: Mouton de Gruyter.

Canavan, Alexandra & George Zipperlen. 1996. CALLFRIEND Farsi LDC96S50. CD-ROM. Philadelphia: Linguistic Data Consortium.

Cheshire, Jenny. 2007. Discourse variation, grammaticalization and stuff like that. *Journal of Sociolinguistics* 11 (2). 155–193.

Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.

Cowper, Elizabeth & Daniel Currie Hall. 2012. Aspects of individuation. In Diane Massam (ed.), *Count and mass across languages*, 27–53. Oxford: Oxford University Press.

Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Daniel, Michael & Edith Moravcsik. 2013. The associative plural. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/36 (accessed on 16 November 2015).

Daniel, Mixail. 2000. *Tipologija associativnoj množestvennosti* [The typology of associative plurals]. Moscow: Russian State University for Humanities.

Erman, Britt. 1995. Grammaticalization in progress: The case of *or something*. In Inger Moen, Hanne Gram Simonsen & Helga Lødrup (eds.), *Papers from the XVth Scandinavian Conference of Linguistics, Oslo, January 13–15, 1995*, 136–147. Oslo: Department of Linguistics, University of Oslo.

Fillmore, C. J. 1988. The mechanisms of "construction grammar". In Shelley Axmaker, Annie Jaisser & Helen Singmaster (eds.), *Proceedings of the fourteenth annual meeting of the Berkeley Linguistics Society*, 35–55. Berkeley: Berkeley Linguistics Society, Inc.

Forbes, Clarissa. 2013. Associative plurality in the Gitksan nominal domain. In Shan Luo (ed.), *Proceedings of the 2013 annual conference of the Canadian Linguistic Association*. http://cla-acl.ca/actes-2013-proceedings/ (accessed 11 March 2017).

Fried, Mirjam. 2015. Construction grammar. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax – theory and analysis. An international handbook*, vol. 1: 974–1003. Berlin & Boston: De Gruyter Mouton.

Gebhardt, Lewis. 2008. Classifiers, plural and definiteness in Persian. In Simin Karimi, Vida Samiian & Donald Stilo (eds.), *Aspects of Iranian linguistics*, 35–56. Newcastle upon Tyne: Cambridge Scholars Publishing.

Gebhardt, Lewis. 2009. *Numeral classifiers and the structure of DP*. Evanston: Northwestern University dissertation.

Ghaniabadi, Saeed. 2008. Optionality and variation: A stochastic OT analysis of M/p-echo reduplication in colloquial Persian. In Simin Karimi, Vida Samiian & Donald Stilo (eds.), *Aspects of Iranian linguistics*, 57–84. Newcastle upon Tyne: Cambridge Scholars Publishing.

Ghaniabadi, Saeed. 2010. *The empty noun construction in Persian*. Winnipeg: University of Manitoba dissertation.

Ghaniabadi, Saeed. 2012. Plural marking beyond count nouns. In Diane Massam (ed.), *Count and mass across languages*, 112–128. Oxford: Oxford University Press.

Ghaniabadi, Saeed, Jila Ghomeshi & Nima Sadat-Tehrani. 2006. Reduplication in Persian: A morphological doubling approach. In Claire Gurski & Milica Radišić (eds.), *Proceedings of the 2006 annual conference of the Canadian Linguistic Association*. http://cla-acl.ca/actes-2006-proceedings/ (accessed 11 March 2017).

Ghomeshi, Jila. 2003. Plural marking, indefiniteness, and the noun phrase. *Studia Linguistica* 57 (2). 47–74.

Ghomeshi, Jila. 2008. Markedness and bare nouns in Persian. In Simin Karimi, Vida Samiian & Donald Stilo (eds.), *Aspects of Iranian linguistics*, 85–112. Newcastle upon Tyne: Cambridge Scholars Publishing.

Ghomeshi, Jila & Diane Massam. 2009. The proper D connection. In Jila Ghomeshi, Ileana Paul & Martina Wiltschko (eds.), *Determiners: Universals and variation*, 67–95. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Goldberg, Adele. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: Chicago University Press.

Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Görgülü, Emrah. 2011. Plural marking in Turkish: Additive or associative? *Working Papers of the Linguistics Circle of the University of Victoria* 21 (1). 70–80.

Haspelmath, Martin. 1993. *A grammar of Lezgian*. Berlin & New York: Mouton de Gruyter.

Hincha, Georg. 1961. Beiträge zu einer Morphemlehre des Neupersischen. *Der Islam* 37 (1–3). 136–201.

Inkelas, Sharon & Cheryl Zoll. 2005. *Reduplication*. Cambridge: Cambridge University Press.

Jackendoff, Ray. 2002. *Foundations of language*. Oxford: Oxford University Press.

Jackendoff, Ray. 2008. Construction after construction and its theoretical challenge. *Language* 84 (1). 8–28.

Jefferson, Gail. 1990. List construction as a task and resource. In G. Psathas (ed.), *Interaction competence*, 63–72. Lanham: University Press of America.

Kahnemuyipour, Arsalan. 2000. On the derivationality of some inflectional affixes in Persian. Paper presented at the annual meeting of the Linguistic Society of America Conference, Chicago, 8 January 2000.

Kahnemuyipour, Arsalan. 2003. Syntactic categories and Persian stress. *Natural Language & Linguistic Theory* 21 (2). 333–379.

Keane, Elinor. 2001. *Echo words in Tamil*. Oxford: Merton College dissertation.

Keane, Elinor. 2005. Phrasal reduplication and dual description. In Bernhard Hurch (ed.), *Studies in reduplication*, 239–262. Berlin & New York: Mouton de Gruyter.

Lazard, Gilbert. 1957. *Grammaire du persan contemporain*. Paris: Klincksieck.

Lazard, Gilbert. 1992. *A grammar of contemporary Persian*. Costa Mesa: Mazda Publishers.

Lidz, Jeffrey. 2001. Echo reduplication in Kannada and the theory of word-formation. *The Linguistic Review* 18 (4). 375–394.

Mauri, Caterina. Forthcoming. Building and interpreting ad hoc categories: A linguistic analysis. In Joanna Blochowiak, Cristina Grisot, Stephanie Durrlemann & Christopher Laenzlinger (eds.), *Formal models in the study of language.* Springer International Publishing.

Moravcsik, Edith. 2003. A semantic analysis of associative plurals. *Studies in Language* 27 (3). 469–503.

Nakanishi, Kimiko & Elizabeth Ritter. 2009. Plurality in languages without a count-mass distinction. Handout of paper presented at the Mass-Count Workshop, University of Toronto, 7 February.

Ourn, Noeurng & John Haiman. 2000. Symmetrical compounds in Khmer. *Studies in Language* 24 (3). 483–514.

Overstreet, Maryann. 1999. *Whales, candlelight, and stuff like that*. New York: Oxford University Press.

Overstreet, Maryann & George Yule. 1997. On being inexplicit and stuff in contemporary American English. *Journal of English Linguistics* 25 (3). 250–258.

Parvaresh, Vahid, Manoochehr Tavangar, Abbas Eslami Rasekh & Dariush Izadi. 2012. About his friend, how good she is, *and this and that*: General extenders in native Persian and non-native English discourse. *Journal of Pragmatics* 44 (3). 261–279.

Sedighi, Anousha. 2010. *Agreement restrictions in Persian*. Leiden: Leiden University Press.

Shaki, Mansour. 1967. *Principles of Persian bound phraseology*. Prague: The Oriental Institute in Academia, Publishing House of the Czechoslovak Academy of Sciences.

Sharifian, Farzad & Ahmad R. Lotfi. 2003. "Rices" and "waters": The mass-count distinction in modern Persian. *Anthropological Linguistics* 45 (2). 226–244.

Sharifian, Farzad & Ahmad R. Lotfi. 2007. "When stones falls": A conceptual-functional account of subject-verb agreement in Persian. *Language Sciences* 29 (6). 787–803.

Singh, Rajendra. 1982. On some "redundant compounds" in Modern Hindi. *Lingua* 56 (3–4). 345–351.

Stilo, Donald. 2004. Coordination in three Western Iranian languages: Vafsi, Persian and Gilaki. In Martin Haspelmath (ed.), *Coordinating constructions*, 269–330. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Wälchli, Bernhard. 2005. *Co-compounds and natural coordination*. Oxford: Oxford University Press.

Wiltschko, Martina. 2008. The syntax of non-inflectional plural marking. *Natural Language and Linguistic Theory* 26 (3). 639–694.

Shahrzad Mahootian and Lewis Gebhardt

13 Revisiting the status of -*eš* in Persian[1]

Abstract: Persian clitics, while not as studied as clitics in some other languages, have generated more interest in recent years. One of the main questions linguists seek to answer about clitics is what category they belong to, both in Persian in particular and across languages in general. This article has three goals regarding the status of Persian clitics, in particular the third person singular -*eš*. First, we review some of the recent literature that argues for at least some clitics being agreement markers. Second, we present data to remind us that -*eš* still exhibits robust clitic properties. Third, we report on a survey of speakers' judgments on the acceptability and meaning of -*eš* in various syntactic positions as well as introduce novel data that show clitics to be more elusive than previously thought.

Keywords: agreement, affixes, clitics, -*eš*, Persian

1 Introduction

Clitics have received increasing attention in recent decades. Some of the main problems in the study of clitics revolve around not only their syntactic and morphological behavior but also their categorial status. It is often difficult to tell whether a particular item is a clitic or an affix, since clitics typically exhibit properties of both categories, not to mention properties of independent words.

In this article we summarize recent work in the study of clitics in Persian and further look into their distribution, particularly that of the third person singular marker -*eš*. Previous work has determined that items like -*eš*, while they may have been clitics in earlier Persian, are or are in the process of becoming affixes, i.e., agreement markers (Rasekh 2011, 2014). Although we do not dispute that -*eš* may be an affix in certain environments, we focus on data that suggest that -*eš* does have robust clitic properties. However, our main

---

contribution is presenting speaker judgments suggesting that the status of *-eš* and other clitics is more complex and subtle than previously thought.

In the next section we review recent literature regarding clitics in Persian. In section 3, we present speaker judgments informally gathered from several native speakers, showing that the nature of *-eš* is even more elusive than previously thought. The data show that speakers vary significantly in their judgments over the grammaticality of sentences with clitics and in what those clitics mean. Nonetheless, a few preliminary generalizations emerge. Finally, we conclude with a look toward continuing with a larger sample of clitic data in Persian with a broader range of verbs, especially of compound verbs.

2 Background

Investigations into Persian clitics, as is the case with clitics cross-linguistically, typically show that the elements under study hover in status between that of agreement affixes and argument-denoting clitics. However, as a matter of terminology, following common practice in the literature, we refer to the items that are clearly agreement affixes as "affixes" and the elements that are not clearly affixes as "clitics", even though members of the "clitic" category are not a uniform category of items with only clitic properties. Indeed, the very point of much of the literature is to determine precisely which, if any, category "clitics" belong to.

First, there is general agreement on the affixes. Persian has a set of them, each agreeing with the sentential subject in person and number, although in Persian, a prodrop language, the subject need not appear overtly. The affixes are identical in present and past, except for third person singular.

Table 1: Persian verbal agreement markers[2]

Person	Present singular	Present plural	Past singular	Past plural
1	*-æm*	*-im*	-æm	*-im*
2	*-i*	*-id* (formal) *-in* (informal)	*-i*	*-id* (formal) *-in* (informal)
3	*-e* (informal) *-æd* (formal)	*-æn(d)*	-Ø	*-æn(d)*

In the present tense examples in (1a), the *mi-* prefix, which is mostly used in generic statements and progressives, is not relevant to the topic of the chapter.

---

2 Transcription is broad IPA with exceptions: with č = IPA ʧ, y = IPA j. In the glosses, Ez = ezafe linker.

The prefix is included because it is typically obligatory in the present. The examples in (1b) show the nearly identical past tense suffixes, the only difference being that in the past the third person singular is null. We assume these agreement markers without comment, except to note that their status as affixes is supported by two facts: they appear obligatorily on the verb and in compound verbs they appear only on the verbal element.

(1) a. Present forms of the verb *xordæn* 'to eat'

    (*mæn*)   *mi-xor-æm*   (*ma*)   *mi-xor-im*
    1SG   Prog-eat-1SG   1PL   Prog-eat-1PL
    'I'm eating'   'We're eating'

    (*to*)   *mi-xor-i*   (*šoma*) *mi-xor-id*
    2SG   Prog-eat-2SG   2PL   Prog-eat-2PL
    'You're eating'   'You're eating'

    (*u*)   *mi-xor-e*   (*unha*) *mi-xor-æn*
    3SG   Prog-eat-3SG   they   Prog-eat-3PL
    'she/he is eating'   'They're eating'³

   b. Past forms of the verb *xordæn* 'to eat'

    (*mæn*) *xord-æm*   (*ma*)   *xord-im*
    1SG   ate-1SG   1PL   ate-1PL
    'I ate'   'We ate'

    (*to*)   *xord-i*   (*šoma*) *xord-id*
    2SG   ate-2SG   2PL   ate-2PL
    'You ate'   'You ate'

    (*u*)   *xord-Ø*   (*unha*) *xord-æn*
    3SG   ate-3SG   they   ate-3PL
    'she/he ate'   'They ate'

The contrast between (2a) and (2b), with a compound verb, shows that the agreement suffix can only appear on the verb.

(2) a. *baz*   *kærd-i*
    open   did-2SG
    'Did you open it?'

   b. *baz-i   *kærd*
    open-2SG   did

---

3 The verb doesn't have to show plural agreement with nonhuman plural subjects.

The forms of the clitics are in Table 2.

Table 2: Persian clitics[4]

Person	Singular	Plural
1	-æm	-emun / -eman
2	-et / -æt	-etun / -etan
3	-eš / -æš	-ešun / -ešan

Clitic constructions parallel full DP constructions as possessives (3a, 3b), objects of prepositions (3c, 3d), and as objects of verbs (3e, 3f). Note also in (3e, 3d) that clitic doubling is possible if the subject *ma* 'we' is pronounced.

(3) a. *doxtær-æm*   b. *doxtær-e*   *mæn*
         daughter-3SG   daughter-Ez   I
         'my daughter'   'my daughter'

     c. *næzdik-æm*   d. *næzdik-e*   *mæn*
         near-1SG   near-Ez   I
         'near me'   'near me'

     e. (*ma*) *did-im-eš*   f. (*ma*) *u-ra*   *did-im*
         we   saw-1PL-3SG   we   3SG-Def.Acc   saw-1PL
         'We saw him/her/it'   'We saw him/her'

In compound verbs, which consist of a nonverbal element and verbal or light-verb element, the clitic may appear on either the light verb (4a) or the nonverbal element (4b).

(4) a. *komæk kærd-im-eš*   b. *komæk-eš kærd-im*
         help   did-1PL-3SG   help-3SG   did-1PL
         'We helped her/him'   'We helped her/him'
         (Mahootian 1997: 139)

As mentioned above, among the important issues in the study of clitics is to determine just what they are. Since clitics correspond to full-word forms but also have properties that overlap with those of affixes, a three-way series of tests must be appealed to in order to distinguish the three categories of affix, clitic,

---

4 The first form in each pair is the colloquial; the second is the formal form. Throughout this chapter we will be using the colloquial forms. The vowels in the clitics appear when the stem ends in a consonant. Also, we uniformly join the affixes/clitics to roots with a single hyphen.

and word (Zwicky 1985). Zwicky (1977, 1985) and Zwicky and Pullum (1983) proposed tests to distinguish clitics from other elements. The criteria for distinguishing clitics from affixes should be read as tendencies rather than strictly necessary or sufficient conditions. (Also see Klavans [1995] and Haspelmath and Sims [2010], among others, for discussion.)

– Clitics have more freedom in choosing a host, while affixes are highly selective in the category of host they attach to.
– Clitics have fuller paradigms than affixes, which often exhibit gaps.
– Clitics are quite regular whereas affixes show more morphophonological idiosyncrasies.
– Clitics are more semantically regular than affixes, which show more idiosyncrasies in their semantics.
– Syntactic rules don't affect clitic groups but may affect affixed words.
– Clitics can attach to clitics but affixes don't attach to clitics.

The more of the criteria that show clitichood for an item, the more clitic-like that item is. Items often don't behave exclusively like clitics or affixes, and, hence, the boundary between what a clitic is and what an affix is remains fuzzy (Haspelmath and Sims 2010: 197–203).

The fuzziness holds in Persian, where in this chapter we are primarily interested in the clitic/affix boundary, particularly with regard to the third person singular -*eš*. Some earlier studies, e.g., Lazard (1957), Mahootian (1997), have classified the elements in Table 2, above, as clitics, partly because of their ability to attach to various hosts. Samvelian and Tseng (2010) refer to the items in Table 2 (although in variant forms) as pronominal (en)clitics but they also point to the difficulty in Persian and other languages in differentiating them from affixes. Despite the difficulty in distinguishing clitics from affixes, their conclusion, based on certain phonological effects and co-occurrence constraints, is that these Persian items are more like suffixes than like "independent syntactic elements" (Samvelian and Tseng 2010: 213).

Rasekh (2011, 2014) takes a historical perspective in analyzing the status of clitics in impersonal constructions. The impersonal construction, illustrated in (5), has been much studied because of its apparently nonstandard agreement, where the verb itself shows the null third person singular agreement and the nonverbal element seems to agree with the overt or covert subject, which might be a topic.

(5)   *xoš-æm*   *'amæd*
      like-Enc.1SG   come.Past.3SG
      'I liked it'
      (Rasekh 2014: 17)

Rasekh determines that the clitic *-æm* in (5) should be treated as an agreement affix, arguing that the clitics are in the process of grammaticalizing into subject agreement markers. The affixal status of the clitic in (5), according to Rasakh, contrasts with its status as a true clitic in other cases, such as *-eš* in (6).

(6) (*mæn*) *did-æm-eš*
      (I)   see.Past-1SG.Su-Enc.3SG
      'I saw him/her'
      (Rasekh 2014: 2)

In effect, Rasekh proposes that an older agreement paradigm is evolving into a new paradigm, as in Table 3, as the clitic *-eš* evolves into an agreement affix that fills the third person singular gap in the overt endings of the old paradigm.

Table 3: Verbal agreement paradigm's change in Persian (Rasekh 2011: 25)

	Old Paradigm	New Paradigm
1SG	*-æm*	*-æm*
2SG	*-i*	*-i*
3SG	*-Ø*	*-eš*
1PL	*-im*	*-im*
2PL	*-id*	*-id*
3PL	*-ænd*	*-ænd*

Kazeminejad (2014) argues that in Persian pronominal complex predicates (i.e., Rasekh's impersonal construction) the pronominal clitic is a phrasal affix and agreement marker (also see Karimi [1997] on complex predicates). In contrast to Rasekh, who calls the null agreement on the verb itself in a sentence like that in (5) a "default / zero morph" (Rasekh 2014: 1), Kazeminejad argues that the null agreement should be interpreted as ordinary third person singular agreement with the subject, which is the theme argument of a state or event, in the case of (5) a state of liking. The enclitic itself agrees with the topic.

In the brief review of recent literature above, the consensus is that Persian clitics are or are evolving into agreement markers, at least in some constructions. In the next section we look at various data involving mostly *-eš* to further examine its status, with an eye toward reminding us of its still robust clitic properties.

3 More data on Persian *-eš*

In this section we recapitulate a number of facts about Persian clitics, focusing on the third person singular *-eš*, and then proceed to other observations.

The item *-eš* is most arguably a clitic when it linearly follows an affix, as in (7), where *-eš* follows the agreement marker *-im* on the verb.

(7) *ma   gorbe-ra   did-im-eš*
     we   cat-Def.Acc   saw-1SG-3SG
     'We saw the cat'

It is also a clitic when it appears in a location that bars agreement. As mentioned above, in compound verbs the agreement affixes can only appear on the light verb. Thus, if *-eš* appears on the nonverbal element in a compound verb, it must be a clitic as in (8a), alternating with the construction where the clitic can appear after the affix on the verbal element (8b).

(8) a. *mæn   be   færzad   nešan-eš   dad-æm*
         I   to   Farzad   indicate-3SG   gave-1SG
         'I showed it to Farzad'

     b. *mæn   be   færzad   nešan   dad-æm-eš*
         I   to   Farzad   indicate   gave-1SG-3SG
         'I showed it to Farzad'

But often it is simply unclear whether *-eš* is a clitic or affix. In (9), the *-eš* could be agreeing with the singular third person subject or functioning as a clitic for the subject. That the overt subject *maman* appears isn't conclusive evidence that *-eš* is an agreement marker since it is well established that clitics can double with overt arguments.

(9) *maman   birun   ræft-eš*
     mom   out   went-3SG
     'Mom went out'

If *-eš* is an agreement marker, then a clitic should be able to attach to its right, but this expectation is not borne out, as shown in (10a) and (10b). However, on the other hand, if *-eš* is a clitic we'd likewise predict (10a) and (10b) to be acceptable, since clitic sequences are typically permitted.

(10) a. *peyda* *kærd-eš-emun*
       find   did-3SG-1PL
       intended: 'She/he found us'

   b. *bær*   *dašt-eš-ešun*
       next to   had-3SG-3PL
       intended: 'she/he took them'

Generally, it seems, *-eš* has a definite/specific interpretation; if *-eš* is a clitic, then it co-refers to, or is co-indexed with, a specific earlier mentioned referent. Example (7) illustrates *-eš* as co-referential with the definite *gorbe-ra* 'the cat' and both examples in (8) show that *-eš* refers to a previously introduced referent, translated as 'it'. However, in asking several native Persian speakers a variety of sentences (discussed below), we did find a few examples where *-eš* seemed to be acceptable with an indefinite reading.

(11) *yek* *gorbe* *yek* *muš*   *xord-eš*
     a   cat   a   mouse   ate-3SG
     'A cat ate a mouse'

In (11), both arguments are indefinite, yet all four speakers that we consulted found the sentence acceptable, with a preference for interpreting *-eš* as referring to the subject. This fact could be evidence that *-eš* in (11) marks agreement, since there is no evidence elsewhere in Persian that (in)definiteness is a factor in agreement. This point is also discussed in Fuß (2005: 133).

    On the other hand, if *-eš* is an agreement marker, it should allow a clitic to follow, but *-eš* seems to resist having a clitic after it, as in (12a). In fact, *-eš* can't be followed by itself (12b) or any of the other clitics (12c).

(12) a. *did-eš-æm*
       saw-3SG-1SG
       intended: 'She saw me'

   b. *did-eš-eš*
       saw-3SG-3SG
       intended: 'She/he saw him/her'

   c. *did-eš-*et/*emun/*etun/*ešun*
       saw-3SG-2SG/1PL/2PL/3PL
       intended: 'She saw you/us/you(PL)/them'

The data in (12a, 12b, and 12c) argue for *-eš* being a clitic.

    Since Persian is a prodrop language, it's possible that the antecedent of a clitic may not be in the same clause, at least not overtly, and may have been

introduced in an earlier clause. In (13), *-eš*, which can be translated as 'it', appears at the end of the second clause, after the agreement marker *-æm* while the shirt it's referring to is mentioned in the first clause, *in piran*.

(13) *in   piran-e   xeyli čerk-e   emruz   bayæd   be-šur-æm-eš*
      This   shirt-Def   very   dirty-is   today   must   Subjunctive-wash-1SG-3SG
      'This shirt is very dirty, I'll have to wash it today'
      (http://persian.nmelrc.org/pvc/pvc.php?verb=shostan&snt,
      accessed 15 October 2015)

The examples so far suggest that *-eš* has both affixal and properties as an object clitic, but it is arguably a subject clitic as well. Of course the problem here is that if *-eš* is related to the subject, it might be an agreement marker, as previous literature suggests. In (9), for example, it's difficult to tell whether the *-eš* simply agrees with the third person subject *maman* 'Mom' or is an argument clitic.

In compound verbs, a curious distribution results in the positions of the clitics. In transitive compound verbs, *-eš* can appear on either the verbal (14a) or nonverbal element (14b).

(14) a. *peyda   kærd-æm-eš*
          find   did-1SG-3SG
          'I found it'

      b. *peyda-š   kærd-æm*
          found-3SG   did-1SG
          'I found it'

As mentioned in the literature review in the previous section, in a subclass of compound verbs, the impersonal constructions, the *-eš* appears only on the non-verbal element as in (15a). Recall that this *-eš* is argued to be an agreement affix by Rasekh, obligatorily appearing on the nonverbal element and never the verbal element. However, we find that (15b) is acceptable, with the sentence-final *-eš* in some kind of relationship with *færzad*. If *-eš* is a third person agreement affix, at least in some contexts, it might be predicted that (15b) should be good, with *-eš* being Rasekh's new-paradigm affix replacing *-Ø*. But if it's not an agreement affix, then it is a subject clitic.

(15) a. *xoš-eš   umæd*
          like-3SG   came
          'She/he liked it'

      b. *færzad   æz   to   xoš-eš   umæd-e*
          Farzad   from you   like-3.SG came-3.SG
          'Farzad liked you'

Sentences like (15) need further investigation and for now we put them aside, accepting the generalization that impersonal verbs have obligatory agreement on the nonverbal element.

Finally, in intransitives, excluding the impersonals, *-eš* appears only on the verbal element, optionally.

(16) a. *leila   rah   oftad(-eš)*
          Leila   way   fell-3SG
          'Leila got going'

      b. **leila   ra(h)-(e)š   oftad*
          Leila   way-3SG   fell
          intended: 'Leila got going'

Let's summarize the patterns. The *-eš* appears optionally on either element of a transitive compound verb. It is barred from the nonverbal element in intransitives and from the verbal element in experiencer intransitives. Finally, it is obligatory on the nonverbal element in experiencer intransitives. The facts appear in Table 4. In standard notation, an element in parentheses is optional and an asterisk outside the parentheses indicates that the element is obligatory.

Table 4: The appearance of *-eš* in different compound verb types

Verb type	Nonverbal element	Verbal element
transitive	(*-eš*)	(*-eš*)
intransitive	**-eš*	(*-eš*)
experiencer intransitive	**(-eš)*	**-eš* (but see example [15])

One potentially interesting asymmetry is the unacceptability of *-eš* on the nonverbal element of ordinary intransitive verbs compared with the unacceptability of *-eš* on the verbal element of the experiencer intransitives. Another asymmetry is that in transitive compound verbs, *-eš* optionally appears on the nonverbal element or on the verbal element, while for both kinds of intransitives there is some restriction on the appearance of *-eš*. Finally, only in the case of the experiencer intransitives is *-eš* required. Indeed this obligatory appearance is one reason for thinking *-eš* here is an agreement affix. But if optionality is taken as a diagnostic, all the other appearances of *-eš* could be as clitics.

Recapitulating, there is mixed evidence as to whether *-eš* is a clitic or affix. Relying on data in the literature and our own judgments, we were nonetheless curious to get input from at least a small sample of nonlinguists. The judgments were gathered by asking four adult native speakers of Persian (three university

students and one homemaker) about the use of *-eš*, through both grammaticality judgments and intuitions about meaning. The sentences were presented in written Persian and participants were asked to give an acceptability judgment and to state whether the *-eš* refers to the subject or the object. Because of the small sample size, the results are not conclusive, but we think they are suggestive. It should be noted that some of the responses were anomalous or nonsensical and we didn't include them. As an example of an anomalous result, for one sentence participants said that *-eš* referred to an object, although the verb was intransitive. Further, although we included them, some judgments didn't agree with ours, with participants claiming ungrammaticality where we considered them acceptable. While of course judgments may vary, there were cases where we felt the participants' judgments of unacceptability stemmed from lack of context, since we presented the sentences in isolation and in written rather than in spoken form, where in some cases intonation would have clarified acceptability or ambiguity. For example, in (xii) in Table 5, the speakers all judged the sentence acceptable and said that *-eš* referred to the object. However, an intonation shift makes it easy to interpret the clitic as referring to the subject. Similarly, a different intonation allows for the clitic to be interpreted as co-referential with the subject instead of the object. Also, in a number of cases participants had a consistent judgment of an unambiguous referent of *-eš* where we found ambiguity. This too must be more rigorously studied, but for now we suggest that some of the unambiguous readings for participants were really preferred readings, with another reading being possible.

In Table 5, we summarize the results of our survey. In the left column is the sentence, with our intended meaning. In each of the other four columns is a speaker's rating of acceptability and meaning. In a few instances participants judged a sentence as unacceptable but commented on what *-eš* was anyway. On the top row of each cell of the responses are their acceptability judgments: G for grammatical, U for ungrammatical. The notation on the second row of each cell of speakers' responses indicates what the speakers thought the referent of *-eš* was: S if referring to the subject and O if referring to the object. S/O indicates that the speaker thought it could refer to either. In sentences where *-eš* appeared twice, we indicated the left *-eš* as 1 and the second *-eš* as 2. So "1-O 2-S/O" means that the speaker thought the first *-eš* referred to the object and the second *-eš* could refer to either the subject *or* the object. Note that speaker 4 didn't complete the questions.

As can be seen in Table 5, the results are mixed, which perhaps isn't surprising given common analyses of *-eš* as affix and/or clitic. Participants were entirely consistent on some sentences: for example, in (xvii) all four participants agreed that *-eš* referred to the object. However, for others there was variation on whether the sentence is grammatical and what the clitics referred to.

