From an encyclopedia of Iranian Folklore to an ontology of Iranian folklore

Mohsen Emadi

Master ‘s Thesis

Master Program in Digital Culture

Department of Art and Culture Studies

University of Jyväskylä

July 2014

University of Jyväskylä

Faculty Faculty of Humanities	**Department** Department of Art and Culture
Author Mohsen Emadi	
Title From an encyclopedia of Iranian Folklore to an ontology of Iranian folklore	
Subject Digital Culture	**Level** Master‘s Thesis
Month and year July 2014	**Number of pages** 60 + 19 appendices

Abstract

The main resource of the thesis came from 37 years of research work by Ahmad Shamlou, an Iranian poet. The body of research should be transformed from manuscript to digital material. Designing and implementing a semantic ontology for such a folk encyclopedia is a basic and essential part of the thesis. As the material bank is significantly large, semantic marking of the materials in their entirety must be done in a group action and this thesis aims to develop the appropriate framework for the encyclopedia.

The model aims to fulfill the philosophy behind Shamlou's approach to the concept of Encyclopedia as an indexical reference to the world by the language and words.

The thesis will analyze different approaches in ontology deployment from a basic lexical ontology based on WordNet to an extended model for ontology development involving different domains of knowledge in the field of Folklore. Deploying such model happens in the real conditions surrounding the faith of the remained materials and therefore the digitalization and indexing process must be discussed on pragmatic bases.

Keywords Digitalization, Folklore, Natural Language Processing, Ontology Development, Semantic Web

Depository University of Jyväskylä

Additional information

Jyväskylän Yliopisto

Tiedekunta Humanistinen tiedekunta	**Laitos** Taiteiden ja kulttuurin tutkimuksen laitos
Tekijä Mohsen Emadi	
Työn nimi Iranin kansanperinteen tietosanakirjasta Iranin kansanperinteen ontologiaan	
Oppiaine Digitaalinen kulttuuri	**Työn laji** Pro gradu -tutkielma
Aika 2014	**Sivumäärä** 58 + 19 s. liitteitä

Tiivistelmä

Opinnäytteen pääaineisto tuli 37 vuotta tutkimustyötä tehneen Iranilaisen runoilijan Ahmad Shamloun aineistosta. Tutkimuksen tarkoitus on muuttaa käsikirjoitus digitaaliseksi aineistoksi. Opinnäytteen perus- ja olennaisin osa on semanttisen ontologian suunnittelu ja toteutus kansanperinne ensyklopediaan. Koska materiaalipankki on huomattavan suuri, on materiaalin semanttinen merkintä tehtävä kokonaisuutena ryhmässä ja tämän opinnäytetyön tavoitteena on kehittää sopiva rakenne tälle ensyklopedialle.

Rakennemalli pyrkii noudattamaan Shamloun ajattelutapaa, jonka mukaan tietosanakirjan tulee viitata kielensä kautta maailmaan indeksikaalisesti.

Opinnäyte analysoi erilaisia lähestymistapoja ontologian kehityksestä, perus leksikaalisesta sananmuodostuksesta, joka perustuu WordNettiin. Laajentaen sen mallin ontologian kehitykseen, johon kuuluu eri aloilla olevaan tietämykseen kansanperinteestä. Tällaisen mallin kehitys tapahtuu todellisissa olosuhteissa, uskollisena säilyneelle materiaalille, ja sen vuoksi digitalisointi ja indeksointi prosessi täytyy toteuttaa käytännöllisesti.

Asiasanat digitalisointi, kansanperinne, luonnollinen kielen prosessi, ontologian kehitys, semanttinen verkko

Säilytyspaikka Jyväskylän Yliopisto

Muita tietoja

Content

List of Figures

List of Tables

Chapter 1: Introduction

According to Britannica the term folklore invented in 1846 by Mr W. J. Thomas as a designation for the traditional learning of the uncultured classes of civilized nations. Collier's new encyclopedia defines folklore as the science which embraces all that relates to ancient observances and customs, to the notions, beliefs, traditions, superstitions, and prejudices of the common people. Henry Glassie asserts that " *The center of folklore is a merger of individual creativity and social order philosophically, politically. [...] it stresses the interdependence of the personal, the social; the aesthetic, the ethical, the cosmological; the beautiful, the good, the true. Practically, folklore is the study of human creativity in its own context*" [Magoulick]. Therefore any category of artistic, musical or literary composition characterized by a particular style, form, content or context belongs to folklore.

The main resource of the thesis came from the research work of Ahmad Shamlou (December 12, 1925 — July 24, 2000), an Iranian poet, translator and researcher. Ahmad Shamlou devoted four decades

of his life to collect and organize Iranian folklore. The research work entitled Ketab Kuche (Translation: The Book of Alleyways) indexes the Iranian folklore through the Persian language. The book is a multi-volume, multi-disciplinary work designed as a major source of information, providing a detailed and accurate picture of an important world civilization over a span of several thousand years. fourteen volumes of the book were printed during his lifetime and about eighty thousand entries and essays on the subject remained unpublished and uncategorized.