Table 5: Speakers' judgments on the use of *-eš*

Persian sentence & intended meaning	Speaker 1 Acceptability Meaning of *-eš*	Speaker 2 Acceptability Meaning of *-eš*	Speaker 3 Acceptability Meaning of *-eš*	Speaker 4 Acceptability Meaning of *-eš*
i. *bidar-eš kærd-eš* 'She woke him up'	U 1-O 2-S/O	U 1-O 2-O	G 1-O 2-O	
ii. *gorbe yek muš xord-eš* 'The cat ate a mouse'	U S	U O	G O	
iii. *širin bidar-eš kærd-eš* 'Shirin woke him up'	U 1-O 2-S	U 1-O 2-O	G 1-O 2-S/O	
iv. *xord-æm-eš* 'I ate it'	G O	G O	G O	
v. *gerɛft-ænd-et* 'They caught you'	G O	G O	G O	
vi. *xanum-ha resid-ænd-ešun* 'The women arrived'	G S	G S	G S	
vii. *xord-id-eš* 'You ate it'	G O	G O	G O	
viii. *ketab-ra xund-eš* 'She read the book'	G S	G O	G S/O	
ix. *yek gorbe yek muš xord-eš* 'A cat ate a mouse'	G S	G S	G S/O	
x. *leila ketab-ra xund-eš* 'Leila read the book'	G S/O	U S	G S/O	
xi. *peyda-š kærd-eš* 'She found it'	G 1-O 2-S/O	U 1-O 2-O	G 1-S/O 2-S/O	G 0-O
xii. *kia xord-eš* 'Kia ate it'	G O	G O	G O	G O
xiii. *xabid-eš* 'She slept'	U S	U ?	G S	G S
xiv. *xab-eš bord* 'She fell asleep'	G S	G S	G S	G S
xv. *peyda-š kaerd* 'She found it'	G O	G O	G O	G O
xvi. *gorbe muš xord-eš* 'The cat ate a mouse'	G S/O	G O	G S	U S
xvii. *bidar-eš kærd* 'She woke him up'	G O	G O	G O	G O
xviii. *xord-eš* 'She ate it'	G O	G O	G O	G S/O
xix. *bidar kærd-eš* 'She woke her up'	G O	G O	G O	
xx. *peyda kærd-eš* 'She found it'	G O	U O	G O	

There was a tendency for speakers to see the -*eš* on the nonverbal element as connected to the object, while the clitic on the light verb was more open to a subject-oriented reading. This is further illustrated in the following contrast. In (17a) speakers interpreted -*eš* on the nonverbal element as referring to the object, *faerhad*, while in (17b) they saw -*eš* on the verbal element as referring to the subject, *širin*.

(17) a. *širin*   *[færhad-o]ᵢ*   *peyda-[š]ᵢ*   *kærd*
Shirin   Farhad-Def.Acc   find-3SG   did
'Shirin found Farhad'

b. *širinᵢ*   *færhad-o*   *peyda*   *kærd-[eš]ᵢ*
Shirin   Farhad-Def.Acc   find   did-3SG
'Shirin found Farhad'

4 Conclusion

In this chapter we noted a tendency in the recent literature to analyze clitics, at least in some constructions, as affixes. We lodged no major objection to the idea that Persian clitics may be becoming affixes, and certainly it is the case that, as clitics in many languages, Persian clitics are not clearly one category or another. Further, we presented data to remind us that there's still a lot of cliticness in clitics. Taking a more neutral stance in collecting data from native speakers in a preliminary informal survey, we simply asked participants "Is -*eš* the subject or the object?", leaving aside as to whether they are more clitic-like or affix-like in any given sentence. Also, though not conclusive, some patterns emerged, particularly a preference for -*eš* on the nonverbal element in compound verbs to be associated with the object and -*eš* on the verbal element to be associated with the subject.

In further study, we will sample a wider range of ordinary verbs and compound verbs to test native speakers' understanding of -*eš*, expecting that a large sample will provide for stronger generalizations that will help to more accurately address the question of what -*eš* is in particular, and what clitics are more generally in Persian.

References

Fuß, Eric. 2005. *The rise of agreement: A formal approach to the syntax and grammaticalization of verbal inflection*. Philadelphia: John Benjamins.

Haspelmath, Martin & Andrea D. Sims. 2010. *Understanding morphology*. London: Hodder Education.

Karimi, Simin. 1997. Persian complex verbs: Idiomatic or compositional. *Lexicology* 3 (2). 272–318.

Kazeminejad, Ghazaleh. 2014. Pronominal complex predicates in colloquial Persian. Lexington: University of Kentucky dissertation.

Klavans, Judith. 1995. The independence of syntax and phonology in cliticization. *Language* 61 (1). 95–120.

Lazard, Gilbert. 1957. *Grammaire du persan contemporain*. Tehran: Institut Français de Recherche en Iran.

Mahootian, Shahrzad. 1997. *Persian*. New York: Routledge.

Rasekh, Mohammad. 2011. The rise of agreement: The case of Persian enclitics. In *International Conference on Languages, Literature and Linguistics. IPEDR 26*. Singapore: IACSIT Press.

Rasekh, Mohammad. 2014. Persian clitics: Doubling and agreement. *Journal of Modern Languages* 24 (1). 16–33.

Samvelian, Pollet & Jesse Tseng. 2010. Persian object clitics and the syntax-morphology interface. In Stefan Müller (ed.), *Proceedings of the HPSG 10 Conference*, 212–232. Paris: CSLI Publications.

Zwicky, Arnold. 1977. On clitics. Bloomington: Indiana University Linguistics Club.

Zwicky, Arnold. 1985. Clitics and particles. *Language* 61 (2). 283–305.

Zwicky, A. & G. K. Pullum. 1983. Cliticization vs. inflection: English *n't*. *Language* 59 (3). 502–513.

Arseniy Vydrin
14 'Difficult' and 'easy' in Ossetic

Abstract: In this article, I will study the peculiarities of the Ossetic dedicated construction conveying the meanings 'difficult/easy to accomplish'. By dedicated I mean that the construction expresses nothing but the intended meaning. The construction is not mentioned in Ossetic standard grammars and is unique among Iranian and Caucasian languages. The construction is close to passive constructions and a modal construction of possibility, however, its origin and the time of appearance in Ossetic is not clear.

Keywords: modality, grammaticalization, Ossetic, Iranian languages, Caucasian languages

1 Introduction

Though the meanings 'easy/difficult to accomplish' can be expressed by grammatical means, e.g., via syntactic constructions, these constructions are usually not dedicated ones and have other meanings. Cf. the so-called Tough Construction (Comrie and Matthews 1990), which besides *easy* and *difficult* can be used with other VALUE adjectives, e.g., *This room is pleasant to sleep in*.

This article deals with a typologically rare grammaticalization of the modal meaning 'easy/difficult to accomplish', which is found in the modern Ossetic language (Eastern Iranian). In section 2, I give general information on the Ossetic language and clarify the sources of my data. Section 3 deals with grammatical ways of expressing the meaning 'easy/difficult to accomplish'. In sections 4, 7, and 8, I describe formation, morpho-syntactic properties, and semantics of the dedicated construction expressing the meaning 'easy/difficult to accomplish'. In sections 5 and 6, I compare the standard passive constructions and complex verbs to the dedicated construction under discussion. Section 9 is an attempt to trace the origin of the construction. Section 10 shows that

none of other Iranian languages and almost none of the Caucasian languages geographically close to Ossetic have dedicated grammatical means to express the meaning 'easy/difficult to accomplish'. Section 11 is the conclusion.

2 Ossetic language

Ossetic is an Eastern Iranian language that is mainly spoken in the Caucasus, in the Republic of North Ossetia-Alania, Russia, and in the Republic of South Ossetia. The total number of Ossetic native speakers in the world is around 529,000 (Ethnologue 2017). Ossetians living in Russia are usually bilingual (Russian and Ossetic).

Ossetic has two major dialects: Iron and Digor. This article focuses on the Iron dialect on which the literary standard Ossetic is based. However, there are some published texts (fiction, poetry, newspapers) in Digor Ossetic as well (cf. our Digor Ossetic written corpus http://corpus-digor.ossetic-studies.org/en).

The Ossetic examples cited in the paper are taken from three sources: Ossetic National Corpus, Ossetic oral texts, and my field data. Ossetic National Corpus (ONC) is available online (http://corpus.ossetic-studies.org/en) and mainly consists of contemporary texts from fiction and the literary journal *Max dug* published in North Ossetia in 2000–2012. All the texts have been automatically morphologically annotated in English and Russian. By the time of submission, ONC had about 5 million tokens. The examples from the Ossetic National Corpus are marked as ONC and have a reference to the source. The oral texts are also available online (http://www.ossetic-studies.org/en/texts/iron) and consist of Iron dialect texts recorded in different parts of North Ossetia in 2007–2012. All the oral texts are transcribed, translated, and interlinearized in English and Russian. By the time of submission, the oral text had about 50,000 words. The examples from the oral texts are marked as Oral text and have the name of the text and the sentence's number. My field data was collected in North Ossetia from 2008 to 2010. Examples obtained from native speakers do not have any special marking.

3 "Difficult to accomplish" and "easy to accomplish" in Ossetic

Ossetic has several grammatical ways to convey the meanings 'difficult to accomplish' and 'easy to accomplish'. In what follows, I will briefly discuss them.

3.1 Dative construction

The most usual way to express the meanings 'difficult to accomplish' and 'easy to accomplish' is to use the construction where the adjective *ᴣnson* 'easy' or *žən* 'difficult' forms a predicate with the verb *wᴣvən* 'to be' in third person singular or with the habitual verb of being *vᴣjjən*. The lexical verb is used in the infinitive (-*ən* ending) and the Experiencer is marked by the dative. The construction is similar to English: "It is difficult / easy for someone to do something". I will call it the dative construction. Compare the examples with *žən* 'difficult' (examples [1])[1] and [3]) and *ᴣnson* 'easy' (example [2]) and with a transitive (example [1]) and an intransitive (example [3]) verb.

(1) *nᴣ   tsᴣ   ᴣrba-jjᴣft-on   ᴣmᴣ mən   žən*
     NEG 3PL.ENCL.GEN PREF-catch.up-PST.TR.1SG and 1SG.ENCL.DAT difficult
     u   tsᴣ   nom   zur-ən
     be.PRS.3SG POSS.3SG name say-INF
     'They had died before I met them and it is difficult for me to say their names' (Oral text. Zangieva_Khabalova_2. 71.1).

(2) *ᴣmᴣ iwərdəgᴣj   mᴣ-χisᴣn   dᴣr təng ᴣnson vᴣjj-ə*
     and from.one.side POSS.1SG-RFL.DAT FOC very easy   be.HAB-PRS.3SG
     wəj   təχχᴣj ᴣmᴣ ž warž-ən   iron   ᴣvžag
     that.GEN POST and I   love-PRS.1SG Ossetic language
     'On the one hand, it is very easy for me [to speak Ossetic] because I love Ossetic language' (Oral text. Alagir school. 2.9).

(3) *χᴣrž   žən   u   mᴣl-ən, me   fšᴣmᴣr*
     very   difficult be.PRS.3SG die-INF POSS.1SG brother
     ᴣmbažəg...   ᴣmbažəg...
     companion.in.arms   companion.in.arms
     'It is very difficult to die, my brother companion-in-arms... companion-in-arms'
     (ONC. Max dug, no. 7, 2003).

---

1 All the examples are transcribed in accordance with the modern standard (Iron) Ossetic pronunciation (for details see Dzahova 2009). In general, the most important phonetic peculiarities are connected to the pronunciation of the consonants (in Ossetic script) *c*, *ᴣ*, *ðᴣ*, and *y̨*. In most of the cases, the letter *c* is pronounced as /š/, *ᴣ* as /ž/, *ðᴣ* as /z/, and *y̨* as /s/. While transcribing the examples I used the IPA symbols. Clitics are used without the symbol = and follow the Ossetic orthography (sometimes they are written separately, sometimes with a hyphen).

Note that in the dative construction the auxiliary is obligatory in third person singular. The construction is not restricted to the meanings 'difficult' and 'easy'; other VALUE adjectives also can be used in it. Compare the following example with the adjective *ɜχson* 'pleasant':

(4) *Asimo-imɜ   nəχaš   kɜn-ən   adɜjmag-ɜn   ɜχson   u*
     Asimo-COM   word   do-NMLZ   man-DAT   pleasant   be.PRS.3SG
     'It is a pleasure for a human being to talk with Asimo (robot's name)'
     (ONC. Max dug, 2003, no. 5, p. 152).

The lexical verb can be omitted in the dative construction. Compare:

(5) *žən   nən   zə   wədi   təng   žən*
     difficult   1PL.ENCL.DAT   3SG.ENCL.INESS   be.PST.3SG   very   difficult

     nən   zə-iw   wədi   fɜlɜ   wɜddɜr
     1PL.ENCL.DAT   3SG.ENCL.INESS-ITER   be.PST.3SG   but   however
     'It was hard there, it was very hard for us [to live there], though (we don't complain)' (Oral text. Aguzarova Izeta. 33.4).

3.2 Facilitive-difficilitive construction

There is another construction in Ossetic, which also consists of the adjective *ɜnson* 'easy' or *žən* 'difficult', the auxiliary *wɜvən* 'to be' (or the habitual auxiliary of being *vɜjjən*), and a verbal derivate of a lexical verb. A Patient-like participant is marked by the nominative while an Agent-like participant[2] or Sole participant is obligatory in the dative. When the construction is formed from a transitive verb, the auxiliary agrees in person and number with a Patient-like participant.[3] Compare the examples (6) to (8) below: example (6) shows the use of the construction

---

2 The term Agent-like participant is broader than A(gent), see example (31) where the unexpressed first participant of the verb 'to forget' is Experiencer rather than Agent. Agent-like participant means "the main participant, the 'hero' of the Situation, who is primarily responsible for the fact that this Situation takes place" (Kibrik 1997: 292). For this participant, Kibrik proposed the hyperrole Principal. Patient-like participant is "the most Effect(Patient)-like participant of a multi-participant event" (Kibrik 1997: 292). Kibrik uses the term *Patientive* for this participant.
3 The basic strategy of argument case marking is nominative for Sole participant of an intransitive verb and for the first participant of a transitive verb (Subject) and nominative/genitive for the second participant of a transitive verb (Object). The choice between the nominative and the genitive depends on animacy and definiteness (the genitive only if an Object is animate).

with the adjective 'easy', while in examples (7) and (8), the adjective 'difficult' is used. In examples (6) and (7), the construction is formed from a transitive verb, in example (8) from an intransitive.

(6) nʒl   fəš-ə   dəmʒg-aw   ʒnson kʒrd-ʒn   štə
    male sheep-GEN fat.tail-EQU easy   cut-NMLZ be.PRS.3PL
    'It is easy to cut them, like a sheep's fat tail' (Abaev 1959: 112).

(7) asə fəš-tʒ   nən   žʒn   ʒrs-aχš-ʒn   štə
    this sheep-PL.NOM 1PL.ENCL.DAT difficult PREF-catch-NMLZ be.PRS.3PL
    'It is difficult for us to catch these sheep'
    (lit. 'these sheep are difficult for catching for us').

(8) ʒnʒ   dʒw   mən   žʒn   sʒr-ʒn   u
    without you.SG 1SG.ENCL.DAT difficult live-NMLZ be.PRS.3SG
    'It is difficult for me to live without you'
    (ONC. Kokaev T. A. Nebesnyj ključ. Vladikavkaz, 2004, p. 91).

The construction differs from the dative construction morphologically, syntactically, and semantically. First of all, the construction is restricted to only two adjectives – ʒnson 'easy' or žʒn 'difficult' – and conveys only 'easy to accomplish' or 'difficult to accomplish' respectively; compare example (6) to (7) and (8). Hereinafter I will refer to it as facilitive-difficilitive construction,[4] or FDC, from Latin *facilis* 'easy' and *difficilis* 'difficult'. Secondly, in FDC, a lexical verb is obligatory used in the non-finite form in *-ʒn* (the properties of this verbal form will be considered in section 4.1 below). Finally, the auxiliary in FDC has full person and number paradigm and agrees in person and number with a Patient-like participant (if the lexical verb is transitive); compare examples (6) and (7) where the auxiliary is in third person plural; an Agent-like participant is obligatory marked by the dative; cf. example (7). If the lexical verb is intransitive, the auxiliary is used in third person singular and a Sole participant is marked by the dative (example [8]). It is important to point out that, when FDC is formed from a transitive predicate, the auxiliary can agree only with a Patient-like participant; cf. (9) where the auxiliary agrees with the Recipient and the sentence is ungrammatical:

---

4 It could be split up into two constructions (the facilitive construction and the difficilitive construction); however, the use of different adjectives does not affect the morphology or syntax of the construction. I have decided to consider it as a single facilitive-difficilitive construction.

(9) *wədon ʒnson gɜrtam ratt-ɜn štə
they easy bribe give-NMLZ be.PRS.3PL
Intended meaning: 'It is easy to give them a bribe'.

Also note that when FDC is formed from a ditransitive predicate with an overt Agent (e.g., 'it was difficult for **us to give her** a great deal of assistance'), the Recipient (which is also marked by the dative in Ossetic) is obligatorily omitted; cf. the following, where example (10) allows the insertion of an Agent (example [10a]), but it does not allow the simultaneous expression of Agent and Recipient (example [10b]).

(10) sɜj žən dɜtt-ɜn dɜ, ud, ɜmɜ sɜj žənary
PRTCL difficult give-NMLZ be.PRS.2SG soul and PRTCL valuable

dɜ, sard
be.PRS.2SG life

'How is it difficult to give you, the soul (lit. 'how are you difficult for giving, the soul'), and how are you valuable, the life!'
(ONC. Degoeva S. M. Pogasšyj luč solnca. Vladikavkaz: Ir, 2002).

a. sɜj žən dɜtt-ɜn **mən** dɜ, mɜ ud
PRTCL difficult give-NMLZ 1SG.ENCL.DAT be.PRS.2SG POSS.1SG soul
'How it is difficult for me to give you, my soul!'

b. *sɜj žən dɜtt-ɜn mən **ɜn** dɜ, ud
PRTCL difficult give-NMLZ 1SG.ENCL.DAT 3SG.ENCL.DAT be.PRS.2SG soul
Intended meaning: 'How it is difficult for me to give you, the soul, to **him**!'
(lit. 'how are you difficult for giving, the soul, for me to him').

The omission of Recipient in this case is explained by the fact that its place is already occupied by the dative participant (Agent).

In FDC formed from a transitive predicate, the auxiliary obligatory agrees with a Patient-like participant in person but not always in number; cf. the following example where the auxiliary is in plural, while the Patient is in singular:

(11) žən dar-ɜn wədəštə utɜppɜt sot
difficult keep-NMLZ be.PST.3PL so.much posterity
'It was difficult to keep so many descendants'
(ONC. Biragova L. H., Agkaceva L. T. Sbornik diktantov. Vladikavkaz: Ir, 2005, p. 127).

Disagreement in number is common in Ossetic if the subject is a collective noun (Ahvlediani 1969: 123).

Though the dative construction is well known to Ossetic grammars, FDC is mentioned neither in Ossetic standard grammars (e.g., Abaev 1964) nor in dedicated studies of Ossetic modality (e.g., Tehov 1970) and the verb (e.g., Takazov 1992). In the following sections, I will discuss only FDC. In the next three sections, I will consider the FDC's constituents, namely, the verbal form in *-ɜn*, the adjectives *žǝn* 'difficult' and *ɜnson* 'easy', and the auxiliary *wɜvǝn* 'to be'. I will show that, though FDC is similar to the passive construction and complex predicates, it differs from both of them and has to be considered a separate construction (sections 5 and 6). In sections 7, 8, 9, and 10 I will discuss FDC's peculiarities, semantics, and origin.

4 FDC's constituents

4.1 Verbal derivate in *-ɜn*

The verbal derivate in *-ɜn*, which is used in FDC, is formed from the present stem of a verb by the suffix *-ɜn*. The origin of the suffix is uncertain. There were some attempts to connect it to the Indo-Iranian verbal suffix *-anā* (Takazov 1992: 108–137); however, *-ɜn* cannot continue *-anā* for phonetic reasons. Thordarson proposes to derive *-ɜn* from the Old Iranian nominal suffix **-ana-* (Thordarson 2009: 145–146). The suffix *-ana-* was used in Old Aryan to derive verbal abstracts, "names of tools and places suitable or intended for some action" (Thordarson 2009: 146): Old Indian *sam-áraṇa*, Old Persian *ham-arana* 'battle' (*ar-* 'to move'). The Ossetic dative case was also derived from the suffix **-ana*. The dative case and the verbal suffix *-ɜn* express close functions: the meaning 'to, for', the direction or destination of the referent. "It is therefore natural to presume that at some stage in the history of the language the derivative suffix **-ana* was grammaticalised as a case ending" (Thordarson 2009: 146).

The verbal derivate in *-ɜn* can be used as a head of a noun phrase (*χiž-ɜn* graze.PRS-NMLZ 'pasture') or an attribute (*fəšš-ɜn zawma-tɜ* write.PRS-NMLZ thing-PL.NOM 'writing-materials'). Besides FDC, the derivate can be used only in the modal construction of participant-external possibility, which will be briefly discussed in section 9.

In the examples, for convenience I will interlinearize the suffix *-ɜn* as NMLZ (nominalization), however, the nominalizing function of the suffix needs a separate study.

4.2 Nouns *ɜnson* and *žən*

The words *ɜnson* and *žən* used in FDC can function either as attributes ('easy', 'difficult') or nouns ('easiness', 'difficulty').

The origin of *žən* 'difficult(y)' is connected to the Avestian *zyā-* 'to damage', cf. Sogdian **žīn-* 'heavy, painful' (Abaev 1989: 322). *ɜnson* 'easiness, easy' is a parallel form of *ɜnsoj* 'rest', which originates from **ham-č(y)āna*, *č(y)ā-*'rest', 'enjoy' and the prefix *ham-* (Abaev 1958: 151–152).

Besides FDC, *ɜnson* and *žən* can be used independently as attributes (12) or heads of an noun phrase (*ɜnson* 'convenience' and *žən* 'difficulty'). The essential peculiarities of the Ossetic noun phrase are as follows (for more details, see Thordarson [2009]): an attribute and any dependent noun phrase precedes its head; an attribute has no case or number marking; possessive proclitics and possessive noun phrases precede the first attribute of the noun phrase; the noun phrase cannot be split by any external material, including clitics or particles; the only inflexional suffix used with adjectives is the comparative suffix *-dɜr*. The example below shows a noun phrase with the possessive noun phrase and the comparative suffix *-dɜr* used with the adjective:

(12) *šərdon ta jɜ sard-ə [žən-dɜr bon-t-ɜm]*_{NP} *ɜr-χɜссɜ...*
Syrdon CONTR POSS.3SG life-GEN difficult-COMPAR day-PL-ALL PREF-reached
'Syrdon again has reached the most difficult days of his life'
(ONC. Džusojty N. G. Sljozy Syrdona. Vladikavkaz, 2004, p. 640).

According to ONC data, when used in FDC, *ɜnson* 'easy' or *žən* 'difficult' function as attributes to the verbal derivate in *-ɜn* and the word combination <*ɜnson* or *žən* + form in *-ɜn*> can be analyzed as a noun phrase.[5] In FDC, *ɜnson* 'easy' or *žən* 'difficult' are always preposed to the verbal form in *-ɜn* and cannot be separated from it by other words, clitics, or particles. The second-position enclitic pronouns cannot be put between *ɜnson* 'easy' / *žən* 'difficult' and the verbal form in *-ɜn;* cf. the following example, where the enclitic is obligatorily used after the verbal form in *-ɜn*:

(13) *ɜnson kuš-ɜn **zə** u?*
easy work-NMLZ 3SG.ENCL.INESS be.PRS.3SG
'Is it easy (comfortable) to work in it (in the tails)?'
(ONC. Max dug, no. 5, 2003, p. 130).

---

5 However, it should be pointed out that, when used in FDC, the head of the noun phrase cannot vary in case and number.

Used in FDC, *ӡnson* 'easy' or *žən* 'difficult' cannot have case or number affixes; however, it can be marked by the comparative suffix *-dӡr*; cf.:

(14) *žӡχχ-ə* *ӡmӡ* *foš-ə* *kʷəštӡg-tӡ* *χӡχbӡšt-ə*
land-GEN and cattle-GEN work-PL.NOM mountainous.terrain-INESS

wӡldaj *žən-dӡr* *kӡn-ӡn* *wəd-əštə*
particularly difficult-COMPAR do-NMLZ be-PST.3PL

'It was more difficult to farm the land and to ranch the cattle in particularly in the mountainous terrain' (ONC. Max dug, no. 8, 2004, p. 91).

It is worth mentioning that, used outside FDC with some verbal forms in *-ӡn*, *ӡnson* and *žən* can form set expressions, e.g., *ӡnson-ӡmbar-ӡn* 'clear, understandable' (easy-understand-NMLZ), *žən-bəχš-ӡn* 'something difficult to endure' (difficult-endure-NMLZ). However, in FDC, *ӡnson* or *žən* do not form a compound with the verbal form in *-ӡn* as they maintain the attributive functions (can be marked by the comparative suffix *-dӡr*).

4.3 Auxiliary

The auxiliary *wӡvən* 'to be' can be used in FDC in all tenses and moods, including the imperative; see example (15). In third person singular of the present indicative, the form *u* is used – see example (13); note that in some other modal constructions the existential copula *i* / *iš* / *j*[6] is used in this case; see section 9. The auxiliary *wӡvən* 'to be' can be used with verbal prefixes; see example (15).

(15) *nӡ* *foš* *žəmӡdž-ə* *ӡnson* *dar-ӡn,* *ӡnson*
POSS.1PL cattle winter-INESS easy keep-NMLZ easy

χӡšš-ӡn *fӡ-wӡnt!*
keep-NMLZ PREF-be.IMP.3PL

'Let our cattle ranch easily during the winter time!'
(ONC. Ajlarov I., Gadžinova R., Kcoeva R. Poslovicy. Vladikavkaz, 2005, p. 606).

In FDC, besides the verb *wӡvən* 'to be', the habitual verb *vӡjjən* 'to be' can be used, which has only third person singular and third person plural forms of the present indicative; cf.:

---

6 The allomorph *j* is used with negation.

(16) *finnag   gorʒt Kuopio-jə   universitet-ə   psiχolog-tʒ*
    Finnish town Kuopio-GEN university-GEN psychologist-PL.NOM

    kʷəd š-bɜrɜg   kod-t-oj,   aftɜmɜj, dam, pessimist-t-ɜn   nɜ,
    how   PREF-known do-TR-PST.3PL so   CIT   pessimist-PL-DAT NEG

    fɜlɜ optimist-t-ɜn   žən-dɜr   sɜr-ɜn   vɜjj-ə
    but   optimist-PL-DAT difficult-COMPAR live-NMLZ be.HAB-PRS.3SG

    'The psychologists of the Kuopio university found out that the life is more
    difficult not to the pessimists but to the optimists'
    (ONC. Max dug, no. 8, 2002, p. 163).

Negation particles always attach to the auxiliary in FDC; cf.:

(17) *…wəj ɜnson ra-χat-ɜn   nɜ   wəzɜn*
    that   easy   PREF-understand-NMLZ   NEG be.FUT.3SG
    'It will be not easy to understand it'
    (ONC. Bicoev G. Kh. Večernjaja zvezda. Vladikavkaz, 2003, p. 156).

The auxiliary can be either postposed (see example [17]) or preposed (see
example [18]) to the noun phrase <*ɜnson* 'easy' or *žən* 'difficult' + the verbal
form in *-ɜn*>. In ONC, a preposition of the auxiliary is rarer than its postposition.

(18) *fɜlɜ ajnɜdž-ə šɜr-t-ə   mary-ɜn **u**   ɜnson a-tɜχ-ɜn,   adɜm,*
    but   rock-GEN top-PL-INESS bird-DAT be.PRS.3SG easy PREF-fly-NMLZ man

    gal-tɜ   ɜmɜ wɜrdɜ-ttɜ   sə   kɜn-oj   tɜχ-ən nɜ žon-əns!..
    ox-PL.NOM and cart-PL.NOM what do-CONJ.3PL fly-INF NEG know-PRS.3PL

    'But for a bird it is easy to fly (lit. 'a bird easily flies = a bird can fly easily')
    across a rock and what should men, oxen and carts do, they cannot fly!'
    (ONC. Džusojty N. G. Sljozy Syrdona. Vladikavkaz, 2004, p. 472).

The auxiliary can be separated from the noun phrase with the verbal form in
-ɜn; cf. the following examples in which the citative clitic and the interrogative
adverb are placed between the verbal derivate in *-ɜn* and the auxiliary:

(19) *žən   ɜmbar-ɜn,   **dam**, štə,   Nafi-jə fəšt-ət-aw   wɜžžau…*
    difficult understand-NMLZ CIT   be.PRS.3PL Nafi-GEN letter-PL-EQU hard
    'They say, it is difficult to understand them as the Nafi's letters, it's hard …'.
    (ONC. Max dug, no. 9, 2003, p. 21).

(20) *kɜd χʷəsaw iš wɜd rɜštzinad aftɜ žən   ar-ɜn   **sɜmɜn** u*
     if God   EXT than truth   so difficult find-NMLZ why   be.PRS.3SG
     'If God exists, why is it so difficult to find the truth'
     (ONC. Aγnatə Gɜštɜn. Temərə kɜštɜr čəžg. Vladikavkaz, 2013).

Also see example (30) where the subordinator stands between the combination of *ɜnson* + the verbal derivate in *-ɜn* and the auxiliary.

5 Voice and FDC

In the previous section, it has been mentioned that in FDC, the auxiliary agrees with a Patient-like participant, while an Agent-like participant is marked by the dative. One can note that FDC is close to passive constructions. Ossetic has a number of passive and modal passive constructions. In the paper, I will mention only the standard passive construction (for details, see Vydrin 2011). The standard passive construction is formed by the past participle of a lexical verb and one of the auxiliaries, *sɜwən* 'to go' or *wɜvən* 'to be' (or the habitual *vɜjjən*), which agrees with Patient in person and number; Agent is marked by the ablative; cf. the active construction (see example [21]) and the standard passive construction (see example [22]):

(21) *kušdžə-tɜ   χɜzar   arɜžt-oj*
     worker-PL.NOM house   build-PST.TR.3PL
     'The workers have built the house'.

(22) *χɜzar arɜžt   u   kuššə-t-ə   amal-ɜj*
     house built.PART.PST be.PRS.3SG worker-PL-GEN means-ABL
     'The house has been built by the workers (lit. 'by the powers of workers').

Besides different verbal forms (past participle vs. the form in *-ɜn*), the standard passive construction differs from FDC in the following:
(a) case marking of Agent-like participant (the dative in FDC and the ablative in the passive construction);
(b) transitivity of the lexical verb (FDC is used both with transitive and intransitive verbs, while the passive construction can be formed only from transitive verbs);
(c) semantics (the standard passive construction does not convey any modal meaning, while FDC does).

6 FDC and Ossetic complex predicates

As with all other modern Iranian, in Ossetic, the bulk of the verbal lexicon is formed by the so-called complex verbs – predicates consisting of an N-constituent (a noun, an adjective or a verbal derivate) and a V-constituent (a finite verb); N is always preposed to V; for example, *š-aχʷər kɜn-ən* (PREF-study do-INF) 'to teach', 'to study'.

The detailed discussion of morphological and syntactic peculiarities of the Ossetic complex verbs goes far from the topic of this article (cf. Grashchenkov 2010; Vydrin 2014: 43–48). According to my understanding of the Ossetic complex verbs, their key features are as follows: the constituents cannot be separated from each other by other words or moved to another position; N cannot have most of the nominal flexion; the verbal negation particles cannot attach to the verb and are placed in front of the whole complex predicate; verbal prefixes are attached only to N.

If we compare the peculiarities of complex verbs with the syntactic features of FDC, one can note that the auxiliary and the combination of 'easy'/'difficult' + a verbal derivate in *-ɜn* cannot be considered a complex predicate. The auxiliary in FDC can be separated from the adjective + a derivate in *-ɜn* combination; it can also be used in front of it; the negative markers are attached only to the auxiliary while the verbal prefixes can attach either to the auxiliary or to the verbal derivate in *-ɜn* (or to the nominal part of the verbal derivate in *-ɜn* in case of a complex verb) (see section 4.3).

7 The use of FDC

I have checked the compatibility of FDC with other voice, valence-increasing and modal constructions. FDC can be used only in the causative construction. The causative construction is formed by the auxiliary *kɜnən* 'to do' used together with the infinitive of the lexical verb; Causee is marked by the nominative / genitive[7] (if the verb is intransitive) or by the dative (if the verb is transitive). For example:

(23) *mad   jɜ   čəžg-ɜn   š-kɜn-ən   kod-t-a   nog kʼaba*
     mother POSS.3SG daughter-DAT PREF-do-INF do-TR-PST.3SG new dress
     'Mother forced her daughter to put on a new dress' (Bagaev 1965: 340).

---

7 Nominative can be used only with inanimate Causee.

Compare the following example of FDC used in the causative construction:

(24) *zonəy*   *jз*   *mid-bənat-ə*   *a-šald*   *зmз,*
     sledge POSS.3SG inside-place-INESS PREF-freeze.PST.3SG and

     зvзccзgзn, *wəsə* *зnз-bon*   *song-зn* *žən*
     probably   that   without-strength hand-DAT difficult

     fe-nkʔʷəš-ən-gзn-зn   *wəd*
     PREF-move-INF-do-NMLZ be.PST.3SG

     'The sledge got frozen to the earth and apparently for his/her weak
     hands it was difficult to move it'
     (ONC. Beštaev G. G. Proizvedenija. 3 vols. Vladikavkaz, 2004, p. 449).

FDC can be used both in assertive and interrogative sentences. The examples of FDC used in assertive sentences have been given above. The following examples show its use in interrogative sentences. It is worth mentioning that it is only the verbal form in *-зn* that cannot function as the question focus (cf. possible translations of example [25]).

(25) *wədon* *žən*   *nəv* *kзn-зn* *štə?*
     they   difficult paint do-NMLZ be.PRS.3PL
     'Is it difficult to paint them?' (Is it difficult to paint THEM or smb else?
     Is it DIFFICULT or easy to paint them? *Is it difficult to PAINT them
     or to find them?).

(26) *wəmзn*   *wədon* *žən*   *š-nəv*   *kзn-зn*   *štə?*
     that.DAT they   difficult PREF-paint do-NMLZ be.PRS.3PL
     'Is it difficult for him to paint them?'

The construction can be used in contrastive sentences, for example:

(27) *asə* *lзppu* *žən*   *nəv* *kзn-зn* *u*   *wəsə* *lзppu*
     this boy   difficult paint do-NMLZ be.PRS.3SG that boy

     ta *nз-w*
     CONTR NEG-be.PRS.3SG
     'It is easy to paint this boy, but not that one'.

(28) *зnз*   *dзw*   *mən*   *žən*   *sзr-зn*   *u,*
     without you.SG.GEN 1SG.ENCL.DAT difficult live-NMLZ be.PRS.3SG

     demз   *ta*   *зnson*
     you.SG.COM CONTR easy
     'It is difficult for me to live without you, but with you, it is easy'.

In FDC, any of the constituents can be omitted, i.e., Patient (example [10]), Agent (example [27]) or Sole participant (example [13]), the auxiliary (example [28]) and the verbal form in -ɜn (examples [27] and [28]). The construction can be used both in the main and in the subordinate clauses; cf. the following example in which FDC is used in a purpose clause:

(29) *sɜmɜj wəj ɜnson-dɜr   ra-mbul-ɜn   wa   wəj*
      for   that easy-COMPAR PREF-win-NMLZ be.CONJ.3SG that

      təχχɜj qɜw-ə   tɜng birɜ arχaj-ən
      POST   need-PRS.3SG very many work-INF
      'One should train a lot to best him easily'.

8 Semantics

Semantically, FDC is used to convey the meaning 'easy or difficult to accomplish'. Sometimes FDC can have a non-epistemic possibility meaning; cf. example (18). When formed from a transitive predicate, FDC conveys the properties of the Patient-like participant, which cause difficulties for the accomplishment of the situation (difficilitive meaning) or make the accomplishment of the situation easy or possible (facilitive meaning). When formed from an intransitive predicate, the construction expresses the situation, which is easy or difficult to accomplish for the Sole participant of the situation. An interesting point is that FDC's Agent-like participant can be either animate or inanimate; cf. the example in which it is inanimate:

(30) *ɜmɜ ištə   nisɜjag   ɜfχɜr-ɜn   nəχaš-ɜn*
      and   something worthless insult-NMLZ word-DAT

      dɜr   ɜnson fe-žnɜt-gɜn-ɜn   kʷə   wa
      FOC   easy   PREF-fury-do-NMLZ if   be.CONJ.3SG
      '(He was afraid that) some worthless, insulting words will easily throw him into a rage' (lit. 'he will be easily thrown into a rage by worthless, insulting words')
      (ONC. Gusalov B. M. I vozdastsja každomu. Vladikavkaz, 2003, p. 102).

FDC can express both agentive and non-agentive situations; cf. the following examples with the non-agentive verbs *roχ kɜnən* 'to forget' and *ulɜfən* 'to breathe':

(31) *šəydзg*   *žзrdз-jə*   *зnkʔar-зn-tз*   *aftз* *зnson* *roχ-gзn-зn*
      pure   heart-GEN   feel-NMLZ-PL.NOM   so   easy   forgotten-do-NMLZ

      ne   *štə*
      NEG   be.PRS.3PL

      'The feelings arisen from a pure heart are not so easy to forget'
      (ONC. Max dug, no. 9, 2003, p. 79).

(32) *ra-jqal*   *dзn,*   *žən*   *ulзf-зn*   *mən*   *kзj*
      PREF-awake   be.PRS.1SG   difficult breath-NMLZ   1SG.ENCL.DAT that

      wəd   *wəmз*   *gзšgз*
      be.PST.3SG   that.ALL   POST

      'I woke up because it was difficult to breathe'
      (ONC. Max dug, no. 4, 2001, p. 131).

9 Origin

There is no diachronic evidence of the FDC's existence before the nineteenth century as Ossetic was an unwritten language till the middle of nineteenth century[8] and there is very little data about the Alanian (the ancestor of the modern Ossetic) language. However, apparently the origin of FDC is connected to the construction of participant-external possibility[9]. Both constructions consist of the verbal derivate in *-зn* and the auxiliary. Agent-like participant is marked by the dative in both constructions. Also note that the verbal derivate in *-зn* is used only in these two modal constructions; cf. the following examples of the participant-external possibility construction formed from an intransitive (example [33]) and a transitive (example [34]) verb.

(33) *Baratašvili-jə*   *qoməšdžən*   *kurdiat-зn*   *passivon* *romantizm-ə*
      Baratašvili-GEN   powerful   talent-DAT   passive   romanticism-GEN

      kʔʷəndзg   *fзlgзt-t-ə*   *ba-sзw-зn*   *nз*   *wəd*
      narrow   frame-PL-INESS   PREF-go-NMLZ   NEG   be.PST.3SG

      'Powerful talent of Baratašvili could not fit in the narrow frames of romanticism'
      (ONC. Beštaev G. G. Proizvedenija. 3 vols. Vladikavkaz, 2004, p. 226).

---

8 The first book in Ossetic was published in 1798 using the Church Slavonic alphabet (translation of the Catechism). Since then there were several unsuccessful attempts to develop an Ossetic alphabet based on the Georgian alphabet (Khutsuri script). Only in the middle of the nineteenth century was the Ossetic alphabet based on Cyrillic letters developed.

9 Participant-external possibility is understood as one of the main meanings of non-epistemic possibility, which "refers to circumstances that are external to the participant, if any, engaged in the state of affairs" and that make this state of affairs possible (e.g., *To get to the station, you can take bus 66*) (van der Auwera and Plungian 1998: 80).

(34) *адзм-зп   адзм   маг-зп   пз-j*
     man-DAT   man   kill-NMLZ   NEG-EXT
     'A human being is not allowed to kill a human being'.

Semantically, the two constructions are also close to each other: FDC can convey non-epistemic possibility; cf. example (18).

However, the constructions differ greatly in their morpho-syntax. Unlike FDC, where the auxiliary agrees with a Patient-like participant, the auxiliary in the participant-external possibility construction is always in third person singular (in the present indicative the existential copula *i* [*iš*] is used). One can hardly imagine the direct evolution of FDC from the participant-external possibility construction. The semantic map of modality (van der Auwera and Plungian 1998) also does not offer the grammaticalization path <participant-external possibility → "difficult / easy to accomplish">.

It is not clear why FDC uses only two adjectives and does not allow other VALUE-adjectives (e.g., 'pleasant', 'long' (time), 'terrible', etc.). There are two logical options of FDC's evolution. Either FDC is an old construction and used to be compatible with an open list of VALUE adjectives, and then the list shortened to only the two adjectives; or FDC appeared in Ossetic recently on the basis of the two adjectives, and later the use of other VALUE adjectives may become possible. Due to the lack of data on the Ossetic language before the nineteenth century, it is impossible to prove or disprove any of the named logical options.

10 Facilitive/difficilitive in other Caucasian and Iranian languages

According to available grammatical studies, besides Ossetic, no other Iranian languages are reported to have a dedicated marker or a dedicated construction to convey facilitive or difficilitive semantics. Among the Caucasian languages geographically close to Ossetic, only Adyghe (North West Caucasian) grammaticalizes the meaning 'easy / difficult to accomplish'. In Adyghe, the suffix *-ʁʷeṣ̂ə* conveys facilitive meaning and the suffix *-ʁʷaje* expresses difficilitive semantics (Rogava and Keraševa 1966: 297–298); cf.:

(35) *ar   bʷew   ṣ̂e-ʁʷeṣ̂ə   aš   ɣe-pλe-ž'ə-ʁʷaj-ep*
     3SG.ABS   very   do-FCL   3SG.ERG   LOC-look.at-REFACTIVE-DFC-NEG
     'This is easy to do, this is not difficult to look at again'
     (Rogava and Keraševa 1966: 298).

Adyghe facilitive and difficilitive suffixes originate in the combinations of the roots *ʁʷe* 'time' + *ŝʷə* 'good' (facilitive) and *ʁʷe* 'time' + *je* 'evil' (difficilitive) (Rogava and Keraševa 1966: 297).

Note that in Kabardian, which is a close relative of Adyghe and which, unlike Adyghe, has been in close contact with Ossetic for a long time, facilitive and difficilitive suffixes are not attested (e.g., Abitov and Balkarov 1957). It is reasonable to assume that Ossetic was not under the areal influence and developed FDC independently.

A separate study is needed to find out where else, beyond the Caucasian region, the grammaticalization of the meaning 'easy / difficult to accomplish' is possible. However, it seems that Ossetic FDC is a typological rare phenomenon.

11 Conclusion

In this article, I have examined the peculiarities of the dedicated construction (facilitive-difficilitive construction or FDC) conveying the meanings 'difficult to accomplish' and 'easy to accomplish' in Ossetic. This construction is not attested in standard Ossetic grammars.

FDC resembles the Tough Construction, which is attested in well-studied European languages; cf. English *John is tough (for Mary) to please; girls are tough to please* (among many others see, e.g., Comrie and Matthews 1990; among recent studies see, for example, Hicks 2009), where the auxiliary agrees with a Patient-like participant, an Agent-like participant is marked by the preposition, which is usually used for Beneficiary. However, unlike FDC, Tough Construction is not restricted to the use of adjectives with the meanings 'easy' and 'difficult' and allows the use of almost all VALUE adjectives, cf. *This room is pleasant to sleep in*.

The narrow semantics of FDC determines its specificity. However, further research is needed to find out where else the facilitive and difficilitive meanings can be grammaticalized to a dedicated marker or construction.

12 Abbreviations

ABL – ablative; ABS – absolutive; ALL – allative; CIT – citative; COM – comitative; COMPAR – comparative; CONJ – conjunctive; CONTR – contrastive; DAT – dative; DFC – difficilitive; ENCL – enclitic; EQU – equative; ERG – ergative; EXT – existential; FCL – facilitive; FDC – facilitive-difficilitive construction; FOC – focus;

FUT − future; GEN − genitive; HAB − habitual; IMP − imperative; INESS − inessive; INF − infinitive; ITER − iterative; LOC − locative; NEG − negation; NMLZ − nominalization; NOM − nominative; NP − noun phrase; ONC − Ossetic National Corpus; PL − plural; POSS − possessive; POST − postposition; PREF − prefix; PRS − present; PRTCL − particle; PST − past; RFL − reflexive; SG − single; TR − transitive.

References

Abaev, Vasilij I. 1958. *Istoriko-etimologičeskij slovar' osetinskogo jazyka*. T. 1. Moscow and Leningrad: Izdatel'stvo akademii nauk SSSR. (Etymology dictionary of the Ossetic language. Vol. 1. In Russian.)

Abaev, Vasilij I. 1959. *Grammatičeskij očerk osetinskogo jazyka*. Ordžonikidze: Severo-Osetinskoje knižnoje izdatel'stvo. (A grammatical sketch of Ossetic. In Russian.)

Abaev, Vasilij I. 1964. *A grammatical sketch of Ossetic.* The Hague: Mouton.

Abaev, Vasilij I. 1989. *Istoriko-etimologičeskij slovar' osetinskogo jazyka*. T. 4. Leningrad: Nauka. (Etymology dictionary of Ossetic language. Vol. 4. In Russian.)

Abitov, Muhob L. & B. H. Balkarov. 1957. *Grammatika kabardino-čerkesskogo literaturnogo jazyka*. Moscow: Izdatel'stvo akademii nauk SSSR. (A grammar of the literal Kabardian-Cherkess language. In Russian.)

Ahvlediani, Georgij S. (red.). 1969. *Grammatika osetinskogo jazyka*. T. 2. Sintaksis. Ordžonikidze: Naučno-issledovatel'skij institut pri sovete ministrov Severo-Osetinskoj ASSR. (A grammar of the Ossetic language. Vol. 2. Syntax. In Russian.)

Bagaev, Nikolaj K. 1965. *Sovremennyj osetinskij jazyk. Čast' 1 (fonetika i morfologija)*. Ordžonikidze: Severo-Osetinskoje knižnoje izdatel'stvo. (Modern Ossetic Language. Part 1. In Russian.)

Comrie, Bernard & Stephen Matthews. 1990. Prolegomena to a typology of Tough Movement. In William Croft, Keith Denning & Suzanne Kemmer (eds.), *Studies in typology and diachrony*, 44–58. Papers presented to Joseph H. Greenberg on his seventy-fifth birthday. Amsterdam and Philadelphia: John Benjamins.

Dzahova, Veronika T. 2009. *Fonetičeskie harakteristiki fonologičeskoj sistemy sovremennogo osetinskogo (ironskogo) literaturnogo jazyka*. Monografija. Vladikavkaz: Izdatel'stvo SOGPI. (Phonetic description of phonological system of modern standard [Iron] Ossetic. In Russian.)

Grashchenkov, Pavel V. 2010. Složnye predikaty v osetinskom jazyke. *Bjulleten' Obščestva vostokovedov RAN*. Vyp. 17 [Special issue: Trudy mežinstitutskoj naučnoj konferencii "Vostokovednye čtenija 2008": Moskva, 8-10 oktjabrja 2008 g.] Moscow: Učreždenie Rossijskoj akademii nauk "Institut vostokovedenija RAN", pp. 115–130 (Complex predicates in Ossetian. In Russian.)

Hicks, Glyn. 2009. Tough constructions and their derivation. *Linguistic Inquiry* 40 (4). 535–566.

Kibrik, Aleksandr E. 1997. Beyond subject and object: Toward a comprehensive relational typology. *Linguistic Typology* 1 (3). 279–346.

Rogava, Georgi V. & Zejnab I. Keraševa. 1966. *Grammatika adygejskogo jazyka*. Krasnodar and Majkop: Knižnoe izdatel'stvo Krasnodar. (A grammar of the Adyghe language. In Russian.)

Takazov, Harum. A. 1992. *Kategorija glagola v sovremennom osetinskom jazyke*. Moscow: Institute of Linguistics of the Russian Academy of Sciences, doctoral dissertation. (Verb in the Ossetic language. Unpublished doctoral dissertation. In Russian.)

Tehov, Fidar D. 1970. *Vyraženije modal'nosti v osetinskom jazyke*. Tbilisi: Mecniereba. (Modality in the Ossetic language. In Russian.)

Thordarson, Fridrik. 2009. *Ossetic grammatical studies*. (Sitszungsberichte der philosophisch-historischen Klasse, Bd. 788, Veröffentlichungin zur Iranistik. Nr. 48). Wien: Verlag der Österreichischen Akademie der Wissenschaften.

van der Auwera, Johan & Vladimir Plungian. 1998. Modality's semantic map. *Linguistic Typology* 2: 79–124.

Vydrin, Arsenij P. 2011. *Sistema modal'nosti osetinskogo jazyka v sopostavitel'nom osveščenii*. Saint Petersburg: Institute for Linguistic Studies of the Russian Academy of Sciences, unpublished dissertation. http://www.ossetic-studies.org/vydrin/index.php/publications/. (Modal system of Ossetic. In Russian.)

Vydrin, Arsenij P. 2014. Glagol v osetinskom jazyke. *Vostokovedenie. Istoriko-filologičeskie issledovanija. Mežvuzovskij sbornik statej*. Vyp. 30 (zaklučitel'nyj). Pamjati akad. M. N. Bogoljubova. Saint Petersburg, pp. 25–80. (Ossetic verb. In Russian.)

Z. A. Yusupova
15 Possessive construction in Kurdish

Abstract: The article contains an analysis of the possessive construction in Kurdish. This construction, employed with structural and semantic variations in all Kurdish dialects, is most widespread in southern Kurdish dialects, where it is formed with enclitic personal pronouns. The northern dialects, which don't have enclitic personal pronouns, use the oblique case forms of personal and sometimes reflexive pronouns for the construction. For the study of the southern Kurdish dialects, I use the *divans* (poetic ontologies) of eighteenth- and nineteenth-centuries' Kurdish poets published in Iraqi Kurdistan, while the study of the northern dialects is based on the folklore texts recorded by O. Mann and I. Zuckermann. Standard romanized transcription for modern linguistic Kurdish studies are used.

Keywords: Kurdish studies, Kurdish linguistics, Kurdish dialects, Iranian languages

Table 1: Symbols used in this article and corresponding IPA symbols:

Symbol used in this article	IPA symbols
C,c	[dʒ]
Ç,ç	[tʃ]
E,e	[æ]
Ê,ê	[e:]
ẍ	[ɣ]
Î,î	[i:]
J,j	[ʒ]
L,l	[l]
ḹ	Velar [l]
R,r	[r]
R̄,ř	[r] (rolled front [r])
Ş,ş	[ʃ]
Û,û	[u:]
Y,y	[j]
'	[ʔ]

As in other Iranian languages, in Kurdish the so-called possessive construction is widely used, its grammatical meaning still not adequately understood. Kurdo-logical studies have actually described certain variants of that construction

Z. A. Yusupova, Institute of Oriental Manuscripts of the Russian Academy Sciences

on the basis of separate Kurdish dialects; their authors perceived it either as a possessive construction as such or as a sentence with the thematically emphasized subject.[1]

The newly reviewed sources, including texts of literary compositions in both southern Kurdish dialects (Gorani, Avramani, Sorani) and northern ones (Kurmanji, Zaza), attest to the fact that the possessive construction varying across these dialects may have quite a few meanings. As follows from the reviewed materials, besides expressing the idea of possessing per se, the construction was widely used to characterize the subject of a sentence, its condition, or a feature it might have. This article analyzes possessive constructions denoting the subject's condition.

In southern Kurdish dialects, the possessive construction is formed with enclitic personal pronouns (first, second, and third person singular and plural). In such cases, the subject is frequently expressed repeatedly: with a noun (or a personal pronoun) usually opening the sentence, as well as an enclitic pronoun, the position of the latter depending on the structure of the sentence. At that, the verb acquires the form of third person singular. The available texts allow for distinguishing the following structural and semantic variants of the analyzed construction.

1 The most common version having a copula verb

Model 1, the most frequent one, where the subject is expressed twice, once with a name or a personal pronoun, and then an enclitic pronoun duplicating it and having a corresponding form. Its position may vary:

It may be merged with the object[2] to express genitive relation (the most common case):

(1)   DiĪ naĪan-iş-en diĪ naĪan-iş-en (DS, 107)[3]
      [The heart is moaning, the heart is moaning
      (lit. 'the heart – its moaning // it has');]
      Nergis to dîde-t nabîna û kor-en (DS, 175)
      [(Oh), narcissus, your blind eyes cannot see;]
      Ce guĪzar buĪbuĪan axêz-şan-en (DS, 69)
      [Nightingales are awakening in the garden;]

---

1 See Tzukerman (1965: 161–165, 1986: 184–187); Eyubi and Smirnova (1968: 138–140); Yusupova (1985: 112–114, 1998: 121–125).

2 Here and below a more precise term would be "object of the state of being".

3 Between the parentheses are the universally accepted abbreviations denoting the quoted sources and their page numbers.

It may be merged with an indirect object (rarely):

(2)   Dirext û dar ce paîz-şan zwîrî-n (DS, 68)
      [(All) forests and trees are offended by the fall;]
      Ya 'Ewdaĺ Seîdî hawar le to-ş-en (DS, 62)
      [(Oh), mountain Awdal, Saidi appeals to you for help;]

It may be merged with the instrumental-directive preposition *pê*:

(3)   Min çunke hîcran[î] mecnûnî pê-m-en (DR, 127)
      [As I feel sad to be away, (like) Majnun (did). . . ;]
      Min ce to fêşter meyl[î] to pê-m-en (XQ, 743)
      [I love you more than you (love me).]

There is one case involving the locative preposition *la*:

(4)   Minîç kemê fam[î] Eyaz la-m-en (DM, 156)
      [I have some of Ayaz's wisdom as well.]

Model 2, where the subject is expressed solely with the enclitic pronoun merged with the object or, less frequently, with a preposition:

(5)   Arezûy dîdar[î] dîn[î] to-şan-en (XQ, 774)
      [They have the desire to see you;]
      Ce dûrî[y] yaran bêqerarî-m-en (DS, 107)
      [Due to my being away from my beloved, I am worried;]
      Pê-m weş-en gêcaw[î] deryay xem (DM, 360)
      [I like to whirl in the maelstrom of sadness.]

Model 3, where the enclitic pronoun duplicates the indirect prepositional object expressed by the words denoting parts of the subject (soul, body, heart, eyes, etc.). Its position may be the following:
      It may be merged with the preposition:

(6)   Dîdey piř esrim xaw ce la-ş dûr-en (DS, 62)
      [My eyes full of tears are sleepless (lit. 'sleep is far from them');]
      Xatirim yend xar[î] meynet tê-ş-da . . . (DM, 262)
      [My soul has so many thorns of grief . . . ;]

It may precede the preposition:

(7)   Ten zam[î] xedeng[î] mujey to-ş pê-we-n
      [My body has wounds caused by the shots of your eyelashes.]

Model 4, where the enclitic pronoun not only duplicates the indirect object, but also the subject:[4]

(8)   Wer to bînaî-t perde-ş ha ne ser … (DM, 72)
      [In case (there is) a cloth (of ignorance) over your eyes …]

In this last example, like in the one with the preposition *tê* (see above), the copula has been omitted.

In the negative construction having a copula verb, the enclitic duplicating the subject may occupy the following positions:

It may be merged with the negative form of the copula:

(9)   Çe ko te bijnewu deng-it nîye-m goş (DS, 38)
      [How can I listen to you, (if) I do not hear you
      (lit. 'if your voice is away from my ears');]
      Çi ser zemîne nîyen-iş qerar (DS, 132)
      [There is no rest for him on this Earth:]

It may be merged with the object or with the word defining it:

(10)   Xeşmit 'ezîmen kerem bêsaman
       Kes tewanay qehr[î] to-ş nîyen (DS, 222)
       [Your rage is great, your generosity, limitless,
       (But) nobody has the strength to be angry with you;]
       Hîç kes ne hîç ca sebûrî-ş nîyen (DR, 67)
       [Nobody has the patience nowhere.]

It is of interest to compare these cases with those where the subject of a negative construction is solely expressed with an enclitic pronoun:

(11)   Ta ser çenî kes nîyen wefa-şan (DS, 85)
       [They are never totally faithful to anybody;]
       Ew dîdey mest-iş coyay xew nîyen (DS, 100)
       [Her intoxicated eyes do not know sleep (lit. 'do not seek sleep').]

---

4 Here and below a more precise term would be "subject of the state of being".

The texts also contain examples of sentences containing copulas, in which the negation is expressed with the particle *ne* preceding the object while the copula remains positive:

(12)   Seîdî tasey sext[î] dûrî[y] yar-iş-en
        Ne aram, ne sebir, ne qerar-iş-en (DS, 61)
        [Saidi suffers terribly (due to) his being away from his beloved one,
        He has neither peace, nor patience, nor rest;]
        Seîdî ne r̂o xurd, ne şew xaw-iş-en (DS, 101)
        [Saidi neither eats during the day, nor sleeps during the night
        (lit. 'neither food at daytime, nor sleep at night he has').]

2 Constructions with various tenses of the verb *bî* 'to be'

Model 1, where the subject is duplicated by an enclitic pronoun in the following positions:
    It may be merged with the object:

(13)   Toîç însaf-it bo henî wese wes … (DM, 26)
        [You, be just, enough tormenting [me], enough!;]
        Ez Uromon mekan-im bê, wulatim (DS, 13)
        [Avraman was my refuge, my motherland;]
        Eger to pîr û pîr-ê-t bo tikakar … (DS, 54)
        [If you are a pir, and you [yourself] require a pir … ;]

It may be merged with the verb:

(14)   To bo-t ne dil hof[î] xudat (DS, 97)
        [May in your heart there be fear (in front) of your God;]
        Dirêẍ min bîa-m dû hezar dîde … (DS, 55)
        [If I only had two thousand eyes!;]
        Cew duma her wext Qeys bê-ş arezû … (DS, 95)
        [After that, every time when Qays developed a desire …]

It may be merged with the preposition:

(15)   Er fam pê-ş bîa diley mecnûn xo
        Biryaş temay lal[î] şîrîn r̂az[î] to (DS, 63)
        [If my mad heart were conscientious,
        It would reject the clandestine desire of your sweet lips.]

Model 2, where the subject is solely expressed through an enclitic pronoun merged with the object:

(16)   Qiblem, ta seher guft-u-go-şan bê,
       Daway hamşanî zuĪfî to-şan bê (DS, 34)
       [Darling, till the (very) morning they chatted, (they)
       Tried to compete with the smell of your locks;]
       BeĪkim xeĪasî-m ne des dûrî-m bo (DM, 45)
       [Maybe, (then) I will be saved from separation.