Due to several conditions surrounding the work and its author from political to religious and institutional, the project and its development remained as an independent research and publishing project. The thesis aims to analyze practical solutions for digitalization and semantic publication of the encyclopedia considering all the resource limitations and the special conditions of the work.

The objective of the thesis is *to purpose a solution for transforming the body of the encyclopedia from manuscript to meaningful digital material*. The digitalization standard purposed and applied in this project can help other independent projects running in Iran or other developing countries facing the same problems. It aims *to follow and analyze the process of design and implementation of a semantic ontology for such a folk encyclopedia as a case study for such projects*. As the material bank is significantly large, semantic marking of the materials in their entirety should be collaborative and the thesis aims *to develop the appropriate semantic framework for the encyclopedia*. Preparing an appropriate ontological framework for semantic marking and automatizing the semantical tagging and indexing of the encyclopedia remain as of one of the major goals of the project. Considering the fact that different languages dialects and traditions currently live in Iran, the overall goal of the project will be an extended dynamic ontological system that could handle the inter-lingual, inter-cultural indexing for such kind of encyclopedias. Therefore the research involves studies on natural language processing in Persian language, ontology development and discussions of the possibilities of collaborative ontology development.

1.1 History of research in Iranian folklore

Ulrich Marzolph describes the history of the research in Iranian folklore in Iranica, a major reference encyclopedia in Iranian studies. In his perspective "*the beginning of folklore studies can be detected in the keen interest early Western travelers took in Persia since the 18th century.[...]While the initial focus of Western scholars since the discovery and translation of the Avesta was on religious studies, by way of linguistic interest in dialect specimens they soon turned to collecting items of folklorist relevance, such as folk-tales, riddles, songs, or narratives of everyday life.[...] Folklore studies by Persian scholars did not occur until the third decade of the 20th century, when strong nationalist feelings coincided with a growing awareness of the phenomenon of the "common people," mingled with a romantic urge for unspoiled tradition, for which the Western researchers had paved the way*" [Marzolph, 71-75].

Marzolph mentions the works of JamalZadeh (January 13, 1892, Isfahan, Iran – November 8, 1997, Geneva, Switzerland) and Hedayat (February 17, 1903, Tehran – 9 April 1951, Paris, France) the fathers of Persian modern novels: "*Hedāyat was the first Iranian to study folklore and outline its scholarly methods*" [Marzolph]. Hedāyat in his Neyrangestān (1933) published a survey of superstition and folk beliefs and practices, while in his essay "Folklor yā farhang-e tūda" (1945), following Pierre Saintyves, he supplied first general outlines on the collection and documentation of folklore.

The research on Iranian folklore was based on the work of individual researchers; However in 1940, Fazl-o-llah Mohtadi Sobhi (1897 – 1962) a writer and story-teller for children in Iranian National Radio started to involve his audience in collecting folktales and later in 1945 and 1946 he published two volumes of Persian folktales. In the year 1958 Edāra-ye farhang-e ʿāmma was found within the context of Ministry of Culture and Arts; a center for popular culture that in 1970 would turn to be the center of studies in Iranian Anthropology and popular culture. Abolghasem Enjavi Shirazi (1921 –

1993), probably the main figure of studies in Iranian folklore, in 1960 used Sobhi's example in a radio program in order to collect folk materials. He trained a considerable staff as his co-workers. The guideline he was using in collecting folklore was based on the outlines suggested by Hedayat. After Islamic revolution his radio program was discontinued. In the context of Islamic Republic in the year 1985 the Organization of Cultural Heritage was found that aims to supervise all kind of cultural activities, archeology, anthropology, and folklore. As is described by Marzolph:

> "Its relevant activities besides the publication of its journal Mīrāṯ-e farhangī (since 1990) include the organization of a first scientific meeting on anthropology and folklore in 1990. The anthropology department, besides educating junior folklorists (up to M.A. level), is conducting field work research within the country on various topics such as water usage and irrigation, children's games, folk medicine, traditional clothing, and storytelling. One of the most recent research projects, conducted for a period of two years, was concerned with a detailed general survey of popular literature (adab-e ʿāmma). No monograph publications have yet resulted from these research projects." [Marzolph].