The negative construction having the verb *bî* standing in preterit can appear in two shapes:
       With the verb having a negative form:

(17)   Axir dawa-şan tan û po-ş nebê,
       Hîç kam boy 'etrî zuĪfî to-ş nebê (DS, 35)
       [Finally, (they understood) that their dispute was senseless,
       (As) none (of them) had the smell of your locks;]
       Seîdî te hîç-it çar nebê, ta ser çenît kes yar nebê (DS, 25)
       [Saidi, you had no other way out,
       (as) nobody remained your friend till the end;]

With the verb having a positive form while the negation gets expressed with the particle *ne* 'neither-nor'

(18)   Ne be şew werar, ne be r̂o xaw-it bê (DS, 86)
       [You had neither peace at night, nor sleep during daytime.]

3 Constructions having the verb *hen* 'to have'

Model 1, where the subject is denoted twice, with a name (or a pronoun) and an enclitic pronoun:

(19)   Aşiqî be diĪ denaĪê meylî giryan-î heye (DN, 584)
       [The loving one whose heart is moaning has a tendency to weep;]
       Ez yarêwe cone-m hene (DS, 26)
       [I have a beautiful sweetheart.]

Model 2, where the subject is denoted solely with an enclitic pronoun:

(20)   Hey hay, cuanim, firêm hen mirad (DM, 4)
       [Alas, my beauty, I have a lot of wishes!;]
       'Adetêk-î heye, hergîz le kesê napirsê (DN, 592)
       [She has a rule. (She) is never curious about anybody.]

The negative construction having the verb *hen*, like the one with *bî*, is formed with *nîye*(*n*) or with the negative particle *ne* whenever the verb remains in the positive form:

(21)   Ax le geĪ ême, Hebîbe ser û peywend-î nîye (DN, 592)
       [Woe to us, Habibe wants nothing to do with us!;]
       Hêz[î] pay r̂eft-û-ama-şan nîyen (DM, 29)
       [They have no (more) strength to move.]

Another construction has been noticed in the texts, in which the negative particle is accompanied by the particle *niho*, also expressing negation:

(22)   Ne fikr-iş hen ce law muĪkî, ne maĪî,
       Niho bak-iş ce ferzend û 'eyaĪa (DS, 151)
       [He has thoughts neither about his property nor his house,
       He has no fear about his kids or his wife.]

Thus, the primary purpose of the reviewed versions of the possessive constructions in southern Kurdish dialects looks like reporting the features of the subject: indication of its physical or psychical condition, or else indicating its temporary or permanent characteristics. With regard to its structure, this construction follows the possessive model proper ("I a son I have"), also having two basic versions: (1) the subject can be expressed twice, once with a noun (or a pronoun) in the direct case (nominative), and once with a personal enclitic pronoun duplicating it; (2) the subject can be expressed solely with an enclitic pronoun. In other words, the main elements of the possessive construction are the enclitic pronouns actually forming an indirect construction grammatically opposing its direct counterpart; cf.: Êmeyç naĪe-man xo bêkar nîyen (DS, 157) [However, our / of ourselves moaning also brings about a result] versus Axir nek êmeyç dax ne diĪ-ênmê (DM, 410) [However, our hearts are sad as well (lit. 'we are with sadness in our hearts')]. In the first case, with the subject expressed with a personal pronoun first person singular, the verbal copula has the form of the third person singular, i.e., it corresponds with the object of the state of

being, while in the second case, the copula reflects both the person and the number of the (grammatical) subject.

Northern Kurdish dialects have lost their enclitic pronouns,[5] and therefore similar constructions use indirect forms of personal pronouns in order to denote the subject of the state of being:

(23) Te dil bê xwab u bê xurd-e (MC, 335)
      [Your heart (is) without sleep or food;]
      Min dilberek wek duř heye (MC, 94)
      [I have a sweetheart (who is) like a pearl . . . ;]
      Me ji zulfên botan řeştir-e bext (MC, 320)
      [My happiness is darker than the locks of Botan's (beauties).]

Compare this also with the negative construction having the verb "to have":

(24) Bi xudê qet te di dilda ji xudê tirs nehin (MC, 535)
      [By God, there is no fear of the Lord in your heart.]

Along with that, in Kurmanji, mostly in folklore texts, there is a construction where the subject of the state of being/possession expressed with a noun (or a pronoun) is duplicated with an indirect form of a personal pronoun or a reflexive-possessive pronoun *xe/xwe* linked with the object via an izafat:

(25) Qewaz lê nihêrî, wekî Zuhre – halêwê t'une (Tz I, 163)
      [The messenger sees that Zuhre is unwell (– she is in no mood);]
      E'slan-padşa – wextekê c'yakî wî hebûye (Tz I, 163)
      [Once upon a time, King Aslan had a mountain;]
      Mervek – kurekî wî hebûye
      [A man had a son.]

The construction having a reflexive pronoun duplicating the subject is typical for Khorasan Kurmanji:

(26) Wan xorak'î xe jî tonnewun (Tz II, 185)
      [They had no food either;]
      Min sed sal omri xe heye (Tz II, 185)
      [I am a hundred years old (lit. 'a hundred years of my own life there are').]

---

5 With the exception of the pronominal enclitic *-ê* corresponding to the indirect form of a personal pronoun third person singular. In southern dialects, this enclitic has lost its semantics.

Some occasions when reflexive pronouns duplicated the subject of the possessive construction have been found in the texts by O. Mann (1932) dealing with a Zaza subdialect (the Kor area): jû cinêy jû dostî xu bî (OM, 337) 'A woman had a friend'; jû kenay xu bî (OM, 342) 'He had a daughter'. As we see, in the second example, the subject is not lexically expressed.

The possessive construction with enclitic pronouns acting as subjects has been noticed in the texts not solely with the verbs denoting being or having. It has also been attested with certain verbs denoting action, most of them intransitive (including compound denominative ones, also expressing the condition of the subject).

4 The constructions with intransitive verbs

Perfect tense (most commonly):

(27) To bê wefayî-t ce hed berşîyen (DM, 337)
      [Your inconstancy has crossed all limits;]
      Wat ce dîdey bed min pê-m neyawan (DS, 126)
      [He said, "(It) did not happen to me due to an evil eye";]
      To çêş pê-t aman, hamderdî saĪan? (DW, 29)
      [What has happened to you (lit. 'has reached you'), old friend?]

Simple past tense (rarely):

(28) Min diĪ-im we tîr[î] to êşa, ya ew? (DM, 405)
      [Was (it) my heart which was hurt with your arrow, or his?]
      Min ce dax[î] to bey tewr pê-m ama (DS, 80)
      [This is what happened to me due to my missing you;]
      Şîřîn cey guftar derûn-iş coşa (XQ, 322)
      [These words made Shirin's soul boil;]

Present-future tense, indicative mood:

(29) Ta şem'î řûy to nûr mewaro lê-ş
      Min kogay 'umr-im aîr meşo lê-ş (DS, 63)
      [As long as the candle of your face radiates light,
      Flame will emanate from my soul (lit. 'the source of my life');]
      Pêşanî-ş pirřşing[î] nûr cê-ş mixêzo (XQ, 44)
      [Her forehead radiates bright light;]
      Mebo pok pî çilê min zêĪ u ceste-m (DS, 54)
      [With the help of this bough my heart and soul will get cleansed;]

Present-future tense, subjunctive mood:

(30)   Êmeyç ba sate bezmê-man cem bo (DM, 185)
        [Let us have some merry time;]
        To wexten we sî biyawo saĪ-it (DR, 115)
        [Soon you must be thirty years old.]

There are cases with the subject expressed solely with an enclitic pronoun:

(31)   Mêşe-m derûn, kize gîon (DS, 25)
        [(Now) my soul is in pain, (and) my heart is moaning;]
        Hoş-iş pê nemend (XQ, 619)
        [He lost consciousness (lit. 'he had no consciousness left');]
        R̄îşe-m berama (DM, 27)
        ["I have started growing a beard".]

This possessive construction with verbs of action has also been found in the "Divan" by Malae Jeziri:

(32)   Min hûn bi co têtin ji dil (MC, 266)
        [Blood is flowing from my heart like a river;]
        Me dil mabû di xeyalêda (MC, 544)
        [My heart stayed dwelling in dreams;]
        Ji agir min nefes hilbû (MC, 554)
        [The flame (of love) strangled my breath.]

As one can see, in all the listed passages, the subject is expressed with an indirect personal pronoun of first person singular. However, folklore texts may contain such possessive structure with state verbs, including not solely the subject of status, but also an indirect personal pronoun duplicating it:

(33)   Siabend û Xecê 'eşq-mih'beta wan ze'f çû bal hev (Tz I, 161)
        [Siaband and Khadje fell in love with each other;]
        Zînê – şew-ro xurê wê bû girî û hêsr (Tz I, 161)
        [Zine did nothing but cried day and night;]
        Koçer – şev deng p-ê k'et, bar kirin (Tz I, 162)
        [At night, a rumor spread among the nomads, and they left.]

This last passage is peculiar, as the indirect object duplicating the subject is expressed with the enclitic pronoun -ê.

The language of the texts in Gorani dialect has yet another version of the possessive construction; there, it may have a transitive verb in the perfect form of third person singular with the subject expressed with a first person singular personal pronoun duplicated with an enclitic pronoun in the same form. This rare construction is meant to express the physical condition of the subject caused by another condition it had in the past:

(34)   Min dûrîy azîz sakin-im senden (DC, 159)
       [Being away from my beloved one deprived me of peace;]
       Min derd[î] dûrî tewana-m berden (DM, 317)
       [The pain of separation deprived me of strength;]
       Min ye 'eşq[î] to quwet lê-m senden (XQ, 387)
       [It was my love to you which deprived me of strength;]
       Min ye 'eşq[î] to zebûn-im kerden (XQ, 243)
       [It was my love to you which made me unhappy;]
       Beḻam Me'dûm heris[î] dîde-ş coş werden (DM, 56)
       [However, tears boiled in Ma'dum's eyes.]

The following example, attested in Maulawi's "Divan", looks grammatically indicative, which proves the structural similarity between the possessive and passive constructions:

(35)   Min dan[î] zîndegî-m kenden (DM, 16)
       [The tooth of my life has been pulled out (at me/my life tooth is pulled out)]

At the same time, the present construction is formally identical with the one having an active transitive verb in the past tense with an enclitic indication of the person. Thus outside a definite context, this sentence may well mean 'I have pulled out the tooth of my life'. Despite that structural similarity, these constructions were used to express different relations between the subject and the object. While a possessive construction has a subject enclitic acting as an indirect object (= logical subject) with the verb coordinated with the object of possession or state of being, the construction having a transitive verb in the past tense, the function performed by the enclitic is totally different. There, it expresses the person and number of the grammatical subject. In other words, enclitic pronouns are actually morphologized in the second construction, even though in some dialects the morphologization process has not yet been concluded. That circumstance has its proof in such features as the occasional use of enclitic subjects, the possibility to conjugate transitive verbs like intransitive

ones, etc.[6] Nevertheless, the available material seems to attest to the fact that these constructions have common roots with their origins, making it possible to trace, via a diachronic approach, the reviewed linguistic facts.

Sources

DC: *Dîwanî Cefayî*. Beẍda [Baghdad]: Çapxaney Şefîq, 1980.

DM: *Dîwanî Mewlewî*. Beẍda [Baghdad]: Çapxaney Elnecah, 1961.

DN: *Dîwanî Nalî*. Beẍda [Baghdad]: Çapxaney Kořî Zanyarî Kurd, 1976.

DR: *Dîwanî Rencûrî*. Beẍda [Baghdad]: Çapxaney Afaq 'Erebîye, 1983

DS: *Dîwanî Seîdî*. Silêmanî: Çapxaney Kořî Zanyarî Kurd, 1971.

DW: Hewramî Usman. *Welî-Dêwane*. Beẍda [Baghdad]: Çapxaney Kořî Zanyarî Kurd, 1976.

MC: *Dîwana Melaê Cizîrî*. Beẍda [Baghdad]: Çapxaney Kořî Zanyarî Kurd, 1977.

OM: Mann, O. *Mundarten der Zâzâ, hauptsächlich aus Siwerek und Kor*. Bearbeitet von Karl Hadank. Berlin: Verlag der Preussischen Academie der Wissenschaften, 1932.

XQ: *Xanay Qubadî. Şîrîn û Xusrew*. Beẍda [Baghdad]: Çapxaney Kořî Zanyarî Kurd, 1975.

Tz I: Tzukerman (1965, 1).

Tz II: Tzukerman (1965, 2).

References

Eyubi, Karim Rahmanovitch & Irayida Anatolyevna Smirnova. 1968. *Kurdsky dialect mukri* [The Kurdish dialect of Mukri]. Leningrad: Nauka.

Tzukerman, Isaak Iosifovitch. 1965. K characteristike predlozheniya s tematicheskim podlezhashchim [On the characteristic of the sentence with a thematic subject]. *Palestinsky sbornik* 13 (76). 161–165.

Tzukerman, Isaak Iosifovitch. 1986. *Khorasancky kurmandzhi* [Khorasani Kurmanji]. Moscow: Nauka.

Yusupova, Zare Alievna. 1985. *Suleymaniysky dialect kurdskogo yazyka* [The Suleymani Kurdish dialect]. Moscow: Nauka.

Yusupova, Zare Alievna. 1998. *Kurdsky dialect gorani* [The Gorani Kurdish dialect]. St. Petersburg: Nauka.

---

6 See Yusupova (1985: 116–118).

Carina Jahani

16 To bring the distant near: On deixis in Iranian oral literature

Abstract: The purpose of this article is to study oral narratives in a number of Iranian languages with a particular focus on how the audience is brought inside the framework of the story. The oral narratives selected for this study are traditional folktales and legends in Koroshi Balochi, Sistani Balochi, Vafsi, and Gorani.

Deictic devices locate an event and its participants in time and space and cannot be fully interpreted without reference to the context. They also bring coherence to the narrative. A deictic center is a point to which the deictic element is anchored. Deixis can be absolute, i.e., place the deictic center at the location and moment of utterance, but the speaker does not necessarily need to adopt his or her own time and location as the deictic center. It is also possible to detach the deictic center completely from not only the temporal and locational setting of the speech, but also from the real world, and to place it at a time and place that never existed or will exist inside an imaginary story (deictic shift).

The four linguistic variants in this study show interesting variation when it comes to deictic shift. It is more common for spatial deixis to be shifted to the story than for tense to be anchored in the story. Koroshi Balochi, Sistani Balochi, and Vafsi present almost total spatial deictic shift, whereas in Gorani the deixis is occasionally moved outside the story. Gorani is the language that has the strongest tense anchoring inside the narrative, with almost exclusive use of the non-past tense. At the other extreme we find Sistani Balochi, which has no tense anchoring in the narrative (only past tense verb forms). Koroshi Balochi uses non-past tense for events in the story line and Vafsi changes between using non-past and past tense.

Keywords: oral literature, deixis, deictic shift, Balochi, Vafsi, Gorani

1 Introduction

The purpose of this article is to study oral narratives in a number of Iranian languages with a particular focus on how deictic devices bring these stories close to the audience in space and time, or rather perhaps how the audience is brought

Carina Jahani, Uppsala University

inside the framework of the story. The oral narratives selected for this study are traditional folktales and legends.

Nyberg (2004: 25) stresses the entertaining rather than moralizing function of oral narratives. Segal (1995: 62) finds that "narrative allows us to vicariously experience phenomena that would be too dangerous or costly to experience directly" and points out that its popularity is obvious since it is "emotionally involving, structurally appealing, and educational". Like with soap operas of the TV age, it is important to keep the audience's attention, and various linguistic and extralinguistic means have been observed in Iranian oral narratives for this very purpose.

In his discussion of genres in Persian literature, Utas (2008: 229) describes the language of both oral and written literature as "normalized, conventionalized and consciously shaped to be remembered" and thereby fundamentally different from spoken language. In this he opposes Ong (1982: 11–14), who discards the concept of oral literature. Ong believes that only written words are of a lasting character, and that "oral tradition has no such residue or deposit".

The language of oral narratives described in this study is thus, according to Utas, to be regarded as a consciously shaped literary language, albeit not a written language, and by no means an ad hoc creation on each occasion of storytelling. It must therefore be assumed that such a language has its own rules, which are rather stable and possible to describe.

This study will investigate how different Iranian languages employ deictic strategies to bring the audience into the story, whether these strategies are employed in all the linguistic varieties under study, and if the strategies are similar or vary considerably from language to language. The focus here will be to investigate deictic pronouns and determiners as well as tense use in four different Iranian varieties, namely Koroshi Balochi, Sistani Balochi, Vafsi, and Gorani. For all these varieties, there are publications with oral tales in transcription and translation into English. For three of the variants, one or more of the texts is fully glossed. There are also grammatical sketches and word lists available, which makes it possible to analyze the texts even when they are not fully glossed.

The narratives selected consist of:

- four tales in Koroshi Balochi with full glossing and English translation published by Nourzaei et al. (2015: 123–209);
- nine tales in Sistani Balochi with full glossing and translation into English published by Barjasteh Delforooz (2010: 286–325, 336–391);
- seven tales in Vafsi with English translation published by Stilo (2004: 26–29, 32–57, 104–123);
- three tales in Gorani, all with English translation and one with full glossing published by Mahmoudveysi et al. (2012: 63–103).

The selected variants are all representatives of northwestern Iranian languages, although a strict division of western Iranian languages into a northwestern and a southwestern group has lately been questioned by Paul (2003: 71) and Korn (2005: 329–330).

Koroshi Balochi is spoken by scattered populations throughout southern Iran, "from Hormozgan all the way to Khuzestan, and onto the Iranian plateau. [...] Three areas with significant concentrations of Korosh are Bandar Abbas, around Shiraz, and across the southern part of Fars Province" (Nourzaei et al. 2015: 21). Although Koroshi shares many features with Southern Balochi,[1] it must be regarded as a distinct subgroup of Balochi with its own dialect division into Northern Koroshi and Southern Koroshi (Nourzaei et al. 2015: 25). The fieldwork for the monograph in which the tales were published was carried out between 2009 and 2014 by Maryam Nourzaei in and around Shiraz in northwestern Fars Province, Iran, and represents northern Koroshi (Nourzaei et al. 2015: 17–18).

Sistani Balochi, which is a variant of Western Balochi, is spoken in Iranian Sistan as well as in adjacent parts of Afghanistan and in scattered pockets throughout northeastern Iran. It is also very similar to the Balochi dialect spoken in Turkmenistan. The fieldwork for the monograph in which the tales were published was carried out between 2000 and 2005 by Barjasteh Delforooz and comprised speakers from both the Iranian and the Afghan part of Sistan (Barjasteh Delforooz 2010: 26).

Vafsi is, according to Stilo (2004: 1), "spoken in four villages in west central Iran: Vafs, Chehreqān, Gurchān, and Fark". These villages are situated southwest of Tehran, between Saveh and Hamedan in Markazi Province, Iran. The folktales published by Stilo were collected by L. P. Elwell-Sutton in the village of Gurchān in 1958. Stilo himself conducted extensive fieldwork on Vafsi in the 1960s and 1970s. This volume was prepared between 1997 and 2000 (Stilo 2004: vii, 5–10).

Gorani, as used by Mahmoudveysi et al. (2012: 2–4), is a general term for a number of vernaculars spoken in pockets in Kermanshah Province, Iran (an area dominated by Southern Kurdish dialects), as well as in adjacent parts of Iraq. The dialect under study in this monograph, Gawraǰūyī, is one of several Gorani variants. Fieldwork for the present monograph was carried out in 2007 and 2008 by Mahmoudveysi in Gawraǰū, "a cluster of four hamlets in the Zimkān river valley" in Kermanshah Province, Iran (Mahmoudveysi et al. 2012: 1).

All the variants under study have a basic split between past and non-past (present-future) tense. They are all pro-drop and have agreement marking by means of person-marking suffixes on the verb in the non-ergative domain

---

1 For dialect divisions in Balochi, see Jahani and Korn (2009: 636–638).

(intransitive verbs and transitive verbs in the non-past tense). Vafsi and Sistani Balochi exhibit a type of differential object marking (DOM) that is common in Iranian languages (Haig 2008: 157–158; Stilo 2004: 232; Barjasteh Delforooz 2010: 286–391), but Koroshi Balochi shows interesting divergences from the normal DOM system in Iranian languages (Nourzaei et al. 2015: 35–36). In the variant of Gorani described by Mahmoudveysi et al. (2012) there is no case marking.

In Vafsi, the subject of intransitive verbs (S), the agent of transitive verbs (A) in the non-past system, and the patient (P) of transitive verbs in the past system take the direct case, whereas the A in the past system and the P in the non-past system take the oblique case. We thus have a split ergative construction in Vafsi, in line with the common tense-split ergativity found in Iranian languages (Stilo 2004: 232). In Koroshi Balochi, alignment is normally non-ergative, but the enclitic pronouns remain as agreement markers in the past tense of transitive verbs, in contrast with person-marking suffixes found on all intransitive verbs and on transitive verbs in the non-past system (Nourzaei et al. 2015: 83). The same system of two different sets of agreement markers is found in Gorani of Gawraǰū (Mahmoudveysi et al. 2012: 27–28). Sistani Balochi also has non-ergative alignment, and all verbs mainly have the same person-marking suffixes in both the non-past and the past tense (Barjasteh Delforooz 2010: 286–391).

2 Deixis

Deictic devices locate an event and its participants in time and space and cannot be fully interpreted without reference to the context. They also bring coherence to the narrative. Anderson and Keenan (1985: 259) define deictic expressions as "those linguistic elements whose interpretation in simple sentences makes essential reference to properties of the extralinguistic context of the utterance in which they occur". They (Anderson and Keenan 1985: 259) recognize only three major categories of deixis: person deixis, spatial deixis, and temporal deixis. However, Fillmore (1997: 61) adds social deixis ("the social relationships on the part of the participants in the conversation"), and crucially for this article, discourse deixis, which deals with "the choice of lexical or grammatical elements which indicate or otherwise refer to some portion or aspect of the ongoing discourse" (Fillmore 1997: 103).

A deictic center is a point to which the deictic element is anchored. Deixis can be absolute, i.e., place the deictic center at the location and moment of utterance. Lyons (1977: 578–579) finds that the speaker does not necessarily

need to adopt his or her own time and location as the deictic center, as in e.g., "I am going abroad next week". It is alternatively possible to adopt the spatio-temporal setting of, e.g., an addressee, as in "look left", to be understood as left of the addressee. Lyons calls this phenomenon deictic projection.[2]

It is also possible for the deictic center not to be grounded in the actual speech situation at all, which in the case of oral literature would be when the story is actually told. In this case, deictic projection is realized as a movement of the deictic center "from the speaker to an imaginary observer in the story world" (Diessel 1999: 95). It is thus possible to detach the deictic center completely from not only the temporal and locational setting of the speech, but also from the real world, and to place it at a time and place that never existed or will exist, inside an imaginary story. Segal (1995: 14–15) calls this "deictic shift" and finds that:

> this act of imagination was commented on over 2,000 years ago by Aristotle in his *Poetics*. He pointed out that poetry (tragedy, comedy, epic) was a mimetic art; its primary mode was to represent actions. The Greek word, *mimesis*, refers to imitation, or representation, or experience of that which is not literally present. [. . .] The deictic shift approach is consistent with phenomenological experience. When reading fictional text, most readers feel they are in the middle of the story, and they eagerly or hesitantly wait to see what will happen next. Readers get inside of stories and vicariously experience them. They feel happy when good things occur, worry when characters are in danger, feel sad, and may even cry, when misfortune strikes. While in the middle of a story, they are likely to use past tense verbs for events that have already occurred, and future tense for those that have not.

Deictic shift is applicable to written as well as oral narration. In oral narratives it is common for the narrator to first provide an introduction outside the frame of the story and with the deictic center in the actual speech situation (Zubin and Hewitt 1995: 131) by an introduction such as "Once upon a time there was . . .". Then the narrator has the option of moving the deictic center into the story by "decoupling the linguistic marking of deixis from the speech situation, and reorienting it to the major characters, the locations, and a fictive present time of the story world itself" (Zubin and Hewitt 1995: 131).

In the following two sections, spatial, discourse, and temporal deixis in the selected corpus will be discussed in detail, with a particular emphasis on demonstrating how spatial and temporal deixis is anchored inside the story rather than in the real world.

---

2 Levinsohn (2015: 144–145) provides a set of questions to determine whether and to what extent a particular language allows deictic projection.

3 Spatial and discourse deixis in the analyzed corpus

In this section, spatial and discourse deixis in the present corpus will be analyzed. All the variants under study have a two-way contrast, proximal versus distal deixis for demonstrative determiners[3] and pronouns. For deictic spatial adverbs, Sistani Balochi (Barjasteh Delforooz 2010: 138) and possibly also Koroshi Balochi (see n. 5), exhibit a three-way deixis with the locational adverbs "here, there1, there2".

It should be noted that the transcription of language data in the sections below follows that of the works in which the stories were originally published. I have, however, removed marking of stress in those texts that had stress marking, since stress marking was found in only two of the corpuses. The translation mainly follows that of the original works, but minor adjustments have sometimes been made, particularly to demonstrate overtly the meaning of the deictic items.

Direct speech has by default its deictic anchoring inside the story and is therefore of less interest in the discussion of deictic shift. Even so, there are interesting conclusions to be drawn from the use of deictic devices in direct speech as well, and they are therefore marked and discussed in the examples in sections 3.1–3.4 below. All forms that contain a deictic element are marked in bold and provided in brackets in the original language.

3.1 Koroshi Balochi

Four Koroshi Balochi stories have been investigated. The findings in two of these will be discussed below. The other two stories show a similar picture. Demonstrative determiners and spatial adverbs found in Koroshi Balochi are given in Table 1 and demonstrative pronouns are given in Table 2 (see also Nourzaei et al. 2015: 49–50).[4]

---

3 In fact, the contrast involving determiners is a three-way one between the absence of a determiner (e.g., "lion") and the presence of a proximal or a distal determiner (e.g., "this lion", "that lion").

4 Most of these forms can add an emphatic *ham* before the actual demonstrative (e.g., *hamē* 'this very', *hamā* 'that very', *hamēšī* 'of this very', *hamēšān* 'these very', *hamīdān* 'right here').

Table 1: Demonstrative determiners and spatial adverbs in Koroshi Balochi

	Proximal	**Distal**
Determiner	*ē, ī* 'this'	*ā* 'that'
Adverb	*īŋa(r), īŋā, ēdā(n), edā(n), eda, īdān* 'here' *ēdānākō, ēdānakō, īdānākō, īdānakō* 'right here'	*āŋa, ādā, odān, ōdān*[5] 'there' *ōdānākō* 'right there'

Table 2: Demonstrative pronouns in Koroshi Balochi

		Nominative	**Oblique**	**Genitive**	**Object**
Proximal	SG	*ī, ē, ēš*		*ēšī, īšī, ešī*	
	PL	*ēšān, īšān, ešān, šān*		*ēšānī, īšānī, ešānī*	*ēšānā*
Distal	SG	*ā*		*āhī, āī, āšī*	
	PL	*āšān*		*āšānī*	*āšānā*

I begin with a detailed investigation of spatial deixis in the story *Goli and Ahmad*, told by an experienced storyteller, Alamdar Samsanian, and published in Nourzaei et al. (2015: 130–146). In this story, the first four discourse units (see Nourzaei et al. 2015: 20) have the character of an introduction with the deictic center outside the story:

> Well, once upon a time … Well, a woman, there was a woman, her name was Goli. [She][6] was actually very bad. [She] was actually very bad. [She] was giving her husband a hard time, you know.

5 There is not enough data to determine if *odān/ōdān* and *ādā* show two different degrees of distal deixis in this corpus, as they do in Sistani Balochi (see Barjasteh Delforooz 2010: 138).

6 Personal and demonstratives pronouns, as well as the adverb "there", that are not found in the original text are given in square brackets. This also applies when there is an enclitic pronoun functioning as agent clitic in the ergative domain. However, non-canonical (dative) subjects expressed by enclitic pronouns are given without brackets (e.g., "to me there is a child", meaning "I have a child"). In order to make the text easier to read, other words that have been supplied in the English translation are placed in square brackets only if they occur in a deictic expression. No brackets are provided in summarized text sections. Other additions are not marked here.

Then there is a tense change to the non-past tense[7] (see also section 3), and the actual story starts. The next five discourse units take place on the very same day.

> So, one day her husband says, "Hey, Goli, **all these people** (*ī hāmmo mardom*) are going to pick green herbs **here [and] there** (*āŋa īŋa*), come on, let the two of us go, too". [She] says, "Very well, let's go". [They] go out into **this wilderness** (*hamī sahrā*), you know, [they] pick green herbs, **like this** (*āŋa īŋa*). The man goes and finds a well. The man finds a well. [He] says, "Hey, my wife, look into **this well** (*ē čāhā*), what is **this** (*ē*) [thing] that shines?" The woman comes, sir, to look into the well. The man pushes her in such a way that she (lit. 'the woman') falls into the well.

In this section, there are three proximal demonstrative determiners, one proximal demonstrative pronoun, and two occurrences of the phrase *āŋa īŋa* 'here and there, like this' (lit. 'there here'), which consists of both a proximal and a distal spatial deictic adverb. Three of the demonstratives[8] occur in direct speech, and one in the narration. The deictic center for the reported speech is, of course, the location of the speaker, but the deictic center for the narration is now inside the story, something that can be seen in the use of a proximal deictic device outside direct speech (i.e., this wilderness). Here we thus see that a deictic shift has taken place. In the next five discourse units the story evolves in the following way:

> [She] falls into the well and [he] comes back home. After four, five days her husband says, "I shouldn't have done like this, I shouldn't have thrown **this one** (*ēšī*) into the well. [She] was my wife". Anyhow, [he] becomes troubled, takes a rope and goes. [He] takes [it] and goes until [he] arrives at the well. [He] throws the rope into the well and says, "Hey, Goli, if you are alive, take hold of the rope, so that I can pull you up". Well, [he] pulls up the rope, **like this** (*hamītaw*). [He] sees the rope is heavy.

Here, there is only one proximal demonstrative pronoun in direct speech, and one adverb with proximal deixis in the narration. Although the husband is now rather far away from the well where he threw his wife, he is still able to use proximal deixis when he is talking about her. It seems that proximal deixis is used here to make a minor participant (in this case Goli) temporarily salient, i.e., the current center of attention (Levinsohn 2015: 139–140). In the next eight discourse units the story takes an unexpected turn.

---

7 Note that in Nourzaei et al. (2015) the story is translated into the past tense, since this is the default tense for narration of past events in English. Here, on the contrary, the translation reflects the actual tense form of the verb. The demonstratives are also translated with the corresponding English proximal or distal demonstrative to reflect the original structure.

8 The phrase *āŋa īŋa* is not counted here.

[He] pulls and pulls until [he] suddenly sees that a dragon came up. A dragon came up. [He] becomes panicky and wants to let go of the rope, but the dragon says, "Don't let go of the rope, I will give you whatever you want". The man says, "Fine, can you get the king's daughter for me?" The dragon says, "Yes, I will get [her] for you". So then **this one** (*ē*) (i.e., the dragon) says, "Very well, tonight I will go and wrap myself around the neck of the king's daughter. Then, no matter who comes (lit. 'came'), I will not unwrap myself except for you. When you come (lit. 'came'), I will unwrap myself. Then say, 'Oh king, if you are going to marry off your daughter, give [her] to me, and I will open up **this dragon** (*ē aždahā*)'". Well, in the evening the dragon goes and wraps itself around the neck of the king's daughter.

There are two proximal demonstratives in this section: one is a pronoun in a narrative section, indicating that the dragon, rather than the man, becomes the center of attention, and the other is a determiner in direct speech. The next five discourse units take place the next morning.

In the morning, when [they] get up, [they] see that the dragon is indeed wrapped around the neck of the king's daughter. Even though [they] bring all kinds of wise men, all kinds from **here [and] there** (*āŋa īŋar*), [it] doesn't unwrap itself. [They] say, "So, is there anyone left?" Someone says, "There is one person, a poor fellow who is called Ahmad, **that one** (*ā*) is left". The king says, "Well, go and bring **that one** (*hamāhī*), too!"

Here, in addition to the phrase "here and there" already discussed above, there are two distal demonstratives, both in direct speech. It seems that there is a need to establish a contrast between the people at the deictic center (the court) and Ahmad, who is in another place, so therefore the distal deictic is used. There is thus both a spatial and a mental distance between those who are speaking and Ahmad.

The next four units tell how Ahmad succeeds in becoming the king's son-in-law.

[They] go and when [they] bring Ahmad, Ahmad says to the king, "O king, if you are going to marry off **this your daughter** (*ē ǰanekat*), give [her] to me, so that I can unwrap **this dragon** (*ē aždahā*) from her neck". Well, then the king has no choice. [He] says, "Very well, I will marry [her] off, I will give [her] to you". Anyway, [he] marries her off, [he] gives her to Ahmad and then Ahmad goes and whispers something in the dragon's ear and the dragon goes away.

There are two proximal demonstrative determiners here; both are used in direct speech and for entities that are close to the speaker.

In the following six units there is yet another development in the story, when the dragon wraps itself around the neck of another king's daughter.

When (lit. '**here** that' *ēdān ke*) the dragon unwraps itself to go, well, it says, "Ahmad, I am leaving right away (lit. '[I] went') but if [I] ever wrap myself around anybody's neck, [you] should not come, you know! If [you] do, then [I] will get angry and eat you". [He] says, "Well, no, I won't come". Anyhow, the dragon goes its way. [It] goes its way, [it] goes to another town and wraps itself around the neck of another king's daughter. Even though [they] bring all the wise men from **here [and] there** (*āŋa īŋar*), sir, [it] doesn't unwrap itself. [It] doesn't unwrap itself until people say, "In such-and-such a town there is a person called Ahmad, [he] is the king's son-in-law. **That one** (*ā*) can unwrap it".

In this passage there is one distal demonstrative pronoun in direct speech. This distal demonstrative indicates a long distance from the present deictic center, i.e., Ahmad who is in another town. In addition, it brings out the contrast between the wise men who failed and Ahmad, who is assumed to be able to solve the problem. There is also one proximal spatial adverb in the conjunction that denotes a temporal relation between the subordinate and the main clause at the beginning of the passage, and one occurrence of the phrase "here and there" discussed above.

In the following six units, Ahmad is approached to solve the problem. He is hesitant at first, because of the dragon's previous threat. There are no demonstratives in this passage.

So [they] go to find Ahmad. But Ahmad, who is dead scared of the dragon, is not coming. Anyway, the king says, "No, you must go. [It] is improper", and such things. [He] sends him away. Ahmad is coming but [he] is worried, you know, and says to himself, "What should I do? The dragon will eat me", things like that. Then, in the middle of the road, suddenly [he] gets an idea. [He] gets an idea about what to say (lit. 'what [I] should say'). [He] says, "Great!" [They] see that Ahmad, who was very worried before, is now laughing and happy. Someone says, "Ahmad, how are you feeling?" [He] says, "Never mind, let's go, I will unwrap the dragon".

In the following ten units the story is resolved.

[They] go and go and when they arrive **here on this side** (*hamī īŋare*) of the court, you know, [they] see that indeed the dragon is wrapped around the neck of the king's daughter. So, **these** (*šān*), go closer. When the dragon's eye falls on Ahmad, [it] gets angry. [It] says, "Well, didn't [I] tell you not to come?!" Ahmad says, "I didn't come to say 'unwrap yourself' actually. [I] didn't come to say 'come loose, and go!'" The dragon says, "So what do you have to say?" Ahmad says, "I only have a message for you". The dragon says, "What is [it]?" Ahmad says, "Goli has come out of the well and she is looking for you!" Sir, the dragon unwraps itself out of fear, and how [it] is running!

In this final part of the story, there is one proximal spatial adverb, one proximal demonstrative determiner, and one proximal demonstrative pronoun. Again,

Ahmad and his people arrive close to the palace; they do not pass by it in relation to where they came from. Where Ahmad (the main protagonist) is, the deictic center is there as well.

The final unit is the formal ending, which does not belong to the actual story; in it the deixis has shifted from inside the story back to the actual occasion of the narration, thus the distal demonstrative determiner.

> Now, may our enemy experience what Goli did and our friend what **that Ahmad** (*hamā ahmad*) did.

In the rest of the material from Koroshi Balochi a similar picture emerges. The longest story in the corpus, entitled *The King's Son* and published in Nourzaei et al. (2015: 162–209), can be summarized as in Table 3:

Table 3: Deixis in *The King's Son*

	Proximal	Distal
Determiner	41	8
Pronoun	26	2
Adverb	13	1

There are many examples where proximal deixis is used for objects and persons physically remote from the deictic center, but still at the current center of attention. Unit 12 (Nourzaei et al. 2015: 165) is said about a foal that has not yet been born, but which is the current center of attention due to the magical power it will possess. The previous two units have described how it should be raised, and the following statement is then made:

> Then **this foal** (*ē korrag*) can provide you with whatever you may want.

In unit 112 (Nourzaei et al. 2015: 200), a woman is talking to her husband about her sisters, who are not physically with them, but who are the reason why she is sad, and thus are important in this context. She says, I know that you are not bald, I know who you are, **these** (*īšān*) are ridiculing me.

Distal demonstratives can create a contrast between two different entities. In units 75–76 (Nourzaei et al. 2015: 188), the six elder sisters, marked with a distal demonstrative, do one thing, and their little sister, marked with a proximal demonstrative, does something else and unexpected, which attracts attention. In fact, they are all at the same location when the action takes place:

So, **those six sisters of hers** (*ā šīš gāhārī*), each one of them hits someone, one, for example, hits the vizier's son in the chest. In short, each one hits a rich person, you know. [They] hit some boys, but **this youngest girl** (*ī kassānoēn ǰanek*) doesn't throw her apple.

Two more instances of a distal demonstrative being used to create a contrast between two entities are found. One is in the reported speech of unit 98 (Nourzaei et al. 2015: 195) and the other is in the narration of unit 111 (Nourzaei et al. 2015: 200).

Then [they] say well, **these six sons-in-law** (*ī šīš dūmād*) say, "Well, there is [another] one too, he has (lit. 'there is to him') a lame mule, [he] comes afterwards, give him the meat, **that one** (*ā*) should bring [it]".

... **these six sisters of hers** (*ē šīš gāhārī*) keep ridiculing **that one** (*āī*) (i.e., the youngest sister) ...

Distal demonstratives are also found as a reactivating device. In unit 70 (Nourzaei et al. 2015: 186), the two entities that are marked with the distal demonstrative determiner are present at the current deictic center, but they are marked with the distal demonstrative for discourse deictic purposes, i.e., to reactivate a previously mentioned topic without changing the deictic center. The proximal demonstratives in units 69 and 70 (*ē/ī bāġā, ī*) (Nourzaei et al. 2015: 185–186) indicate the current center of attention.

She sees, dear Lord! [There] is a rider on a horse in **this garden** (*ī bāġā*), [he] is riding around, it is as if [there] is an angel riding around in **this garden** (*ē bāġā*). After a while [he] stops, [he] stops. [She] sees that [he] gets off the horse. [She] goes towards [him] and **this one** (*ī*) came towards [her] too. [He] pulls **that same aforementioned stomach** (*hamā komaokā*) over his head. [She] sees that it is **that very bald one** (*hamā kačal*) who works in their garden.

Two distal demonstrative determiners are found in the very last unit of the story, unit 141 (Nourzaei et al. 2015: 209), where the spatial anchoring is no longer located within the story but is at the moment of narration:

Now, may it happen to our friend **like to that king's son** (*čō hamā šāhay bačā*) and to our enemy **like to those six servants** (*čō hamā šīš nawkarā*).

There are three more distal demonstrative determiners in the text, which will be discussed in section 4.1 because they are all connected to temporal adverbs.

3.2 Sistani Balochi

For Sistani Balochi nine stories have been included in this study (Barjasteh Delforooz 2010: 286–325, 336–391). The results for two of them will be presented here. The rest provide a similar picture. The demonstrative determiners and spatial adverbs found in the Sistani Balochi texts in this corpus are presented in Table 4, and the Sistani Balochi demonstrative pronouns in Table 5 (see also Barjasteh Delforooz 2010: 138, 150–151).[9]

Table 4: Demonstrative determiners and spatial adverbs in Sistani Balochi

	Proximal	Distal 1	Distal 2
Determiner	*ē, ī* 'this'	*ā* 'that'	
Adverb	*idā* 'here'	*ōdā* 'there'	*ā(d)dā* 'far away there'

Table 5: Demonstrative pronouns in Sistani Balochi

		Nominative	Oblique	Genitive	Object
Proximal	SG	*ē, yē, ī, ēš*	*ēšī, ešī*		*ēšā, ešā, ēširā, ēšīrā*
	PL	*ē, ī, ēš*	*ēšān*	*ēšānī, ešānī*	*ēšānā*
Distal	SG	*ā*	*ā(y)ī*		*ā(y)irā*
	PL	*ā*	*āwān*	*āwānī*	*āwānā*

One of the Sistani Balochi tales, *The Story of the Lion and the Three Bulls*, told by the experienced storyteller Paraddin Gorgej (Barjasteh Delforooz 2010: 378–383), is here presented in its entirety. Like those in Koroshi Balochi, the stories in Sistani Balochi exhibit deictic shift, and in the story below proximal deixis is overwhelmingly predominant, to such an extent that nouns are used more frequently with a determiner than without.

> Sir, [they] say, there were three cows. **All these three cows** (*ī har say gōk*) were in unity (lit. 'in one heart'), there was a black one, there was a light brown one, and there was a white one. **These** (*ē*) were in unity from **that [old] time** (*amā waxtā*). Wherever [they] grazed no beast of prey had any power against **these** (*ešānī*). If there was a lion, if there was a leopard, if there was a wolf, that attacked one [of them], all three attacked [it] and no beast of prey attacked **these** (*ēšānī sarā*), because they were of one heart and of one mind. What did a certain lion do? [It] was stalking **these** (*ēšānī*), [it] was lying in ambush for **these** (*ēšānī*). It said, "Unless I change the mind of each one of **these** (*ēšānā*), I won't

---

9 These forms can add an emphatic *am* before the actual demonstrative (e.g., *amē* 'this very', *amā* 'that very', *amēšī* 'of this very', *amēšān* 'these very', *amidā* 'right here').

be able to eat **these** (*ī*) (lit. 'these won't be eaten')". **This lion** (*ē šīr*) came and lied to **these** (*ēšānā*), told **these** (*ēšān*) a lie, "O fellows, you are **this kind of [good] friends** (*ē rangēn rapēγ*), I am your fourth brother. I have seen a pasture in a place. [It] is very green (lit. 'spring'). You . . . let's go **there** (*ōdā*), I will take you there, you eat **that grass** (lit. 'spring') (*ā bahārā*). I will watch over you", the lion said. [It] deceived (lit. 'made donkey') **these** (*ēšānā*) and took [them]. When [it] took [them], [it] said, "Now I will watch over you **on this mound** (*ē dikkayay sarā*), you eat **this grass** (*ē bahārānā*)". [It] passed, that day (lit. 'today') passed and the next day (lit. 'tomorrow') passed, [it] whispered in the light brown cow's ear and in the white one's. [It] said, "O friend, your hair and my hair are the same colour, **that black [one]** (*amā siyāhēn*) is ill-matched (lit. 'unripe') among us. You, don't help [it], I will eat **that [one]** (*āyirā*), a lot of grass will remain for you". [It] confused **these ones'** (*ēšānī*) minds and one of the cows said, "Fine". So it seized the black one. **Those** (*ā*) didn't help. When [they] didn't help, [it] overpowered **this** (*ešī*) single one. [It] ate **this one** (*ēšā*) up. When [it] ate and finished it . . . , when [it] finished **this** (*ešā*) [it] whispered in the light brown cow's ear and said, "Your hair and mine are the same colour, you, don't help [it], when I eat **this** (*amē*) white one, all the grass will remain for you. Then I will watch over you **here** (*idā*), you can eat!" [It] said, "Fine". When [it] seized **this one** (*ēširā*) too, **that one** (*ā*) (i.e., the light brown one) didn't help. The lion ate **this one** (*ēširā*) (i.e., the white one) too. When [it] finished **this one** (*ēšā*) too, [it] said to **that one** (*āyirā*), "Now [I] alone am more powerful than you". [It] ate **that one** (*āyirā*) too. In **this manner** (*ē rangī*), with **this trick** (*ē sīyāsat*), [it] destroyed **these ones** (*ēšānā*).

Proximal deixis must be seen as the default in this story. It is interesting to note that even in the final comment "in this manner, with this trick, it destroyed these ones", where a shift out of the actual story may have been expected, proximal deixis remains. In addition to the proximal deictics, the story provides some interesting examples of distal deixis. The first three, "**that** (*ā*) old time", "**there** (*ōdā*)", and **that grass** (*ā bahārā*) can easily be explained as indicating a distant time and a distant place, far from the present deictic center. The following five ones, though, all indicate entities that are present at the deictic center; however, a contrast needs to be established between the cow that is to be eaten and the other cow(s).

Another story in Sistani Balochi, *Xarmizza* (Barjasteh Delforooz 2010: 286–293) contain the following deictic items (Table 6):

Table 6: Deixis in *Xarmizza*

	Proximal	Distal
Determiner	20	—
Pronoun	30	1
Adverb	1	3

As in the previous Sistani Balochi tale, here, too, proximal deixis is completely predominant when compared to distal deixis. This time, however, the proximal

determiner is used less frequently than the simple noun.[10] There are several examples of how a proximal deictic determiner or pronoun is used even if the entity referred to is not present at the deictic center, such as when the king in his palace refers to a dragon far away from the palace with proximal deixis (units 14–20, Barjasteh Delforooz 2010: 287). The use of the verb "to bring" clearly shows that the dragon is not present at the palace:

> [They] came back. One of them said, "Lord king, [it] is a dragon". The king said, "Oh …
> **this dragon** (*ē aždiyā*) has something to say, [it] has a petition, who can bring **this one's**
> (*ēšī*) petition to me?"

In another very interesting example, proximal deixis is used to refer to the palace, although the story has now moved from the palace to the mountain, where the dragon obviously lives (units 34–43, Barjasteh Delforooz 2010: 288–289). However, the passage starts out with a distal deictic adverb indicating a movement from the former deictic center (the town, where the dragon has come to seek help for a problem) to the new deictic center (the mountain, where the dragon takes the carpenter who has come to help).

> The dragon gave the carpenter one hint after another and went onto the mountain. When
> [he] went **there** (*ōdā*), good heavens ohhh … **this dragon** (*ē aždiyā*) has a mother as well.
> **This one's** (*ēšī*) mother, there are (lit. 'aren't') **this [kind of] wild mountain goats** (*amē
> kōhay pāčin*) with big horns, it has caught one of **these** (*amēšān*) and **these** (*amē*) its horns
> have got stuck in the mother's throat and [it] is short of breath, the dragon has come **here**
> (*idā*) and informed the king.

In this example, "has come **here** (*idā*)" refers to the previous deictic center, i.e., the town where the previous scene was located, rather than to the mountain, where the hero and the dragon have now gone.

There are a few additional distal demonstratives to account for, apart from the one in the example above "**there** (*ōdā*)". Two more occurrences of "**there** (*ōdā*)" are found in the text, one referring to a place that is not the current deictic center (Barjasteh Delforooz 2010: 287) and one at the very end of the story "the name *xarmizza* ('melon', lit. 'donkey tasted') remained from **there** (*ōdā*)" (Barjasteh Delforooz 2010: 293). The latter occurrence must be seen as an addition to the actual story where the narrator shifts the deictic center from within the story to the moment of narration to account for why this fruit is called what it is even now. The one distal demonstrative pronoun in this text (Barjasteh Delforooz 2010: 289) is found in a context where there is a need to create a contrast between two entities that both are present at the deictic center, namely

---

10 For example, the noun for "king" occurs once with the proximal determiner and six times with no determiner.

the carpenter, who rescues the mother dragon and is the participant through whom the story will continue to develop, and the mother dragon herself, who has no further part to play in the story:

> **This one** (*ē*) (i.e., the carpenter) sawed the wild goat's horns and **that one** (*ā*) (i.e. the mother dragon) was rescued.

3.3 Vafsi

Seven stories have been investigated for Vafsi. The results from four of them are presented here. The other stories show similar results. The demonstrative determiners and adverbs found in the Vafsi texts are presented in Table 7 and the demonstrative pronouns in Table 8 (see also Stilo 2004: 225, 227).

Table 7: Demonstrative determiners and adverbs in Vafsi

	Proximal	Distal
Determiner	*in* 'this'	*an* 'that'
Adverb	*indi, indiænæ* 'here' *ena* 'this way'	*andi, andiænæ* 'there' *ana* 'that way'

Table 8: Demonstrative pronouns in Vafsi

		Nominative	Oblique
Proximal	SG	*in*	*tini, intine, intini*
	PL	*ine*	*tinan*
Distal	SG	*an*	*tani, tane, antane*
	PL	*ane*	*tanan*

Like in Koroshi and Sistani Balochi, deictic shift and an overwhelming predominance of the proximal demonstratives are found in the Vafsi narratives. This can clearly be seen in the tale entitled *Moses, the holy hermit and the infidel chieftain* (Stilo 2004: 110–115). This is the story in full:

> Once his holiness Moses[11] went to Mt. Sinai and asked, "Oh God, is a merciful infidel[12] better (lit. 'good')? Or a merciless Muslim?" A voice proclaimed from the heavens, saying, "Of course, a merciful infidel is better than a merciless Muslim. Now go out from **this mountain** (*in ku*) and see for yourself". When Moses went out off **this mountain** (*in ku*),

---

11 The phrase *hæzræt-e musa* 'his holiness Moses' with the honorific *hæzræt* will be translated as only 'Moses' in the rest of the story.

12 There are two words translated 'infidel' in this story. In the general discussion at the beginning of the story the word *kafær* is used. Later on in the story the word *gæbr* is used as well, sometimes in combination with *kafær*.

he went up to **this top of a mountain** (*in ku-kællæ*) and saw that a holy hermit – he wasn't actually **there** (*andi*) himself – had put some bushes under **this one's** (*tini*) cooking pot as firewood, and the fire had gone out. Moses said, "Well, since **this [fire] under the pot** (*in zer-dizi*) went out, I will light **this [fire] under the pot** (*in zer-dizi*). When **this holy hermit** (*in abed*) comes back, maybe [he] will give me a little of **this lunch of his** (*in nahares*), let me take the trouble to light the fire myself". When Moses came to light **this [fire] under the pot** (*in zer-dizi*), the bushes got bumped and everything in the pot spilled out. Just as it spilled out, **this holy hermit** (*in abedæ*) came back and said, "Why did you spill the pot out?" [Moses] said, "Honestly, I wanted to make **this** (*in*) better, but [it] turned out bad to rekindle the fire under the pot". "No, what business was [it] of yours?" And the hermit threw two or three punches at Moses and Moses threw two or three punches at the hermit, and through an act of the Lord the backs of **these** (*tinan*) two stuck together. After the backs of the two had stuck together, as [they] walked along, Moses carried **this one** (*tini*) on his back for a while and then the hermit carried Moses for a while. Until [they] arrived, that is to say, to a plain where [they] saw, wow, a hundred tents were pitched. **These infidels** (*in gæbre*)[13] . . . [these] were the tents of infidels.

[They] got up to **these infidels** (*in gæbran*) – their chieftain himself was 90 years old . . . was 100 years old and his wife 90 years old. **These** (*tinan*) hadn't had any children. But the Lord had just given **these** (*tinan*) a newborn in swaddling clothes. [He] had just given **these** (*tinan*) a newborn when **these** (*ine*) two arrived and **here** (*indi*), so to say, were **this one's** (*tine*) guests and **this infidel chief** (*in sær-kærde-ye-gæbri, gæbr-e kafæri*) said, "Well, how is it that you have got **like this** (*æzin*)?" [They] said, "Well, our story is **like this** (*æzin*), it happened **like this** (*æzin*). [We] got into a fight and we ended up **like this** (*æzin*)". [He] said, "well, you . . . is there no cure for **this** (*in*)?" [He] said, "Yes, there is a cure". Moses turned and said, "[There is] a cure. But [it] should be an infidel. [He] should be the chief of 100 tents. He should be 100 years old himself. His wife should be 90. [They] should have had no son, but the Lord should just have given **that one** (*tani*) a son. [They] should take **that son** (*an lazey*) and slit his throat (lit. 'cut his head'), right between the two of us, then we will come apart from one another". **This infidel's** (*in gæbri*) wife had gone to the bathhouse and had put **this baby** (*in zarru*) to sleep in the cradle. **This infidel** (*in gæbri*) considered it for a bit and said, "Well, I didn't have a child until now. And now, God has willed [it] (lit. 'it pleases God'). Now, let me slit **this one's** (*tini*) throat so **these** (*in*) can get apart from one another, I don't want **this child** (*in zarru*) any more". **This infidel** (*in gæbr*) goes and gets a pen knife and slits his son's throat, right between Moses and the hermit, and **these** (*ine*) come apart from one another and start walking away.

[They] start walking away and [he] quietly puts **this newborn** (*in qondaq*) back into the cradle and covers his head, but when the wife returns from the bathhouse, **this infidel** (*in gæbr*) says, "Oh God, what should [I] say in answer to my wife?" **This one** (*in*) starts to beseech the Lord. The wife comes back and says, "Husband, the baby hasn't woken up, has he?" [He] says, "No, [he] hasn't woken up yet. All of a sudden, [he] sees the baby crying. As the baby is crying, the woman comes and picks up the newborn to breast-feed [him] and **this infidel** (*in gæbr*) comes over to inspect **this newborn's** (*in qondaqi*) throat. The woman asks him, "Husband, what are you looking at so closely?" [He] tells her, "Wife, to tell the truth, [it] happened **like this** (*æzin*). I slit **this one's** (*tine*) throat and [it] happened **like this** (*æzin*). The wife says, "Husband, **those two people** (*an do næfær*)

---

13 The different case forms of the noun *gæbr* found here are direct singular (*gæbr*), oblique singular (*gæbri*), direct plural (*gæbre*), oblique plural (*gæbran*) (see Stilo 2004: 223).

were people of very high standing. Run and bring **those ones** (*tanan*) back for a feast". **This one** (*in*) runs, actually goes and brings **these ones** (*tinan*) back. [He] says, "[You] are people of very high standing". [He] brings **these ones** (*tinan*) back and puts on (lit. 'gives') a feast. When [he] brings **these ones** (*tinan*) back, **this one** (*ine*) says, "I am Moses". So **this infidel** (*in gæbr*) and the people in the hundred tents all become Muslims. When Moses goes back to Mt. Sinai God asks, "O Moses, is a merciful infidel better? Or a merciless Muslim? [He] answers, "O God, a merciful infidel". And [that] is the end. **This** (*in*) is the Vafsi for it.

Deictic shift applies to the whole story, even in the comment at the end "**this** (*in*) is the Vafsi for it". In this story, the vast majority of spatial and discourse deictic lexical items thus denote proximal deixis and place the narrator and the audience inside the story. There are only five distal deictic items. The first one, "he wasn't actually **there** (*andi*) himself", is an explanatory comment to the audience outside the story. The two following ones, "the Lord should [just] have given **that one** (*tani*) a son" and "they should take **that son** (*an lazey*) and slit his throat" occur in a description of an imagined situation. The final two deictic devices are "**those two people** (*an do næfær*) were people of very high standing. Run and bring **those ones** (*tanan*) back for a feast". In this case, distal deixis is in fact used for persons who are far from the deictic center, but it could also fulfill the function of highlighting a climactic statement.

In the three stories *The Needle Dirties Himself*, *The Molla and the Jew*, and *Shangol, Mangol and Dastegol* (Stilo 2004: 49–57) there are no distal deictics whatsoever, except for the adverb *ana* 'that way' in the expression *ena-o ana* 'this way and that way' (unit 41, Stilo 2004: 56). There are, however, several instances of a proximal deictic demonstrative being used to denote an entity that is not present at the place where the speaker is, such as in unit 33 (Stilo 2004: 48–50), where the needle says "[I]'ll go tell **this mouse** (*in muši*) [and] he'll come make a hole in your bed". The presence of the verbs "to go" and "to come" show that the mouse is not where the needle is, though it is the center of attention within the reported speech.

3.4 Gorani

Three texts in the Gawraǰūyī dialect of Gorani have been analyzed in this study. Spatial and discourse deictic demonstratives and adverbs in this dialect are presented in tables 9 and 10 (see also Mahmoudveysi et al. 2012: 15, 17).

Table 9: Demonstrative determiners and adverbs in Gorani of Gawraǰū

	Proximal	Distal
Determiner	*ī* 'this'	*ā* 'that'
Adverb	*īnā, īnahā* 'here'	*āna, ānā* 'there'

Table 10: Demonstrative pronouns in Gorani of Gawraǰū

Proximal	SG	*īn(a), īnī*
	PL	*īnān(a), īn(ak)ānī*
Distal	SG	*ān(a), ānī*
	PL	*ānān(a), ānānī*

The three stories *Tītīla and Bībīla*[14] (Mahmoudveysi et al. 2012: 63–77), *The Tale of Bizbal* (Mahmoudveysi et al. 2012: 81–88), and *Mard and Nāmard* (Mahmoudveysi et al. 2012: 96–103) will be analyzed here. The text *Mard and Nāmard*, which contains the largest number of deictics, is presented in full. Parts of the story where there are no deictics are summarized in brackets.

> Well, where should [we] begin, where should [we] hear [it], the story of two friends, two men. Both of them go looking for work. [They] are together; their names are Mard and Nāmard. Both of them make a contract together; one says, "Brother". The other says, "Yes?" The first one says, "[We] will go find work to do and a town, a place, where we may earn a morsel of bread for our children, and [we] will come back again together". [They] say, "All right". From home, **this one** (*ī(n)*) wraps up bread and [other] victuals and ties [it] to his back. **That one** (*ānī*), simply brings bread and [other] victuals and ties [it] to his back. (They eat and then they fall asleep. Nāmard wakes up first. He takes off with all the foodstuffs. When Mard wakes up, he realises that he has been abandoned.) [He] goes a long way until he reaches the inside of a mill, a machine. [He] goes inside **there** (*āna*); [it] is old, nothing anymore, for example, [they] do not work in it anymore. Eh, [he] sits down inside **there** (*āna*), [he] hides himself. (A bear, a wolf, and a lion come.)[15] One of them says, "[Here] is the scent of a human being!" One of them looks around and says, "There is nothing, believe [me], no, [there] is no human being in **this place** (*ī dawray*)". [They] sit down and – like me now – they tell a story.
>
> **This one** (*īn*) (i.e., the wolf) says, "Brother". [They] say, "Yes?" [It] says, "The king's daughter has gone insane. Do [you] know what the cure for her is?" It is **this** (*īna*). The wolf speaks. **These** (*īnakānī*) say, "No". The wolf says, "They tried all kinds of medicine and remedies, but there has not been a cure for her. The dog with the flock, if I only were a human being, [I] would have killed **that dog** (*ā tūta*) with the flock, [I] would have taken out its brain, [I] would have left it out in the sun, so [it] would have become dry. [I] would have ground [it], [I] would have brought [it], [I] would have steeped [it] like tea, [I] would have given [it] to the king's daughter, so she becomes completely well again". The man says: "Well, **this** (*īna*) is the first of the stories.[16] The wolf says, "As for me, I would eat,

---

14 The story *Tītīla and Bībīla* is a longer version of the same story as *Shangol, Mangol and Dastegol* in Vafsi, and *The tale of Bizbal* is also structurally very similar to *The needle dirties himself* in the Vafsi corpus.

15 The storyteller changes the three animals throughout the story. The bear turns into a leopard, and on one occasion a dog is also mentioned.

16 An alternative translation of *īna* here and below is 'it is so/thus' (personal communication, Denise Bailey).

be full with its meat, of the flock". It is not my concern anymore, finally then, **that one** (*āna*), **that one** (*āna*) tells another story, thing. The lion answers, he says: "Have you seen **this tree** (*ī dāra*) **outside this mill** (*az ī bar āsyāw*)? [It] has become dry, **this** (*īna*) has not brought forth fruit for several years. If only I were a human being, if [I] could find a way for the tree to spread its roots. There were three royal vases in it. [They] are full of gold and precious stones. If only I were a human being, if [I] would have found [it], **this tree** (*ī dāra*) too would have then born fruit. **This** (*īna*) is the second story. The leopard answers, saying, "Inside the mill, whatever they did, it has not worked. You must find it. There are also two vases in it. If the owner would come for attending **this mill** (*ī āsyāwa*), he would put it to work, it would start to work". **This** (*īna*) is all three stories.

Brother, as for the man, Mard, [he] simply listens until the early morning becomes day. (Then he goes, finds the vases, puts them in a place where he can find them later, finds the flock, kills the dog and takes out its brain.) [He] takes [it] out in **that same way** (*ā jūr(a)*) the wolf said, [he] puts it out in the sun, [it] becomes dry, and [he] grinds [it] and puts [it] into his bag. [He] sets off on his way, [he] goes. [He] goes, [he] reaches the city, where [he] sees that, yes **this** (*īna*) is [it]. The king whose daughter has become insane is from **this city** (*ī šāray*). Finally, [he] reaches **there** (*ānā*) and says, he goes and knocks on the door of the king's house, and **this one** (*īn*) (i.e., a person at the door) says, "Who is [it]?" **That one** (*ān*) (i.e., another person) says, "Who is [it]?" And [Mard] says, "I have come to cure your daughter, [I] am a doctor". **These ones** (*īnān*) in turn say, the people in the king's house say, "So many medicines and doctors came and [they] brought remedies, and the doctor gave medicine; his medicine did not bring about healing. (Mard claims to be different and promises to cure the king's daughter within a few days.) The king says, "What is **this** (*īna*)?" The servant says, "By God, a young man has come, [he] says, 'I will cure his daughter'. Your highness, what do [you] command?" The king says, "Let [him] come upstairs, no problem, [he] is welcome, **this one** (*īn*) too, up like those doctors". The king says, "Well, doctor". (Mard introduces himself and asks what he will get if he cures the king's daughter.) The king says, "My daughter, as a gift, [I] will give [her] to you, [I] will also give **this crown and throne** (*ī tāǰ-u taxt*) of mine to you". Mard says, "No, may your crown and throne be a gift to yourself. But if [I] cure your daughter, then [I] want your daughter in marriage". The king says, "So be [it], may [she] be a gift to you". So [they] make a contract **there** (*āna*). **This one** (*īn*) (i.e., Mard) also goes, [he] goes a little way to attend to the girl. (He gives her the medicine, trying to remember what the wolf had said.) "After that, for example, then, anoint her back and **these** (*īnān*) (i.e., some other body parts) with it, put the medicine on it until she is well again". (The girl gets well and is given to Mard in marriage. Then Nāmard turns up.)

Nāmard says, "I recognized you. You are Mard, [you] are indeed Mard, what have [you] done that you reached **this [high] position** (*ī pāya*)? I wander about in **this state** (*ī jūra*) without purpose, [I] still have achieved nothing, nothing at all". Mard says, "You are not a good man, [you] have proven yourself. We were friends, you yourself stole the bread and went your way. [You] did not wait right at that moment. I was so hungry, [I] ate earth. You man without a conscience! Nevertheless, now I will also give you **this advice** (*ī řāwēza*), listen! Me, from then on, **this God** (*ī xwiyā*) had mercy on me, [he] placed **this much good** (*ī hamkay xayřša*) in front of me (lit. 'my mouth'). Go into the mill, to a corner high up, a leopard and a dog and a lion, [they] come back in the evening, [they] talk. Listen to their stories". Nāmard says, "Fine!" Brother, **this one** (*īnī*) goes at once. They have a pipe for the

stove, Nāmard goes and just sits up on that stovepipe and makes himself very comfortable. In the evening, the wolf and the lion and the leopard return. [They] say, "[Here] is the scent of a human being!" [They] grab Nāmard by his leg, bring him down and tear him to pieces. A bouquet of flowers, a bouquet of narcissus, may I never see your death, never.

The proximal demonstrative determiner is not used as frequently in this text as in the Koroshi Balochi, Sistani Balochi, and Vafsi texts. However, it is still clear that we find deictic shift and that the deictic center is mostly inside the story. The four occurrences of the deictic adverb "**there** (*āna, ānā*)", however, seem to break this pattern and move the deixis outside the story. The four occurrences of the distal demonstrative pronoun can be explained as establishing a contrast between Mard and Nāmard, between the first and the second animal who tell their stories, and between the first and the second person at the king's door, while the two occurrences of the distal demonstrative determiner can be explained as indicating spatial and temporal distance. When the wolf tells the story, **that dog** (*ā tūta*) is not present, and **that same way** (*ā jür(a)*) refers back in the discourse to when the wolf had told its story. There is a formal ending to this story as well, but there are no deictic elements in it.

The other two stories *Tītīla and Bībīla* and *The Tale of Bizbal* contain the following deictic elements (Table 11):

Table 11: Deixis in *Tītīla and Bībīla* and *The Tale of Bizbal*

	Proximal	Distal
Determiner	12	5
Pronoun	8	3
Adverb	—	—

Distal deixis with a determiner is found four times in the expression "the other side" of the river (units 6, 7, 12, Mahmoudveysi et al. 2012: 63–64), thus indicating a contrast to the side of the river where the deictic center is located:

> [They] go to the Zimkān [river]; the flock goes to **that** (*ā*) [other] side.

A distal demonstrative pronoun is found once in connection with the main protagonists, where they are probably both contrasted with "the flock" and also highlighted (unit 4, Mahmoudveysi et al. 2012: 63):

> [They] are at home; the flock goes to the mountains and **those** (*ānān*) stay at home.

Another distal demonstrative pronoun (unit 75, Mahmoudveysi et al. 2012: 73) occurs after a dialogue, and refers back to the protagonist who was not the

last speaker (the wolf), indicating that the wolf remains the current center of attention. Thus, the distal demonstrative creates a contrast between the wolf and the goat:

> The goat says, "Morning, at midday [there] will be war. [I] will come to the square and [we] will fight". The wolf says, "All right". **That one** (*ānī*) (i.e., the goat) comes, comes to Lālo Pāydar ... (units 73–75, Mahmoudveysi et al. 2012: 73).

However, in another example (unit 40, Mahmoudveysi et al. 2012: 84), the distal demonstrative pronoun does refer back to the last speaker (Auntie Tahmineh). The effect is to direct attention back to the previous speaker (the cat), who is the main character of the story:

> So the cat goes and says to the mother of Čiman, the cat says, "Auntie Tahmineh, please send your daughter, [she] should dance for one hour". **That one** (*ānī*) (i.e., Auntie Tahmineh) says, "Brother, she has no shoes, you must go and make shoes for her". The cat says, "Sure". (units 39–41, Mahmoudveysi et al. 2012: 84)

4 Temporal deixis in the analyzed corpus

The main means of anchoring a story in time is the tense form of the verbs. In the following subsections, the use of tense in Koroshi Balochi, Sistani Balochi, Vafsi, and Gorani narratives will be investigated. Interesting observations about temporal adverbs will also be noted.

4.1 Koroshi Balochi

In Koroshi Balochi, "the default tense of narration of past events [...] is the non-past form" (Nourzaei et al. 2015: 20). The introduction of a story, as well as background material in the story, are in the past tense, whereas all the events in the main story line are in the non-past tense. This can be observed in the story *Goli and Ahmad*, published in Nourzaei et al. (2015: 130–146), which was presented in full in section 3.1 above, as well as in the other stories in the corpus. The verb forms at the beginning of *Goli and Ahmad* are presented below as they are in the original text. (Note that the translation of the verb forms was somewhat modified in section 3.1 to give a more idiomatic translation into English.) Verb forms in direct speech have not been analyzed, since they are not part of the story line. In fact, it could be argued that verbs in subordinate nominal clauses after verbs of perception are not part of the story line either, but they have been included in the analysis because they exhibit an interesting variation in tense

use. The same pattern as in this extract is found in all the tales in Koroshi Balochi.[17]

> Well, once upon a time (lit. 'Well, **there was** [*ad*] one, **there was** [*nayad*] no one, except for God, **there was** [*nayad*] no one') well, a woman, **there has been** (*boda*) a woman, her name **has been** (*boda*) Goli. Actually she **has been** (*boda*) very bad. Actually she **has been** (*boda*) very bad. She **has been troubling her husband** (*azzīyate šūay makanā boda*), you know.
>
> So, one day her husband **says** (*ašī*), "Hey, Goli, all these people are going to pick green herbs here and there, come on (*byā*), let the two of us go, too". She **says** (*ašī*), "Very well, let's go". They **go** (*arran*) out into this wilderness, you know, they **have been picking** (*mačenēn boda*) green herbs, like this. The man **goes** (*arra*) and **finds** (*pēdā akant*) a well. The man **finds** (*pēdā akant*) a well. He **says** (*ašī*), "Hey, my wife, look into this well, what is this thing that shines?" The woman **comes** (*akay*), sir, **to look** (*say kan*) into the well. The man **pushes her** (*lohe adā*) in such a way that she **goes** (*arra*) into the well.
>
> She **goes** (*arra*) into the well and he **comes** (*akay*) back home. After four, five days her husband **says** (*ašīt*), "I shouldn't have done like this, I shouldn't have thrown her into the well. She was my wife". Anyhow, he **becomes** (*abī*) troubled, **takes** (*azo*) a rope and **goes** (*arra*). He **takes** (*azo*) it and **goes** (*arra*) until he **arrives** (*arasī*) at the well. He **throws** (*aprēnī*) the rope into the well and **says** (*ašī*), "Hey, Goli, if you are alive, take hold of the rope, so that I can pull you up". Well, he **pulls** (*akašī*) up the rope, like this. He **sees** (*agennī*) the rope **is** (*en*) heavy.
>
> He **pulls** (*akašīd*) and **pulls** (*akašīd*) until suddenly he **sees** (*agennīt*) that a dragon **came** (*āk*) up. A dragon **came** (*āk*) up. He **becomes** (*abī*) panicky **to let go** (*wel dā*) of the rope, but the dragon **says** (*ašī*), "Don't let go of the rope, I will give you whatever you want". The man **says** (*ašī*), "Fine, can you get the king's daughter for me?" The dragon **says** (*ašī*), "Yes, I will get her for you". Then the dragon **says** (*ašī*), "Very well, tonight I will go and wrap myself around the neck of the king's daughter. Then, whoever came, I will not unwrap myself except for you. When you came, I will unwrap myself. Then say, 'Oh king, if you marry off your daughter, give her to me, so that I may open up this dragon'". Well, in the evening the dragon **goes** (*arra*) and **wraps** (*apēčī*) itself around the neck of the king's daughter.

All the verbs in the story line are in the non-past tense. The only verbs found in the past tense are those that introduce the story in the first paragraph, and those that describe background imperfective (ongoing) events (**they have been picking** (*mačenēn boda*) green herbs). Once there is also a verb in the past tense in the subordinate nominal clause after a verb of perception (**he sees** (*agennīt*) that a dragon **came** (*āk*); i.e., 'had come'). In another nominal clause, however, the non-past is used (he **sees** (*agennī*) the rope **is** (*en*) heavy).

---

17 Subject pronouns and other words that have been supplied for syntactic reasons or for the sake of better comprehension are not marked in this section, since the discussion here concerns the verbs.

For tense forms of the verbs, we thus have a deictic shift in Koroshi Balochi and the tense is anchored in the very story, which is portrayed as ongoing at the very moment by means of using the non-past tense. There are, however, three time adverbials in the corpus that modify this picture slightly, since they all contain a distal demonstrative determiner: **that day (***ā rōč***)** (unit 18, Nourzaei et al. 2015: 168), **that day** (*ā rō*) (unit 68, Nourzaei et al. 2015: 185), and **that hour** (*ā sāhat*) (unit 100, Nourzaei et al. 2015: 196).

4.2 Sistani Balochi

The picture that emerges for Sistani Balochi is somewhat different. Here all the verb forms are in the past tense and therefore the story is temporally anchored at the time of narration rather than within the story itself. All verb forms in the narrative parts of the story (direct speech excluded), which is the same story as that in section 3.2, are given below. Only the evidential marker *gušīt* at the very beginning of the story is in the non-past tense, which is logical, since it refers to a non-past event (Barjasteh Delforooz 2010: 378–383). All the other verb forms are in the past tense.

> Sir, they **say** (lit. 'he says') (*gušīt*), **there were** (*atant*) three cows. All these three cows **were** (*atant*) in unity (lit. 'in one heart'), **there was** (*at*) a black one, **there was** (*at*) a light brown one, and **there was** (*at*) a white one. These **were** (*atant*) in unity from that old time. Wherever **they grazed** (*čartant*) no beast of prey **had** (*nadāšt*) any power over them. If **there was** (*būtēn*) a lion, if **there was** (*būtēn*) a leopard, if **there was** (*būtēn*) a wolf, that **attacked** (*alma kurtēn*) one of them, all three **attacked** (*amlaa kurtant*) it and no beast of prey **attacked** (lit. 'went' on') (*našut*) them, because they **were** (*atant*) of one hear and of one mind. What **did** a certain lion **do** (*kurt*)? It **was** (*būt*) stalking them, it **was** (*būt*) lying in ambush for them. It **said** (*guštī*), "Unless I change the mind of each one of them, I won't be able to eat them". This lion **came** (*āt*) and **lied** (*drōg ǰat*) to them, **told them a lie** (*drōgē ǰat*), "O fellows, you are such good friends, I am your fourth brother. I have seen a pasture in a place. It is very green (lit. 'spring'). You . . . let's go there, I will take you there, you eat that grass, I will watch over you", the lion **said** (*gušt*). It **deceived** (lit. 'made donkey') (*xarē kurt*) them and **took** (*burt*) them. When it **took** (*burtē*) them there, it **said** (*guštī*), "Now I will watch over you on this mound, you eat this grass". It **passed** (*būt*), that day **passed** (*būt*) and the next day **passed** (*būt*), it **whispered** (lit. 'entered') (*putrit*) in the light brown cow's ear and in that of the white one. It **said** (*guštī*), "O friend, your hair and my hair are the same colour, that black one is ill-matched (lit. 'unripe') among us. You, don't help it, I will eat it, a lot of grass will remain for you". It **confused** (lit. 'ruined') (*xarābē kurt*) their minds and one of the cows **said** (*guštī*), "Fine". So it **seized** (*čalāpt*) the black one. They (i.e., the other cows) **didn't help** (*kumak nakurtant*). When they **didn't help** (*kumak nakurtant*), it **overpowered** (lit. 'was strong') (*zōr at*) this single one. It **ate** (*wārtē*) this one up. When it **ate** (*wārt*) and **finished it** (*alāsē ku*) . . . , when it **finished it** (*alāsē ku*), it **whispered** (lit. 'sneaked') (*putrit*) in the light brown cow's ear and **said**

(*guštī*), "Your hair and mine are the same colour, you, don't help it, when I eat this white one, all the grass will remain for you. Then I will watch over you here, you can eat!" It **said** (*guštī*), "Fine". When it **seized** (*gipt*) this one too, that one (i.e., the light brown one) **didn't help** (*komak nakurt*). The lion **ate** (*wārt*) this one (i.e., the white one) too. When it **finished it** (*alāsē kurt*) this one too, it **said** (*gu*) to that one, "Now I alone am more power- ful than you". It **ate** (*wārt*) that one too. In this manner, with this trick, it **destroyed** (*ziyānē kurt*) them.

When it comes to time adverbials, unlike tense, they sometimes demonstrate proximal deixis, i.e., anchoring in the story (deictic shift). There are a number of instances of proximal time deictics in the corpus, such as the demonstrative deictics in unit 114 (Barjasteh Delforooz 2010: 343, units 113–119 are translated below) and unit 15b (Barjasteh Delforooz 2010: 358), which refer to time rather than to space:

> From the first day when this girl had been born, until **this very [time]** (*amē*) when she was mature and had reached puberty, except for her mother and father who knew that this was a girl, even the neighbours didn't know that this was a girl.

> ... the Baloch nomad, **at this moment** (lit. 'here') (*idā*) (i.e., at the moment when his camel was dying), experienced the feeling of desperation. ...

Also the time adverbials "today, tomorrow, and the day after tomorrow" are used in these texts in narrative parts to refer to "the same day, the next day, the day after the next" (unit 10, Barjasteh Delforooz 2010: 343; units 42–43, Barjasteh Delforooz 2010: 380; unit 28, Barjasteh Delforooz 2010: 385; unit 50, Barjasteh Delforooz 2010: 387). Units 41–44 (Barjasteh Delforooz 2010: 380–381) read as follows:

> It passed, **that day (**lit. 'today') (*mrōčī*) passed and **the next day** (lit. 'tomorrow') (*bāndā*) passed, it whispered in the light brown cow's ear and in that of the white one.

Another passage that would indicate temporal anchoring in the actual narration is units 72–73 (Barjasteh Delforooz 2010: 299) where the word *šapī* is translated "tonight". However, in view of the fact that a parallel adverb *sōbī* has been translated "in the morning" (unit 119, Barjasteh Delforooz 2010: 374; unit 72, Barjasteh Delforooz 2010: 388), it seems likely that *šapī* can also be translated as "at night"[18]: "Surely I will die tonight (or 'at night') (*šapī*). Death did not come to him tonight (or 'at night') (*šapī*). ..".

---

18 This translation, or the translation 'this night' was also suggested by one of my Baloch friends.

4.3 Vafsi

The verb forms in Vafsi exhibit yet another pattern. Most stories start out in the past tense, not only for the initial background description, but also for foreground events in the event line at the beginning of the story. In Tale B7 (Stilo 2004: 110–113), given in full above, there is one tense shift to the non-past tense, which takes place between units 24 and 27.[19] In Tale A1 (: 26–29), Tale A3 (Stilo 2004: 32–47), Tale A4 (Stilo 2004: 48–51), Tale A9 (Stilo 2004: 104–109), and Tale B8 (Stilo 2004: 116–123), there are several shifts between past and non-past tense. All the tales start out in the past and end in the non-past tense, except for A1, which ends in the past tense. Tale A5 (Stilo 2004: 52–53) is in the past tense from the beginning until the end. This is also true for Tale A6 (Stilo 2004: 54–57). The point where tense shift in Tale B7 occurs is presented here:

> The infidel **considered** (*molazæs bækærde*) it for a bit and **said** (*va*), "Well, I didn't have a child until now. And now, God has willed it (lit. 'it pleases God'). Now, let me slit his throat so they can get apart from one another, I don't want this child any more". This infidel **goes** (*ætari*) and **gets** (*ærgiri*) a pen knife and **slits** (*ærbirine*) his son's throat, right between Moses and the hermit, and **they come apart** (*joda -rbuænd*) from one another and **start going** (*bæna -rkærende siæn*).

In Vafsi, the same word is used for 'tomorrow' and 'the next day' (*soæy*) (Stilo 2004: 277). In Text A5 (unit 9, Stilo 2004: 52), a proximal demonstrative determiner is even added before *soæy*. An enclitic pronoun (=*s*) is added to the phrase *soæy šo* 'tomorrow night' to denote 'the next night' (unit 91, Stilo 2004: 40):

> **The next day** (lit. 'this tomorrow') (*in soæy*) the Jew came …

> Then it turned into **the next night** (lit. 'it's tomorrow night') (*soæy šos*). On **the next night** (lit. 'it's tomorrow night') (*soæy šos*), this vizier got up again.

It is also interesting to note the proximal demonstrative determiner in the temporal expression "at first" in Tale A9 (unit 31, Stilo 2004: 106):

> **At first** (lit. 'this first') (*in ævvæl*) they said no.

4.4 Gorani

In the three Gorani tales, when direct speech is excluded, the verbs are almost invariably in the non-past tense both in the introduction and in the event line. One or two past tense verb forms are, however, used for background informa-

---

19 Units 25–26 are direct speech.

tion, e.g., in a subordinate adverbial clause (unit 13, Mahmoudveysi et al. 2012: 64) and in a relative clause (unit 64, Mahmoudveysi et al. 2012: 100). Units 2–8 of the tale *Tītīla and Bībīla* (Mahmoudveysi et al. 2012: 63–64), where both the introduction and verbs in the event line are found, read like this:

> Tītīla and Bībīla, in the Kurdish language we **say** (*mwāžām*), "The lame goat and the lame ram". **There is** (*mawu*) a goat and **there is** (*maw(u)*) a ram; they **are** (*mawin*) lame. They **are** (*mawin*) at home; the flock **goes** (*mašu*) to the mountains and they (i.e., the goat and the ram) **stay** (*mamanin*) at home. They . . . The front of the gate **is** (*mawu*) open; they **go** (*mařawin*) out of the courtyard and they say, "Let's go, let's reach the flock". **They go** (*mašin*) to the Zimkān river; the flock **goes** (*mašu*) to that other side. From the side of the Zimkān river, the flock **crosses** (lit. 'does') (*makarī*) to that other side. The flock **crosses** (lit. 'does') (*makarī*) to that side. . . . Suddenly they (i.e., the goat and the ram) **say** (*mwān*), "Hey, brother!" The ram **says** (*mwāy*), "Yes?"

When it comes to time adverbials, Mahmoudveysi et al. (2012: 81 n. 43) note that the difference between "that/this evening" and "the following evening/night" is *īšaw* versus *ī šaw*. What is relevant to this discussion is that according to this definition, there is no difference between "that evening" and "this evening". It is also interesting to note that the phrase "the following evening/night" contains the proximal demonstrative determiner (*ī*).

5 Summary and comparison with other Iranian languages

The four linguistic variants in this study show interesting variation when it comes to deictic shift (see Table 12). First of all, it is more common for spatial deixis to be shifted to the story than for tense to be anchored in the story. Koroshi Balochi, Sistani Balochi, and Vafsi present almost total spatial deictic shift, whereas in Gorani there are a few cases of the deictic adverb 'there (*āna*, *ānā*)', which move the deixis outside the story. It also seems that proximal deixis is the norm in at least Koroshi Balochi, Sistani Balochi, and Vafsi to repeatedly mark the current center of attention with deictics.[20]

Gorani is the language that has the strongest tense anchoring inside the narrative, with almost exclusive use of the non-past tense. At the other extreme

---

20 Thomas Jügel (personal communication) also notes the same phenomenon in German, and comments that distal pronouns are rarely used, their function being taken over by proximal pronouns and elements that are neutral with respect to deixis.

we find Sistani Balochi, which has no tense anchoring in the narrative (only past tense verb forms), but does have some adverbials that move the deictic center into the story. Vafsi is ambivalent, with constant shifts between the past and non-past tenses. Normally the tales start out in the past tense and end in the non-past tense. As for Koroshi Balochi, tense use is consistent with background material in the past tense and foreground material in the event line in the non-past tense. On the other hand, there are a few temporal adverbials with distal demonstrative determiners in Koroshi Balochi.

Table 12: Summary of spatial and temporal deixis in the four variants under study

Language	Spatial deixis inside narrative	Tense anchoring inside narrative[21]
Koroshi Balochi	+	+
Sistani Balochi	+	−
Vafsi	+	+/−
Gorani	+/−	+

Studying deixis in Persian, Roberts (2009: 233) argues for "a bias or preference for proximal deixis over distal deixis". Persian prefers, as an example, to use the proximal deictics 'now, here, this, yesterday, today, tomorrow etc., both with the speech moment as the deictic center (he will come tomorrow) and with another deictic center than the speech moment (he told me that he would come the next day (lit. 'tomorrow') (Roberts 2009: 235–241). Further studies are needed to determine to what extent the preference for proximal deixis in Persian should be attributed to deictic shift in this language.

Deictic shift has also been reported for Talyshi folktales by Paul (2011: 93), who shows that the deictic center sometimes "is projected onto the narrative's chief protagonists" and sometimes "onto the central locational reference point" of a certain episode in the narrative. Barjasteh Delforooz (2010: 146) reports the same phenomenon for Sistani Balochi. He finds that the spatial deictic center in the stories he analyzes in depth is normally where the main participant of the story is located, but that the deictic center sometimes moves ahead of the major participant to the scene where an important event or the climax of the story will take place. The point is that it stays inside the story, which means that we are dealing with deictic shift. Barjasteh Delforooz concludes that in his data, proximal deixis "is much more frequent than distal deixis" and that in fact it is even more frequent in Sistani Balochi than in Persian (Barjasteh Delforooz 2010: 159).

---

21 Here only the verbs in the event line are considered.

The distal demonstratives are mainly used for establishing a contrast between two different entities, for highlighting purposes, and for reactivation. The use of a distal demonstrative pronoun to establish contrast is also noted in Kumzari by Wal Anonby (2015: 67), who finds that "a secondary participant is referenced by the pronoun *ān* [...] in place of third person singular *yē*, to distinguish it from a primary participant".

Other features of the stories under study also contribute to the deictic shift of the listener from the present world to the world of the narrative. One of these is the total predominance of direct speech. Now and then skillful narrators also address the audience with phrases such as "listen brothers" (Barjasteh Delforooz 2010: 303, unit 1), "sir" (Nourzaei et al. 2015: 123, unit 2), or an even more personal address such as "dear doctor", as found in Barjasteh Delforooz (2010: 391, unit 115).

It is thus clear that there are numerous devices to make the oral narration lively and exciting by bringing the distant near, thereby giving the audience a fascinating experience of journeying to "a time and a land that is not".[22]

6 Acknowledgments

References

Anderson, Stephen R. & Edward Keenan. 1985. Deixis. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. III: 259–308. Cambridge: Cambridge University Press.

Barjasteh Delforooz, Behrooz. 2010. *Discourse features in Balochi of Sistan. Oral narratives* [Studia Iranica Upsaliensia 15]. Revised version. Uppsala: Acta Universitatis Upsaliensis. http://uu.diva-portal.org/smash/record.jsf?pid=diva2:345413 (accessed 10 December 2015).

Diessel, Holger. 1999. *Demonstratives. Form, function, and grammaticalization.* Amsterdam and Philadelphia: John Benjamins.

Fillmore, Charles J. 1997. *Lectures on deixis.* Stanford, CA: CSLI Publications.

Haig, Geoffrey L. J. 2008. *Alignment change in Iranian languages. A constructional grammar approach.* Berlin & New York: Mouton de Gruyter.

22 See www.poemhunter.com/poem/the-land-that-is-not/ for a poem with the English title *The Land That Is Not*, originally written in Swedish by the poet Edith Södergran.

Jahani, Carina & Agnes Korn. 2009. Balochi. In Gernot Windfuhr (ed.), *The Iranian languages*, 634–692. London & New York: Routledge.

Korn, Agnes. 2005. *Towards a historical grammar of Balochi* [Beiträge zur Iranistik, 26]. Wiesbaden: Reichert.

Levinsohn, Stephen H. 2015. *Self-instruction materials on narrative discourse analysis*. SIL International. http://www-01.sil.org/~levinsohns/narr.pdf (accessed 23 November 2015).

Lyons, John. 1977. *Semantics*. Vol. 2. Cambridge: Cambridge University Press.

Mahmoudveysi, Parvin, Denise Bailey, Ludwig Paul & Geoffrey Haig. 2012. *The Gorani language of Gawraǰū, a village of west Iran* [Beiträge zur Iranistik, 35]. Wiesbaden: Reichert.

Nourzaei, Maryam, Carina Jahani, Erik Anonby & Abbas Ali Ahangar. 2015. *Koroshi. A corpus-based grammatical description* [Studia Iranica Upsaliensia, 13]. Uppsala: Acta Universitatis Upsaliensis. http://uu.diva-portal.org/smash/record.jsf?pid=diva2%3A810250 (accessed 10 January 2016).

Nyberg, H. S. 2004. *Muntlig tradition, skriftlig fixering och författarskap*. Compiled and edited by Bo Utas. Uppsala: Kungl. Humanistiska Vetenskaps-Samfundet i Uppsala.

Ong, Walter. 1982. *Orality and literacy. The technologization of the word*. London & New York: Methuen.

Paul, Daniel. 2011. A glance at the deixis of nominal demonstratives in Iranian Taleshi. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in Iranian linguistics*, 89–102 [Beiträge zur Iranistik, 34]. Wiesbaden: Reichert.

Paul, Ludwig. 2003. The position of Balochi among western Iranian languages: The verbal system. In Carina Jahani & Agnes Korn (eds.), *The Baloch and their neighbours: Ethnic and linguistic contact in Balochistan in historical and modern times*, 61–71. Wiesbaden: Reichert.

Roberts, John R. 2009. *A study of Persian discourse structure* [Studia Iranica Upsaliensia, 12]. Uppsala: Acta Universitatis Upsaliensis. http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A285628 (accessed 28 October 2015).

Segal, Erwin M. 1995. A cognitive-phenomenological theory of fictional narrative. In Judith F. Duchan, Gail A. Bruder & Lynne E. Hewitt (eds.), *Deixis in narrative: A cognitive science perspective*, 61–78. Hillsdale, NJ, & Hove, UK: Lawrence Erlbaum Associates.

Stilo, Donald L. 2004. *Vafsi folk tales*. Supplied with folklorist notes and edited by Ulrich Marzolph [Beiträge zur Iranistik, 25]. Wiesbaden: Reichert.

Utas, Bo. 2008. *Manuscript, text and literature. Collected essays on Middle and New Persian texts.* [Beiträge zur Iranistik, 29]. Edited by Carina Jahani and Dariush Kargar. Wiesbaden: Reichert.

Wal Anonby, Christina van der. 2015. *A grammar of Kuzmari. A mixed Perso-Arabian language of Oman*. Leiden: Leiden University dissertation.

Zubin, David A. & Lynne E. Hewitt. 1995. The deictic center: A theory of deixis in narrative. In Judith F. Duchan, Gail A. Bruder & Lynne E. Hewitt (eds.), *Deixis in narrative: A cognitive science perspective*, 129–155. Hillsdale, NJ, & Hove, UK: Lawrence Erlbaum Associates.

Katarzyna Marszalek-Kowalewska

17 Extracting semantic similarity from Persian texts

Abstract: The study of semantics has always attracted researchers from the fields of philosophy, linguistics, and communication theory. Recent developments in corpus linguistics and in computational linguistics have enabled an empirical study of semantics on a large scale. The goal of this article is to present procedures for extracting semantic information from a corpus of Persian newspaper texts. The analysis – modeled on Mason's (2006) study – focuses on applying the procedures for identifying semantic information to the Persian language. First, two procedures – *collocational overlap* and *usage patterns* – are introduced. Next, a comparative study of English loanwords in Persian, and their native Persian counterparts is presented. The aim of this study is to examine whether these two – the loanword and its native counterpart approved by the Academy of Language and Literature – are semantically similar.

1 Introduction

Semantics, the study of meaning, has been one of the core concepts and is widely discussed in philosophy, linguistics, and information theory. John Firth (1957: 190), in emphasizing the importance of semantics, wrote "[t]he study of meaning is a permanent interest of scholarship". This article tries to deal with semantics from a linguistic perspective. Since the study of meaning involves both linguistic and nonlinguistic worlds, it has been rather neglected (not to say ignored) in linguistic analyses for a long time. Linguists have claimed, for instance, that it is not possible to analyze semantics empirically, e.g., "word meanings are not among the phenomena which can be covered by empirical, predicative scientific theories" (Sampson 2001: 206).

Although some of the critics were right in their beliefs, e.g., that meaning changes and is inseparably connected with speakers, the idea that meaning should not be analyzed empirically seems rather unconvincing, especially nowadays. Recent advances in corpus and in computational linguistics, and particularly in

Katarzyna Marszalek-Kowalewska, Adam Mickiewicz University

natural language processing, make it possible to not only analyze semantics empirically, but to conduct those analyses on a scale that was not possible before. To name just a few examples, let's consider: Brown et al.'s (1991) study on word sense disambiguation; Fellbaum's (1998) work on WordNet; Magnini and Cavigliá's (2000) proposal on integrating subject field codes into WordNet; Danielsson's (2001) study of the automatic identification of meaningful units in language; and Girju, Badulescu, and Moldovan's (2003) research on semantic constraints in the automatic discovery of part-whole relations.

This article presents a study of the process for identifying certain semantic information in the Persian language and, to be more precise, it focuses on extracting semantic similarities. It presents two procedures for extracting semantic information presented by Mason (2006). Mason's work focuses on the English language, but here his procedures are applied to Persian. Firstly, two introductory sections are presented. The former describes general notions about corpus and about computational linguistics, with respect to a semantic analysis. The latter section focuses on Persian corpus linguistics and its contribution to the study of meaning. Then the two procedures mentioned above – *collocational overlap* and *usage patterns* – are introduced. The methodology and corpus details are also provided, and a comparative study of English loanwords in Persian, along with their native counterparts approved by the Academy of Persian Language and Literature (Farhangestāne Zabān va Adabiyāte Fārsi), is presented. Finally, this article concludes with a discussion of the results.

2 Extracting semantic information from a corpus

Since this article deals with extracting semantic information from a corpus of Persian texts, certain terms connected to corpus and computational linguistics need to be defined. A corpus is usually defined as a systematic collection of naturally occurring texts. The field of corpus linguistics therefore enables the analysis of naturally occurring texts, on the basis of an analysis often carried out with the use of specialized software. Computational linguistics deals with the statistical or rule-based modeling of natural language from a computational perspective. Finally, the area of information retrieval (IR) can be briefly described as the science of finding objects in any media relevant to a user query.

The literature and the examples of corpus-based semantic information retrieval are enormous. First of all, a corpus can be annotated with semantic information. Although adding semantic metadata to a corpus is a complex task, nowadays more and more corpora provide semantic information, including: CLEF Corpus (Roberts et al. 2007), GENIA (Kim et al. 2003), PropBank (Palmer, Gildea, and

Kingsbury 2005), FrameNet (Baker, Fillmore, and Lowe 1998), Penn Discourse TreeBank (Prasad et al. 2008), and OntoNotes (Hovy et al. 2006).

Computational semantics is the area of computational linguistics that focuses on linguistic meaning within a computational approach to the study of a natural language. It focuses on an approach where the meaning of words, phrases, and sentences can be computed systematically from the meaning of their syntactic constituents. This subdiscipline of computational linguistics is currently gaining more and more interest. Within the Association of Computational Linguistics, there is a Special Interest Group on Computational Semantics (SIGSEM), which hosts three conferences on computational semantics: International Workshop on Computational Semantics, *SEM Joint Conference on Lexical and Computational Semantics and Inference in Computational Semantics.[1] Another example demonstrating the popularity of computational semantics is FraCaS,[2] the Framework for Computational Semantics, which is a European Union project that deals with convergences in computational semantics. There are also numerous semantic web projects, tools, and platforms that deal with semantic data management. To name just a few, these are: Semantic Web, DataVersity, PreDose, Semantic Bookmarking Platform, GoNTogle, the Knowledge and Information Management Platform, SemTag, OntoMat, MnM,[3] etc.

One of the main foci in computational semantics is word-sense disambiguation (WSD). This is a process for identifying the meanings of a word in a manner that is determined by its context. In the area of computational linguistics and natural language processing, it becomes clear that semantic disambiguation at a lexical level is necessary for even a basic understanding of language (Stevenson and Wilks 2003). Some early trials to deal with the problem of ambiguity were undertaken by Wilks (1972), with his preference semantic systems; Small (1980) with word expert parsing; and Hirst (1987) with polaroid words. In the dictionary-based approaches, researchers have made use of machine-readable dictionaries for word-sense disambiguation, e.g., Lesk (1986) and McRoy (1992). Another way of dealing with the problem has been the connectionist approach. Examples of this include: Waltz and Pollack (1985), who independently developed a model based on psycholinguistic observations; and Ide and Veronis (1980), who constructed large neural networks in order to solve the problem of language

---

1 For more information on the activities of SIGSEM, see http://www.sigsem.org/w/.

2 For more information, see http://www.lt-world.org/kb/players-and-teams/projects/obj_62590.

3 For more information on the semantic web and semantic web platforms, see Reeve and Hyoil (2005) *Survey on Semantic Annotation Platforms*, and Kiryakov et al. (2004) *Semantic annotation, indexing and retrieval*.

ambiguity. Currently, most approaches to word-sense disambiguation use statistical and machine learning approaches. These include Yarowsky (1992) and Schütze (1992).

In the area of information retrieval, information is stored in various thesauri and ontologies. The most popular database that allows the retrieval of semantic information is WordNet (Fellbaum 1998), a large lexical database of the English language, in which nouns, verbs, adjectives, and adverbs are grouped into synsets (sets of cognitive synonyms).[4]

The semantic information extracted via natural language processing, as well as corpus and computational linguistic approaches, are useful in many areas, e.g., text summarization, advertisement targeting, biomedical text mining, etc. This very general description of the methods for identifying semantic information in the fields of corpus linguistics, computational linguistics, and natural language processing aims to provide a background for the possibilities of finding semantic information in texts. For a more detailed description, please refer to *The Oxford handbook of computational linguistics* (Mitkow 2003), the *Handbook of natural language processing* (Dale, Moisl, and Somers 2000), and *Word sense disambiguation* (Eneko and Edmonds 2007).

3 Persian corpus linguistics and semantic information retrieval

The Persian (also known as Farsi) language belongs to the Indo-Iranian branch of the Indo-European language family, and is spoken by more than 100 million people throughout the world. It is the official language of Iran, Afghanistan, and Tajikistan. Despite the large number of speakers, Persian has, unfortunately, been among the less resourced and least analyzed languages from a computational point of view. Although there are more and more complex developments in the areas of Persian natural language processing, corpora, and computational linguistics, it is still not as fully researched as, for example, English. One reason for this may be the fact that the Persian language poses certain problems for computational and corpus linguists (Megerdoomian 2010; Seraji, Megyesi, and Nivre 2012; Shamsfard 2011; Ghayoomi, Momtazi, and Bijankhan 2010; Ghayhoomi and Momtazi 2009; Farajian 2011). Some of the main challenges in Persian text processing are:

---

4 For more information on WordNet, see https://wordnet.princeton.edu/.

(a) Unclear word boundaries and resulting unclear phrase boundaries: One of the most problematic issues in processing texts in Farsi is the fact that word boundaries are often not clear. First of all, there is the problem of pseudospaces. Pseudospaces or zero-width non-joiner (ZWNJ, \u200c in Unicode) indicate boundaries of words or compound parts. Users often use white spaces instead of pseudospaces, which causes words such as زبان شناسی *zabānshenāsi* 'linguistics' to be processed as the separate words زبان *zabān* 'language' and شناسی *shenāsi* 'knowledge', instead of as the single word 'linguistics'. What is more, white spaces are not used when they should be, e.g., توازماگرفت *toazmāgereft* 'YouTookFromUs'. As a result this phrase would be processed as one lexeme instead of as four separate items. Finally, the inconsistency in the use of white spaces in terms with detached morphemes creates another challenge. The number of possibilities resulting from such situations can be observed in the following table:

Table 1: Word boundary ambiguity (adapted from Ghayoomi, Momtazi, and Bijankhan 2010; examples mine)

Affix	Attached	Pseudospace	White space
می	میخرید	می‌خرید	می خرید
ترین	کوچکترین	کوچک‌ترین	کوچک ترین
ها	زنها	زن‌ها	زن ها

(b) Diacritics: The following diacritic marks are used in Farsi:
- *zabar* (*fathe*) – used to mark the vowel [a] after a consonant, e.g., دَر *dar* 'door';
- *zir* (*kasre*) – used to mark the vowel [e] after a consonant, e.g., گِل *gel* 'mud';
- *piš* (*zamme*) – used to mark the vowel [o] after a consonant, e.g., سُر *sor* 'slide';
- *sokun* – used to mark that the consonant is not followed by any vowel, e.g., گَنْج *ganj* 'treasure';
- *tašdid* – indicates a double consonant, e.g., رَدّ *radd* 'dismissal';
- *tanvin* – used on a final alef to mark the final [an] of certain adverbs بعداً *ba'dan* 'later',
- *hamze* – used to mark ezafe construction after the vowel [e], e.g., خانهٔ من *xāne-ye man* 'my house'.

The fact that diacritics are usually not written poses several problems. To begin with, this can cause a lot of ambiguity, particularly in the case of homographic lexemes. As a result, it blurs the statistics and frequencies, as after tokenization, these words are counted as one, whereas the real frequency of each of them is difficult to establish.

(c) The Ezafe construction. The Ezafe marker is a short vowel added between prepositions, adjectives, and nouns in a phrase in order to determine the relation between nouns and their modifiers, e.g., کتاب من *ketāb-e man* 'my book'. The Ezafe marker is always pronounced but whether it is written is arbitrary. When not written, it can lead to ambiguity. The ambiguity here refers to problems in chunking, as well as to the semantic and syntactic processing of a sentence.

(d) Complex tokens. This category refers to multi-element lexemes that consist of a lexeme itself and an attached part that represents a separate lexical category or a part of speech from the one that it is attached to. The number of possibilities resulting from such a situation can be observed in the following table:

Table 2: Complex tokens (adapted from Ghayoomi, Momtazi, and Bijankhan 2010)

Word	Type	White space	Pseudospace	Attached
به	Preposition	به شیوه	به‌شیوه	بشیوه
هم	Prefix	هم کلاس	هم‌کلاس	همکلاس
این	Determiner	این مرد	این‌مرد	اینمرد
آن	Determiner	آن قدر	آن‌قدر	آنقدر
را	Postposition	شرایط را	شرایطرا	شرایطرا
که	Relativizer	چنان که	چنان‌که	چنانکه

(e) Encoding. Since texts in Farsi are written in an Arabic script, some online materials are written in mixed Arabic and Persian codes, i.e., as well as the Unicode characters for Farsi, Arabic or ASCII characters tend to be used as well. To be more precise, letters ک and ی can be expressed in Persian (\u06a9 for ک and \u064a for ی) or Arabic encoding (\u0643 for ک and \u06cc or \0649 for ی).

Despite the problems presented here, linguists dealing with the Persian language have undertaken steps to provide corpora, as well as tools for analyzing the Persian language. There is a great variety of Persian corpora. In the area of monolingual corpora, the following examples should be cited: The Persian Linguistic Database,[5] prepared under the supervision of Professor Seyyed Mostafa Assi at the Institute for Humanities and Cultural Studies; the Hamshahri Collection,[6] with data from *Hamshahri*, the first online newspaper in Iran; The Peykareh Text Corpus (Bijankhan et al. 2011), designed and developed by Bijankhan; and finally the Bijankhan Corpus,[7] the first linguistically annotated Persian corpus.

---

5 For more information, see http://pldb.ihcs.ac.ir/.

6 For more information, see http://ece.ut.ac.ir/dbrg/hamshahri/.

7 For more information, see http://ece.ut.ac.ir/dbrg/bijankhan/.

There are also a few examples of multilingual corpora. The Comparative Persian-English corpus (Hashemi et al. 2010), compiled by the Intelligent Systems Research Laboratory of Tehran University, consists of two news sources: Persian news from the Hamshahri News agency and English news from the BBC news agency. Other examples are the English-Persian Parallel Corpus (Farajian 2011), which belongs to the ELRA project, and the Persian 1984 corpus,[8] which consists of a translation of George Orwell's novel *1984*, and belongs to the MULTEX-East parallel corpus. The Shiraz corpus is a bilingual parallel-tagged corpus developed from a large Persian corpus of online material. There are also syntactically annotated corpora like: The Persian Treebank[9] (Per TreeBank); The Persian Dependency Treebank[10] (PerDT); and The Uppsala Persian Dependency Treebank.[11] A few corpora have also been compiled and used for speech recognition, such as the Farsi Speech Database[12] (FARSDAT), which was built in 1996 by the Research Centre of Intelligence Signal Processing (Ghayoomi, Momtazi, and Bijankhan 2010). Another speech corpus, the large Persian Speech Database, contains data on over 1,000 hours of speech, which was recorded by one hundred native Persian speakers representing ten different dialects. There are a few corpora focusing on the linguistics of telephone calls, in particular the Linguistic Data Consortium, which offers two telephone speech corpora – OGI Multilingual Corpus containing 175 calls; and CALLFRIEND containing 109 calls. Another corpus, the Persian Telephone Database, comprises one hundred hours of dialogue between 200 native speakers of Persian (Ghayoomi, Momtazi, and Bijankhan 2010).

In order to facilitate studies of the Persian language, certain computational tools have also been developed, including: Persian LG (Dehdari and Lonsdale 2008), PerStem (Jadidinejad, Mahmoudi, and Dehdari 2010), PersPred (Samvelian and Faghiri 2013), PerLex (Sagot et al. 2011), SteP-1 (Shamsfard, Jafari, and Ilbeygi 2010), etc.

There have also been developments in the semantic resources available in Persian Natural Language Processing research. One such work is undoubtedly FarsNet,[13] a lexical ontology for the Persian language. FarsNet aims to provide lexical, syntactic, and semantic information on words and phrases organized in sets of cognitive relations. It consists of the Persian WordNet and Persian Net of

---

8 For more information, see http://catalog.elra.info/product_info.php?products_id=1124.

9 For more information, see https://hpsg.fu-berlin.de/~ghayoomi/PTB.html.

10 For more information, see http://dadegan.ir/en/perdt.

11 For more information, see http://stp.lingfil.uu.se/~mojgan/UPDT.html.

12 For more information, see http://catalog.elra.info/product_info.php?products_id=18.

13 For more information, see http://dadegan.ir/catalog/farsnet.

Verb (PeVNet) frames and includes two types of connections: inner- and inter-language relations. The former consists of the relations between different senses and synsets, e.g., synonymy, hyperonymy, hyponymy, meronymy, cause, and antonymy. The latter consists of equal-to and near-equal-to relations. The following lexical resources were used to construct FarsNet: two monolingual dictionaries (the Sokhan Dictionary and the Sadir Afshar Dictionary, two corpora (the Persian Linguistic Database and Peykareh), bilingual dictionaries, the Persian Thesaurus, and a dictionary of Persian synonyms and antonyms (Shamsfard, Fadaei, and Fekri 2010). Numerous research studies have been conducted on the FarsNet framework, including: Dehkharghani and Shamsharf (2009) on the mapping of Persian words to the WordNet synsets; Shamsharf et al. (2010) on extracting lexico-conceptual knowledge for the development of the Persian WordNet; Rouhizadeh, Yarmohammadi, and Shamsfard (2008) on WordNet for Persian verbs; Taheri and Shamsfard (2011) on mapping FarsNet to suggest an upper merged ontology, etc.

There is also serious work being published that relates to word-sense disambiguation in the Persian language. As examples, consider: Soltani (2010) on statistical word-sense disambiguation in Persian; Hamidi, Borji, and Ghidary (2007) on Persian word-sense disambiguation; Makki and Homayounpour (2008) on the word-sense disambiguation of Persian homographs; Miangah and Khalafi (2005) on the usage of a learner corpus in a machine translation system; and Makki and Homayounpour (2008) on using a decision list for Farsi word-sense disambiguation.

4 The analysis of semantic similarity: A corpus-based study

Similarity is a complex concept widely discussed in philosophy, linguistics, and information theory. So far, many different measures of semantic similarity between word pairs have been proposed. Some use lexical databases, some statistical or distributional techniques, and some a hybrid approach, which combines both statistical and lexical techniques. This section presents two methods of measuring semantic similarity proposed by Mason (2006) – "collocational overlap" and "usage patterns". Then a case study on semantic similarity between English loanwords in Persian and their native Persian equivalents will be presented. Yet, before describing the procedure and case study, let's first focus on the data that underwent scrutiny.

In order to identify semantic information, a new corpus of Persian modern newspaper data was compiled. The compilation process consisted of collecting news articles from popular Iranian online newspapers (*MehrNews* and *Hamshahri*), which were preprocessed (e.g., removing links, dates, etc.) and prepared for linguistic analysis. It consists of 10 million tokens and in terms of features, it is as follows:

- specialized: it contains data from one source: newspaper texts
- synchronic: it contains data from 2010
- written: it contains written texts
- monolingual: the collected data concern the Persian language
- full text: it contains full samples
- static: it is meant to represent the language at a particular point in time

To carry out the analysis, *WordSmith*[14] 6.0 tool was used. It is a complex program for corpus analysis prepared by Mike Scott. It allows extraction of word lists, keyword lists, collocations; analysis of concordances; and provision of numerous cluster extraction functions, as well as some statistical measures.

4.1 Collocational overlap

Collocation is one of the core concepts in linguistics and can be described with Firth's (1957: 179) words as the "company a word keeps". To be more precise, a collocation is a sequence of words that tend to co-occur with one another more often than it might be expected. Collocational overlap is based on the idea that "collocations are the words which occur within the context of a node word more significantly than expected by chance. Therefore collocates of a word describe its context in a condensed way, and through analysing the distribution of collocates across the environment of a number of words we can assess their semantic similarity" (Mason 2006: 228). The whole procedure starts with a choice of words[15] to be investigated, i.e., the node word and candidate words hypothetically chosen as similar ones. Then collocates of the node and candidates are examined and the overlap of collocates is analyzed. This overlap was determined by the following formula:

$$d(i,j) = 1 - \frac{2 \times (ci \cap cj)}{ci + cj}$$

---

14 For more information, see http://www.lexically.net/wordsmith/.

15 It would also be possible to conduct the analysis without preidentified words using natural language processing techniques.

i.e., the distance between two words i and j equals twice the number of shared collocates divided by the sum of the sizes of each word's collocates. This is subtracted from 1 in order to yield a distance value, with 1.0 being the maximum distance, and 0.0 being identity. On the basis of data analysis (Mason 2006: 229), it was concluded that words having overlap distance less than 0.8 are classified as similar, i.e., two words will be classified as semantically similar when the overlap distance between them is less than 0.8. If the overlap distance is more than 0.8, two words are not classified as semantically similar. As an example, consider رایانه *rāyāne* 'computer' and its collocational overlap with the following words:

Table 3: Collocational overlap

	Distance between رایانه **and**	Qualified as similar (YES < 0.8 > NO
کامپیوتر 'computer' (English loanword)	0.5	YES
ماشین 'machine'	0.36	YES
فنی 'technological'	0.79	YES
اسباب 'device'	0.87	NO

Words that qualify as similar are then collected in one set. Another set is made from all collocates these candidates have. Then for each candidate, the co-occurrence frequencies of each word from the collocates set are counted. The counts (as well as the n) are entered into a contingency table that is then analyzed using Correspondence Analysis (for more details, see Mason [2006: 230]).

One very important aspect that needs to be remembered here is that the result will depend on the corpus used in the analysis, e.g., whether it is a reference corpus (designed to provide a comprehensive reflection of a language) or a specialized (compiled to reflect a very specific part of a language) one. What is more, the test used for extracting collocates also leads to different results (for example, the Mutual Information test, which was used here, measures the strength of association between words whereas t-score measures the confidence with which we can assert that there is an association).

4.2 Usage patterns

This method is based on the assumption that similar entities have more properties in common. The starting point here is the ready set of triples – the output of grammatical usage pattern extraction – two words and the relation between

them. For each of these words there is a set of significant arguments – words that have similar sets of partners in one or more relationships. These particular arguments are responsible for defining the *shared environment* between the investigated words. What is more important here, these words provide a link to those words that themselves have one or more of those arguments, which are classified as semantically related candidates (Mason 2006). The set of arguments of each candidate word is prepared, and later it is compared with a set of arguments of a target word. We may illustrate this with the example of the NN relation of کامپیوتر *kāmpyuter* 'computer' (English loanword). The results are آزمایش *āzmāyesh* 'test', خواص *khavās* 'properties', based on the argument مدل *modl* 'model'. Another set of substitutes would be اپل *apl* 'apple' and اطلاعات *etelāat* 'information' based on the argument سیستم *sistem* 'system'.

The grammatical relations provide information on lexical co-occurrence within syntactic relations. The set of co-occurrence of a word within syntactic relations provides a strong restriction of its semantic properties.

4.3 Comparative analysis of English loanwords and their Persian counterparts

The application of measuring semantic similarity will be presented as a comparative study between English loanwords and their Persian equivalents. Thus, the study focuses on English loanwords present in Persian and their Persian counterparts approved by the Academy of Language and Literature. The reason
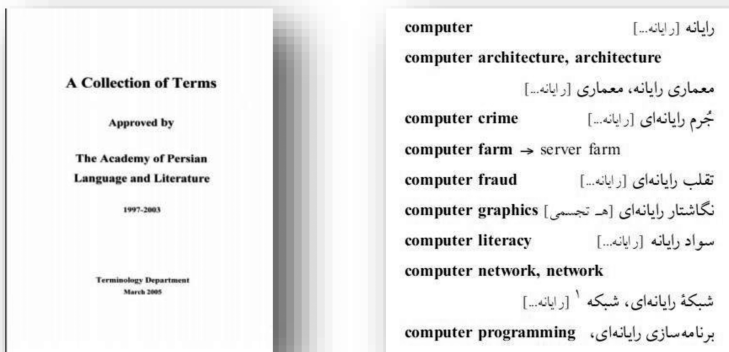
Figure 1: A sample of A Collection of Terms Approved by the Academy of Persian Language and Literature.

behind such an analysis was the Iranian language policy toward loanwords, which can be characterized by linguistic purism. The beginnings of Iranian language policy can be traced back to the tenth century and the times of Abu Ibn Sina. However, since this thesis deals with the modern Persian language, we will describe briefly only the most recent period of the purist movement (for more on Iranian language policy, see Marszałek-Kowalewska [2010]). In 1991, the Third Supreme Council of the Iranian Revolution established the Farhangestāne Zabān va Adabiyāte Fārsi (Persian Academy of Language and Literature).

The academy members, language experts, and professors are responsible for studying grammar, orthography, manuscripts, and various Iranian dialects. The policy of the Third Academy is as follows:

- In coining and choosing a new word, Persian phonetic rules and learned speakers' way of talking and Islamic points of views should be regarded as the main criterion.
- Phonetic rules should be followed according to the Persian way of talking.
- New words should follow the Persian grammatical rules for coining nouns, adjectives, verbs, and so on.
- New words should be chosen or coined out of the most common or frequent words that have been used since AD 250.
- New words can be chosen from among the most frequent and common Arabic words as used in Persian.
- New words can be chosen from the Middle and Old Persian stages of the language.
- There should be only one equivalent in Persian for any of the Latin words, particularly technical ones.
- It is not necessary to adapt or create new Persian words for those Latin words that have been used internationally and globally (Farhangestan-e Zaban [2001] as quoted in Monajemi 2011: 5).

Therefore, the decision was taken to compare English loanwords and their Farsi counterparts. The following section presents the research procedure on the example of COMPUTER.

Both an English loanword کامپیوتر and its Persian equivalent رایانه are used by native speakers. Firstly, let's consider the raw frequency in the corpus. As can be seen, the native word is much more frequent; it appears 799 times in the corpus while the English loanword appears only eighty times.

Table 4: Raw frequency of loanwords in a corpus

English loanword frequency	Persian counterpart frequency
کامپیوتر	رایانه
80	799

The first part of the analysis consists of comparing n-grams. N-gram is a sequence of words that appear consecutively in the text. It can consist of two words (bi-gram), three words (tri-gram), four words (4-grams), etc. They are easy to formulate and extract and they provide useful reflections of lexical, semantic, and syntactic relations. The following tables present n-grams frequency, loan-word n-grams, and native word n-grams, respectively. There was a condition that as n-gram we define every sequence of words that appeared a minimum of five times in a corpus. There were eighteen loanword n-grams and fifty-eight native n-grams:

Table 5: N-gram frequency

English loanword n-gram number (minimal n-gram frequency=5)	Farsi counterpart n-gram number (minimal n-gram frequency=5)
18	58

Table 6: Loanword n-grams

N	Translation	N-gram	Freq.	Length
1	*computer and*	کامپیوتر و	31	2
2	*and computer*	و کامپیوتر	17	2
3	computer in	کامپیوتر در	16	2
4	electronic and computer	الکترونیک و کامپیوتر	8	3
5	computer engineering	مهندسی کامپیوتر	8	2
6	*computer + direct object*	کامپیوتر را	7	2
7	computer sciences	علوم کامپیوتر	7	2
8	*use of the computer*	استفاده ازکامپیوتر	7	3
9	*one/a computer*	یک کامپیوتر	7	2
10	Internet and computer	اینترنت و کامپیوتر	6	3
11	computer electronics	الکترونیک کامپیوتر	6	2
12	*from computer*	از کامپیوتر	6	2
13	*personal computer*	کامپیوتر شخصی	6	2
14	computer dimension	ابعاد کامپیوتر	6	2
15	computer protecting products	محصولات حفاظت کننده کامپیوتر	5	4
16	computer protection	حفاظت کننده از کامپیوتر	5	4
17	computer electronics and	الکترونیک کامپیوتر و	5	3
18	*with computer*	با کامپیوتر	5	2

Table 7: Persian word n-grams

N	Translation	N-gram	Freq.	Length
1	*a computer*	رایانه ای	291	2
2	some computers	رایانه های	109	2
3	computers	رایانه ها	65	2
4	cloud computing	ابر رایانه	60	2
5	computer games	بازیهای رایانه	51	2
6	this computer	این رایانه	49	2
7	*from computer*	از رایانه	31	2
8	computer crime	جرایم رایانه	28	2
9	to computer	به رایانه	26	2
10	*use of the computer*	استفاده از رایانه	24	3
11	corporate computer	صنفی رایانه	22	2
12	tablet PC	لوح رایانه	22	2
13	*computer and*	رایانه و	22	2
14	association of computing	نظام صنفی رایانه	22	3
15	*and computer*	و رایانه	21	2
16	computer trade organization	سازمان نظام صنفی رایانه	20	4
17	in computer	در رایانه	20	2
18	*computer + direct object*	رایانه را	19	3
19	fighting with computer crime	مبارزه با جرایم رایانه	18	4
20	with computer crime	با جرایم رایانه	16	3
21	from computers	از رایانه های	15	2
22	national computer	رایانه ملی	15	2
23	national foundation of computer games	بنیاد ملی بازیهای رایانه	14	4
24	in computer	روی رایانه	14	2
25	national computer games	ملی بازیهای رایانه	14	3
26	national supercomputer	ابر رایانه ملی	13	3
27	computer game	بازی رایانه	13	2
28	personal computers	رایانه های شخصی	13	2
29	PCs	رایانه های خانگی	12	2
30	first computer	اولین رایانه	12	2
31	computer is	رایانه ای است	11	2
32	world computer	رایانه جهان	11	2
33	*with computer*	با رایانه	11	2
34	computer virus	ویروس رایانه	10	2
35	on computer	بر روی رایانه	10	3
36	thousands of computers	هزار رایانه	8	2
37	special computer	ویژه رایانه	8	2

38	computer that	رایانه که	8	2
39	*personal computer*	رایانه شخصی	8	2
40	foreign computer	رایانه ای خارجی	8	2
41	production of national supercomputer	تولید ابر رایانه ملی	8	4
42	infected computers	رایانه های آلوده	7	2
43	computer industry	صنعت رایانه	7	2
44	portable computers	رایانه های قابل	7	2
45	world computers	رایانه های دنیا	7	2
46	first supercomputer	اولین ابر رایانه	6	3
47	computer bug/worm	کرم رایانه	6	2
48	computer system	سیستم رایانه	6	2
49	PC	طریق رایانه	6	2
50	home computer	رایانه خانگی	6	2
51	company computer	رایانه شرکت	6	2
52	father of computer	پدر رایانه	5	2
53	telecom computing	رایانه شرکت مخابرات	5	3
54	oldest computer	قدیمی ترین رایانه	5	3
55	computer information	اطلاعات رایانه	5	2
56	computer fraud and scums	رایانه ای جعل و کلاهبرداری	5	5
57	computer crime hoax	با جرایم رایانه ای جعل	5	4
58	computer for	رایانه ای برای	5	2

There are eight common n-grams (presented in italics in the tables) between the English loanword and its Persian native counterpart, and it is the Persian word that tends to have more of them. Yet, when looking at Figure 2 presenting n-gram relative frequency, the difference is not that big, and in fact it is the English loanword that has more n-grams with respect to its frequency in the corpus:
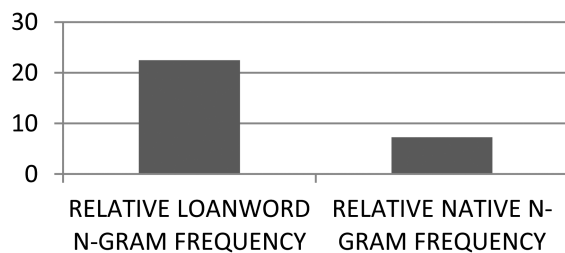
Figure 2: Relative n-grams frequency with respect to common n-grams

In order to establish the semantic similarity between the loanword and its Farsi counterpart based on their n-grams, the following formula (collocational overlap measure) was used:

$$d(i,j) = 1 - \frac{2 \times (ci \cap cj)}{ci + cj}$$

N-grams overlap = 0.79
Semantic similarity: YES

There are eight common n-grams that stand for 0.79 of n-gram overlap. Given that semantically similar lexemes have the overlap less than 0.8, the two lexemes in question can be classified as similar in terms of n-grams.

To answer the question whether the difference between the number of n-grams is statistically significant, the Chi² test was applied. The Chi² value is calculated from the formula:

$$X^2 = (df, N) = \sum \frac{(fo - fe)^2}{fe}$$

In this example

$$X^2 = (1, 879) = 21.37$$
$p < 0.01,$

which proves that there is a significant difference between the number of loanword and its native counterpart n-grams.

Another level of analysis consists of studying phrasemes or multiword expressions of both lexemes. A phraseme can be here very generally described as a conventional, repeated multiword unit.[16] Tables 8 and 9 present loanword and Persian phrasemes respectively:

Table 8: Phrasemes with loanword

1.	computer engineering	مهندسی کامپیوتر
2.	computer sciences	علوم کامپیوتر
3.	personal computer	کامپیوتر شخصی
4.	computer dimension	ابعاد کامپیوتر

---

16 For more on phrasemes, see, e.g., Dobrovol'skij & Piirainen (2005).

Table 9: Phrasemes with Farsi word

1.	supercomputer	ابر رایانه
2.	cloud computing	رایانه ابری
3.	computer crime	جرایم رایانه
4.	corporate computer	صنفی رایانه
5.	tablet PC	لوح رایانه
6.	computer game	بازی رایانه
7.	PCs	رایانه های خانگی
8.	computer virus	ویروس رایانه
9.	personal computer	رایانه شخصی
10.	infected computers	رایانه های آلوده
11.	computer industry	صنعت رایانه
12.	portable computers	رایانه های قابل
13.	computer bug/worm	کرم رایانه
14.	father of computer	پدر رایانه

In order to establish the semantic similarity between the loanword and its Farsi counterpart, this time based on their phrasemes, the formula of the collocational overlap was used:

$$d(i,j) = 1 - \frac{2 \times (ci \cap cj)}{ci + cj}$$

Phraseme overlap = 0.89
Semantic similarity: NO

There is only one common phraseme that stands for 0.89 of phraseme overlap. Given that semantically similar lexemes have the overlap at most 0.8, the two lexemes in question cannot be classified as similar in terms of phrasemes.

Again, the Chi^2 value was calculated to answer the question whether the difference between the numbers of phrasemes was statistically significant. Chi^2 value is calculated from the following formula:

$$X^2 = (df, N) = \sum \frac{(fo - fe)^2}{fe}$$

In this example:

$$X^2 = (1, 879) = 3.62$$
$p > 0.5,$

which proves that there is no statistically significant difference between loan-word and Persian phrasemes since if the chi square value results in a probability that is more than 0.05 (i.e., more than 5 percent), it is not considered statistically significant.

The above example shows that in terms of n-grams, this particular loanword and its native Persian counterpart can be classified as similar. Yet, the results of their phrasemes give contradictory results. Let's now consider other words that hypothetically could be considered as similar:

Table 10: Semantic similarity based on phrasemes

Lexeme translation	Synonymic lexemes	Collocational overlap	Classified as semantically similar
strategy	استراتژی راهبرد	1	NO
system	سیستم نظام	1	NO
candidate	نامزد کاندیدا	0.8	NO
model	الگو مدل	1	NO
touristic	توریستی گردشگر	0.82	NO
symbol	سمبل نماد	1	NO
computer	رایانه کامپیوتر	0.89	NO
game	بازی گیم	1	NO
digital	رقمی دیجیتال	1	NO

Table 11 presents the results of analyzing semantic similarity of hypothetically similar words. As can be seen, none of them can be classified as semantically similar in terms of phrasemes and only five can be classified as similar in terms of n-grams. Therefore, the study reveals that words that at first glance could be treated as similar cannot be treated as such, at least not in terms of the phra-semes they form. Of course, it needs to be remembered that corpus results always depend on the type of corpora used. In this case, the results are confined to the corpus whose data represent the newspaper domain.

Table 11: Semantic similarity based on n-grams

Lexeme translation	Synonymic lexemes	Collocational overlap	Classified as semantically similar
strategy	استراتژی راهبرد	0.83	NO
system	سیستم نظام	0.77	YES
candidate	نامزد کاندیدا	0.64	YES
model	الگو مدل	0.86	NO
touristic	توریستی گردشگر	0.68	YES
symbol	سمبل نماد	0.76	YES
computer	رایانه کامپیوتر	0.79	YES
game	بازی گیم	0.98	NO
digital	رقمی دیجیتال	0.82	NO

5 Conclusion

Dr. Samuel Johnson, when emphasizing the importance of semantics in the preface to his pioneering dictionary in 1799, wrote: "It is not sufficient that a word is found, unless it be so combined as that its meaning is apparently determined by the track and tenor of the sentence". Therefore, the core of studying meaning is to analyze words in context.

In this article, certain procedures for extracting semantic information from texts have been discussed. I have tried to present it with the example of the Persian language. The loanword-native equivalent analysis aimed at outlining semantic differences and similarities between the two with respect to their n-grams and phrasemes. The analysis revealed that words that at first glance seem to be similar, in terms of their collocates or phrasemes, cannot necessarily be treated as such. Consider, for example, the *computer* case. It seems that both English loanword and Persian native word favor different collocates, e.g., مهندسی کامپیوتر *mohandesi kāmpiuter* 'computer engineering' or علوم کامپیوتر *elme kāmpiuter* 'computer sciences' in the case of the loanword and رایانه ابری *rāyāne abri* 'cloud computing' or جرایم رایانه *jarāme rāyāne* 'computer crime' in the

case of the native word. The fact that hypothetically similar words tend to form different phrasemes and the possibility of identifying these different phrasemes in a corpus seem to be particularly interesting. This finding seems to be of great importance to translators and students of Persian who often struggle with the lexical richness of the Persian language. Although this article presents only an extract on the abounding study of meaning, it provides empirical analysis and therefore it attempts to contribute to the acquisition of meaning – "the Holy Grail of lexical acquisition" (Manning and Schütze 2000: 294).

References

Baker, Collin F., Charles J. Fillmore & John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, 86–90. Montreal: Association for Computational Linguistics.

Bijankhan, Mahmood, Sheykhzadegan Javad, Bahrani Mohammad & Masood Ghayoomi. 2011. Lessons from building a Persian written corpus: Peykare. In *Language Resources and Evaluation*, 45 (2). 143–164. Secaucus: Springer.

Brown, Peter, Stephen Della Pietra, Vincent Della Pietra & Robert Mercer. 1991. Word sense disambiguation using statistical methods. In *Proceedings of the 29th annual meeting of the Association for Computational Linguistics*, 264–270. Stroudsburg: Association of Computational Linguistics.

Dale, Robert, Hermann Moisl & Harold Somers. 2000. *Handbook of natural language processing*. New York: Marcel Dekker.

Danielsson, Pernilla. 2001. *The automatic identification of meaningful units in language*. Göteborg: Göteborg University dissertation.

Dehdari, Jon & Deryle Lonsdale. 2008. A link grammar parser for Persian. In Simin Karimi, Vida Samiian & Donald Stilo (eds.), *Aspects of Iranian linguistics*, vol. 1: 19–34. Cambridge: Cambridge Scholars Press.

Dehkharghani, Rahim & Mehrnoush Shamsfard. 2009. Mapping Persian words to WordNet Synsets. *The International Journal of Interactive Multimedia and Artificial Intelligence* 1 (2). 6–13.

Dobrovol'skij, Dmitri O. & Elisabeth Piirainen. 2005. *Figurative language: Cross-cultural and cross-linguistic perspectives*. Amsterdam: Elsevier.

Eneko, Agirre & Philip Edmonds. 2007. *Word sense disambiguation: Algorithms and applications*. Dordrecht: Springer.

Farajian, Mohammad Amin. 2011. PEN: Parallel English-Persian news corpus. In *Proceedings of 2011 International Conference on Artificial Intelligence* (ICAI'11), 523–528. Las Vegas: ICAT.

Fellbaum, Christiane (ed.). 1998. *WordNet: An electronic lexical database and some of its applications*. Cambridge: MIT Press.

Firth, John. 1957. A synopsis of linguistic theory, 1930–1955. In John Firth (ed.), *Studies in linguistic analysis. Special volume of the Philological Society*, 1–32. Oxford: Blackwell.

Ghayoomi, Masood & Saeedeh Momtazi. 2009. Challenges in developing Persian corpora from online resources. In Min Zhang, Haizhou Li, Kim-Teng Lua & Minghui Dong (eds.), *Proceedings of 2009 IEEE International Conference on Asian Language Processing*, 108–113. Los Alamitos: IEEE Computer Society Press.

Ghayoomi, Masood, Saeedeh Momtazi & Mahmood Bijankhan. 2010. A study of corpus development for Persian. *International Journal on Asian Language Processing* 20 (1). 17–33.

Girju, Roxana, Adriana Badulescu & Dan Moldovan. 2003. Learning semantic constraints for the automatic discovery of part-whole relations. In *Proceedings of the Association for Computational Linguistics – Human Language Technologies*, 1–8. Stroudsburg: Association for Computational Linguistics.

Hamidi, Mandana, Ali Borji & Saeed Shiry Ghidary. 2007. Persian word sense disambiguation. In *Proceeding of 15th Iranian Conference of Electrical and Electronics Engineers*, 114–118. Teheran: ICEE.

Hashemi, Homa Baradaran. 2010. Creating a Persian-English comparable corpus. In Maristella Agosti, Nicola Ferro, Carol Peters, Maarten de Rijke & Alan F. Smeaton (eds.), *Proceedings of the 2010 International Conference on Multilingual and Multimodal Information Access Evaluation: Cross-language evaluation forum lecture notes in computer science*, 27–39. Padua: Springer.

Hirst, Graeme. 1987. *Semantic interpretation and the resolution of ambiguity*. Cambridge: Cambridge University Press.

Hovy, Eduard, Mitchell Marcus, Martha Palmer, Lance Ramshaw & Ralph Weischedel. 2006. OntoNotes: The 90% solution. In *Proceedings of the Human Language Technology Conference of the NAACL*, 57–60. New York: Association for Computational Linguistics.

Ide, Nancy & Jean Veronis. 1990. Word sense disambiguation with very large neural networks extracted from machine readable dictionaries. In Hans Karlgren (ed.), *Proceedings of the 13th Conference on Computational linguistics*, 389–404. Helsinki: Association for Computational Linguistics.

Jadidinejad, Amir Hossein, Fariborz Mahmoudi & Jon Dehdari. 2010. Evaluation of Perstem: A simple and efficient stemming algorithm for Persian. In Carol Peters, Giorgio Di Nunzio, Mikko Kurimo, Thomas Mandl, Djamel Mostefa, Anselmo Peñas & Giovanna Roda (eds.), *Multilingual information access evaluation I. Text retrieval experiments*, *lecture notes in computer science*, 98–101. Heidelberg: Springer.

Kim, Jin-Dong, Tomoko Ohta, Yuka Teteisi & Jun'ichi Tsujii. 2003. GENIA corpus – a semantically annotated corpus for bio-textmining. In *Bioinformatics*, 180–182. Oxford: Oxford University Press.

Kiryakov, Atanas, Popov Borislav, Terziev Ivan, Manov Dimitar & Damyan Ognyanoff. 2004. Semantic annotation, indexing, and retrieval. In *Web Semantics: Science, Services and Agents on the World Wide Web* 2 (1). 49–79. Amsterdam: Elsevier.

Lesk, Michael. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In Virginia DeBuys (ed.), *Proceedings of the 5th Annual International Conference on Systems Documentation*, 24–26. New York: ACM.

Magnini, Bernardo & Gabriela Cavagliá. 2000. Integrating subject field codes into WordNet. In *Proceedings of LREC-2000, Second International Conference on Language Resources and Evaluation*, 1413–1418. Athens: European Language Resources Association.

Makki, Raheleh & Mohammad Mehdi Homayounpour. 2008. Word sense disambiguation of Farsi homographs using thesaurus and corpus. In *Advances in natural language processing, Lecture notes in computer science*, 315–323. Gothenburg: Springer.

Makki, Raheleh & Mohammad Mehdi Homayounpour. 2008. Using decision list for Farsi word sense disambiguation. In *Proceedings of 2nd Joint Congress on Fuzzy and Intelligent Systems*. Teheran: Malek Ashtar University.

Manning, Chris & Hinrich Schütze. 2000. *Foundations of statistical natural language processing*. Cambridge: MIT Press.

Marszałek-Kowalewska, Katarzyna. 2010. Iranian language policy: A case of linguistic purism. In *Investigationes Linguisticae*, 89–103. Poznan: Institute of Linguistics UAM.

Mason, Olivier. 2006. *The automatic extraction of linguistic information from text corpora*. Birmingham: University of Birmingham dissertation.

McRoy, Susan. 1992. Using multiple knowledge sources for word sense discrimination. In *Computational linguistics*, 1–30. Cambridge: MIT Press.

Megerdoomian, Karine. 2010. Developing a Persian part-of-speech tagger. In *Zabane Farsi va Rayane* [Persian language and computers]. Tehran: SAMT Publishers.

Miangah, Mosavi & A. Delavar Khalafi. 2005. Word sense disambiguation using target language corpus in a machine translation system. In *Literary and linguistic computing*, 237–249. Oxford: Oxford University Press.

Mitkow, Ruslan (ed.). 2003. *The Oxford handbook of computational linguistics*. Oxford: Oxford University Press.

Monajemi, Ebrahim. 2011. Can ethnic minority languages survive in the context of global development? http://www.sil.org/asia/ldc/parallel_papers/ ebrahim_monajemi.pdf (accessed 5 January 2011).

Palmer, Martha, Dan Gildea & Paul Kingsbury. 2005. The proposition bank: A corpus annotated with semantic roles. In *Computational Linguistics Journal* 31 (1). 71–106.

Prasad, Rashimi, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi & Bonnie Webber. 2008. The Penn Discourse Treebank 2.0. In Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC), 2961–2968. Marrakech: European Language Resource Association.

Reeve, Lawrence & Han Hyoil. 2005. Survey of semantic annotation platforms. In Lorrie M. Liebrock (ed.), *SAC'05: Proceedings of the 2005 ACM Symposium on Applied Computing*, 1634–1638. New York: ACM.

Roberts, Angus, Robert Gaizauskas, Mark Hepple, Neil Davis, George Demetriou, Yikun Guo, Jay (Subbarao) Kola, Ian Roberts, Andrea Setzer, Archana Tapuria & Bill Wheeldin. 2007. The CLEF corpus: Semantic annotation of clinical text. In *Proceedings of the AMIA Knowledge Centre*, 625–629. Chicago: American Medical Informatics Association.

Rouhizadeh, Masoud, Mahsa A Yarmohammadi & Mehrnoush Shamsfard. 2008. Building a WordNet for Persian verbs. In Attila Tanács, Dóra Csendes, Veronika Vincze, Christiane Fellbaum & Piek Vossen (eds.), *Proceedings of The Fourth Global WordNet Conference*, 406–412. Szeged: University of Szeged.

Sagot, Benoît, Géraldine Walther & Pegah Faghiri, Pollet Samvelian. 2011. Développement de ressources pour le persan : le nouveau lexique morphologique PerLex 2 et l'étiqueteur morphosyntaxique MEltfa. In *Proceedings of TALN 2011, Conference on Natural Language Processing*. Montpellier: TALN.

Sampson, Geoffrey. 2001. *Empirical linguistics*. London & New York: Continuum.

Samvelian, Pollet & Pegah Faghiri. 2013. Introducing PersPred, a syntactic and semantic database for Persian complex predicates. In Valia Kordoni, Carlos Ramisch & Aline Villavicencio (eds.), *Proceedings of the 9th Workshop on Multiword Expressions, NAACL-HLT 2013*, 11–20. Atlanta: Association for Computational Linguistics.

Schütze, Henrich. 1992. Dimensions of meaning. In Robert Werner (ed.), *Proceedings of the 1992 The Association for Computing Machinery (ACM) and the IEEE Computer Society conference on Supercomputing*, 787–796. Los Alamitos: IEEE Computer Society Press.

Seraji, Mojgan, Beáta Megyesi & Joakim Nivre. 2012. A basic language resource kit for Persian. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Ugur Dogan, Bente Maegaard, Joseph Mariani, Jan Okij & Stelios Piperidis (eds.), *Proceedings of the Eight International Conference on Language Resources and Evaluation* (LREC'12), 2245–2252. Istanbul: European Language Resource Association.

Shamsfard, Mehrnoush. 2011. Challenges and open problems in Persian text processing. In *Proceedings of Language Technology Conference 11*, 65–69. Poznań: LTC.

Shamsfard, Mehrnoush, Hakimeh Fadaei & Elham Fekri. 2010. Extracting lexico-conceptual knowledge for developing Persian WordNet. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner & Daniel Tapias (eds.), *Proceedings of* Language Resources and Evaluation (LREC 2010), 3798–3802. Malta: European Language Resource Association.

Shamsfard, Mehrnoush, Akbar Hesabi, Hakimeh Fadaei, Niloofar Mansoory, Ali Famian, Somayeh Bagherbeigi, Elham Fekri, Maliheh Monshizadeh & S. Mostafa Assi. 2010. Semi automatic development of FarsNet: The Persian WordNet. In *Proceedings of 5th Global WordNet Conference*. Mumbai: GWA.

Shamsfard, Mehrnoush, Hoda Sadat Jafari & Mahdi Ilbeygi. 2010. STeP-1: A set of fundamental tools for Persian text processing. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner & Daniel Tapias (eds.), *Proceedings of Language Resources and Evaluation (LREC 2010)*, 859–865. Malta: European Language Resource Association.

Small, Steven. 1980. *Word expert parsing: A theory of distributed word-based natural language understanding*. College Park: University of Maryland dissertation.

Soltani, Mehdi. 2010. A statistical approach on Persian word sense disambiguation. In Informatics and Systems (INFOS), 1–6. Cairo: IEEE Computer Society Press.

Stevenson, Mark & Yorick Wilks. 2003. Word sense disambiguation. In Ruslan Mitkov (ed.), *Oxford handbook of computational linguistics*, 249–265. Oxford: Oxford University Press.

Taheri, Aynaz & Mehrnoush Shamsfard. 2011. Mapping FarsNet to suggested upper merged ontology. In Mohamed Mohamed Salem, Khaled Shaalan, Farhad Oroumchian, Azadeh Shakery & Halim Khelalfa (eds.), *Proceedings of Asia Information Retrieval Societies Conference 2011*, 604–613. Dubai: Springer.

Waltz, David & Jordan Pollack. 1985. Massively parallel parsing: A strongly interactive model of natural language interpretation. In *Cognitive Science*, 51–74. Norwood: Ablex Publishing.

Wilks, Yorick. 1972. *Grammar, meaning and the machine analysis of language*. London & Boston: Routledge.

Yarowsky, David. 1992. Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. In *Proceedings of COLING-92*, 454–460. Stroudsburg: Association for Computational Linguistics.

List of contributors

Mohammad Abdolhosseini
Isfahan University, Isfahan, Iran
mabdolhosseini@gmail.com

Shadi Dini
Drexel University, School of Education
One Drexel Plaza, 3141 Chestnut Street,
Philadelphia PA, 19104, USA
shadi.dini@drexel.edu

Zohreh R. Eslami
Texas A&M University
College Station, Texas 77845, USA
zeslami@tamu.edu

Lewis Gebhardt
Department of Linguistics, Northeastern
Illinois University
5500 North St. Louis Avenue, Chicago,
IL 60625, USA
l-gebhardt@neiu.edu

Jila Ghomeshi
Department of Linguistics, University of
Manitoba
Winnipeg, MB, R3T 5V5, Canada
Jila.Ghomeshi@umanitoba.ca

Shinji Ido
Graduate School of Humanities,
Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-8601,
Japan
ido@nagoya-u.jp

Youli Ioannesyan
Institute of Oriental Manuscripts of the
Russian Academy of Sciences
Dvortsovaya emb., 18, Saint Petersburg,
191186, Russia
youli19@gmail.com

Carina Jahani
Department of Linguistics and Philology,
Uppsala University
Box 635, SE-751 26 Uppsala, Sweden
carina.jahani@lingfil.uu.se

Alireza Korangy
Societas Philologica Persica
438 West 37th Street, Room 5H, New York,
NY 10018, USA
korangy@gmail.com

Agnès Lenepveu-Hotz
Department of Persian Studies, University
of Strasbourg
22 rue René Descartes, 67084 Strasbourg,
France
a.hotz@unistra.fr

Behrooz Mahmoodi-Bakhtiari
University of Tehran, College of Fine Arts,
Department of Performing Arts
Enghelab Avenue, Tehran, Iran
mbakhtiari@ut.ac.ir

Shahrzad Mahootian
Department of Linguistics, Northeastern
Illinois University
5500 North St. Louis Avenue, Chicago,
IL 60625, USA
s-mahootian@neiu.edu

Katarzyna Marszalek-Kowalewska
Institute of Linguistics, Adam Mickiewicz
University
al Niepodległości 4, 61-874 Poznań
k.marszalek.kowalewska@gmail.com

Corey Miller
The MITRE Corporation
7515 Colshire Drive, McLean, VA 22102, USA
camiller@mitre.org

Lutz Rzehak
Berlin Humboldt-University, Central-Asian
Seminar
Unter den Linden 6, 10099 Berlin, Germany
lutz.rzehak@gmx.de

Adriano V. Rossi
Department of Asian, African and
Mediterranean Studies
University L'Orientale – Piazza S.
Domenico M. 12, 80134 Naples, Italy
arossi@unior.it

Hooman Saeli
University of Tennessee, Knoxville
2344 Dunford Hall, Knoxville, TN 37996, USA
hsaeli@utk.edu

Martin Schwartz
Department of Near Eastern Studies,
University of California, Berkeley
250 Barrows Hall, Berkeley, CA 94720, USA
martz@berkeley.edu

Toon Van Hal
Department of Linguistics, University of
Leuven
Blijde-Inkomststraat 21, pb 3008, BE-3000
Leuven, Belgium
toon.vanhal@kuleuven.be

Arseniy Vydrin
Institute for Linguistic Studies of the Russian
Academy of Sciences
Tuckhov pereulok, 9, Saint Petersburg,
199053, Russia
senjacom@gmail.com

Z. A. Yusupova
Institute of Oriental Manuscripts of the
Russian Academy of Sciences
Dvortsovaya emb., 18, Saint Petersburg,
191186, Russia
youli19@gmail.com

Index

ablative 287, 293
address forms 135–41, 143–7, 149–53, 155–
    61
address pronoun 138
adjuncts x, 101, 103, 105, 107, 109, 111, 113,
    115, 117
adverbs 197, 199, 210, 231, 239, 315–16,
    319, 322, 324, 326, 329, 342–3
– spatial 314–15, 321
Adyghe 294
Afghanistan x, 67, 71, 101–2, 104, 111, 114,
    118–19, 121, 123, 131, 311, 342
agreement affixes 264, 268–9, 271–2
allophones 120, 125, 127
Amerindian languages 3
animacy 87–90, 97, 222, 238, 280
Arabs 73, 75, 110, 114, 117
Aramaic 69, 72–3, 76
argots 70, 72–3, 75–7
Armenian 4, 217–18
associative plurals 233–6, 240–2, 247–8,
    258, 260, 262

Badakhshani 121, 127, 132
Balkh 110–11, 144
Balochi xi, 54–9, 61, 63–4, 66–7, 309, 311,
    337–8
– Koroshi 309–12, 314–15, 319, 321, 329–
    32, 335–6
– lexicography x, 53, 55, 57, 59, 61, 63, 65,
    67
– Sistani 309–12, 314–15, 321–2, 324, 329–
    30, 332, 335–6
Bengali 245
Benveniste, Emile 62, 65, 82, 99
bilingualism 188, 211
Brāhūī 65

Caucasian languages 277–8, 292
Central Asia 50, 69–71, 73–5, 77, 119, 122
Cheshire, Jenny 247, 249, 252, 260
Cheung, Johnny 77–9
Chierchia, Gennaro 216–18, 227, 231

Chinese 21–2, 24, 32, 37, 42, 44, 50–1, 231–2
– Middle 44, 51
circumposition 81–8, 91–2, 94–5, 97–8
classical Persian 76, 90, 100, 119–20, 129–
    30, 190
classifiers 213–20, 222, 224–8, 230–2, 237,
    259, 261
cliticization 276
clitics xi, 193, 196, 198–9, 204, 259, 263–4,
    266–73, 275–6, 279, 284
collocations 76, 347–8, 357
colloquial Dari x, 101, 103, 105–7, 109, 111,
    113, 115, 117
colloquial Persian 113, 182–4, 188, 190–201,
    204–12, 214, 233, 261, 276
complex verbs 85, 87, 91, 276–7, 288
compound verbs 85, 198, 205, 264–6, 269,
    271, 275
computational linguistics 339–42, 358–61
Construction Grammar 234, 237, 254, 256,
    260
Construction Morphology 233, 254, 256, 260
corpus linguistics 339–40, 342

Dari x, xii, 101–3, 105, 107, 110, 113–23, 126,
    130
– colloquial x, 101, 103, 105–7, 109, 111, 113,
    115, 117
– literary 124, 126–7
dative 92, 94, 279–83, 287–8, 291, 293, 315
deictics 223, 314, 327, 335
deixis xi, 309, 312–13, 319, 322, 329, 335–8
Delforooz, Barjasteh 310–12, 314–15, 321–3,
    332–3, 336–7
demonstrative determiners 314, 319–21,
    324, 329, 332, 336
demonstrative pronouns 314–15, 318, 321,
    323–4, 329–30, 337
derivation 54, 72, 152, 221, 230–1, 256, 294
determiner 216, 221, 223–4, 261, 310, 314–
    15, 317, 319, 321–4, 326, 329, 344
demonstrative 314, 319–21, 324, 329, 332,
    336