Political and personal concerns could be detected in the narration of Marzolph especially when he neglects to speak about political concerns and their impact on the research of folklore in Iran. The first trend of Iranian researchers starting from Jamalzadeh and Hedayat pursue their works under the influence of nationalistic propaganda of Reza Khan, the king of Iran (1878-1944). In the year 1925, Reza Khan came to power and established Pahlavi monarchy that lasted until the Islamic revolution. Reza Khan, tried to modernize Iran by force and insisted on ethnic nationalism and cultural unitarism. In the light of such period, Hedayat writes several books in praise of ancient Persia and its culture, criticizing Iranian society and culture of being corrupted by Islamic and Arabic influences. Therefore his approach toward the folk-believes majorly are based on nationalistic, anti-Arab and anti-Islam interpretations. Few decades later the Islamic government on the contrary tried to redefine a religious narration for Iranian identity based on "Persian" language and "Shi'i" culture. Such

approaches in identity-making restricted researchers of working on or publishing marginal texts. Marzolph perhaps due to his political and personal concerns neglects the major work of Ahmad Shamlou, while even the website of the faculty of Literature & Humanities in University of Tehran could not forget to mention Ketab Kuche as a major work on Persian Folklore[1]. The reason Marzolph and many other official Iranian folklorists ignore the book could be traced back in Iranian newspapers during 1976-1978 period and his controversial impact on political and cultural sphere of Iranian society before and after the revolution. For example, in the year 1977, Ehsan Yarshater the founder and organizer of Iranica in the University of Columbia, invites Shamlou to work on Ketab Kuche and Shamlou first accepts and later rejects to work with him because of the close relationship existed between Yarshater the monarchy system of Iran. He tries to continue his research on the folklore independent from any political and religious forces, therefore his works always remain controversial [See. Appendix II].

1.2 Ketab Kuche, The book of alleyways

By the book Ketab Kuche, Ahmad Shamlou intended to index Persian folklore on the axis of Persian language. In his perspective if according to Heidegger the language is the house of existence, the existence of a nation puts its traces on the language. Traditions, beliefs, rituals, festivals and all the living elements of a culture will not be saved in monuments or just in the written texts but the language itself preserve their traces and memories. Therefore the language could be considered as the only capital of the people that carries the traces of their living existence. Ketab Kuche developed as an independent work, consists of 14 published volumes and nearly 80000 unpublished indexical entries and essays to demonstrate how the Persian folklore can be indexed by the Persian language and how Persian language remembers the lived life of a nation.

1 http://literature.ut.ac.ir/book-int

Ahmad Shamlou in the year 1961 publicly startes to work on Ketab Kuche by publishing a weekly section entitled Ketab Kuche dedicated to Iranian folklore in a literary magazine called "Ketab Hafte". In the year 1971 the National Academy of Persian language invites him to continue his research on Ketab Kuche within the context of the academy. In the year 1976 he includes other existing resources about the Persian folklore hold in the University of Princeton to his research work. In the year 1978 the first volume of the encyclopedia is published in Iran. In 1979 he uses another literary magazine and continue to publish a section dedicated to Ketab Kuche. In the year 1982 the book is banned for twelve year. And finally fourteen volumes of the work gets published by 2000 so far. [2]

The book's references come from previously published researches on Persian folklore (up to 1999), published diaries, collections of folk-materials, traveling monographs, collections of religious talks and books and so on. The reputation of Ahmad Shamlou, summoned lots of independent researchers to dedicate their personal collection or researches to his work.

Neither the published volumes, nor the manuscripts was digitalized until the official website of Ahmad Shamlou[3] with the sole right for digital publishing of the entire work of Shamlou started the digitalization process.

The book uses a semi-thesaurus structure to index lexical entries and also defines a tagging system to tag the entries according to their folk-genres or their lingual structures. Essays on each entry consist of genre or structure tags, descriptions(meaning, history, use, contextual information), photographs or drawings, variations of the entry, interlinks and references. The existing material covers three main dimensions of folklore: culture and anthropology, spoken and written language and the genres of folklore.

Tags among the encyclopedia appears in two different categories :

language structures (for example verbal compositions) : tarkibAt e masdari, tarkibAt e jomleyi

2 http://ketab-kuche.com/index.php?title=احمد_شاملو

3 Www.shamlou.org

and etc.

genres of folklore (for example fairytale or belief) : Matal, bAvar and etc.

The digitalization process of the work consist of the following running processes:

1. scanning, categorizing and archiving the entire published books and the manuscripts (approx. 75% of the task is done.)

2. transforming the archived materials to typed texts and indexing (approx. 40% of the task is done.)

3. online publishing of the typed content (approx. 12% of the task is done.)

4. semantic marking (in process)

According to the digital publishing contract for Ketab Kuche, the entire work must be published under a free license similar to Creative Commons or GNU public license [See Appendix I].

1.3 Researches on NLP and the Persian language

The Persian language, known as Parsi or Farsi, is a southwestern Iranian language within the Indo-Iranian branch of the Indo-European languages. It is the official language of Iran, Afghanistan and Tajikistan with more than 100 millions speakers. Four major phases are distinguished in its development, namely, Old, Middle, Classic and Contemporary(or Modern) Persian. *Old Persian* is represented in the inscriptions and cuneiforms of the Achaemenid era, dating from the 6th century B.C.E. *Middle Persian* is the language spoken during Sassanid empire and plenty of texts remained in middle Persian in the form of religious writings of Zarathushtrian texts. *Classical Persian* was used after Islamic-Arab invension of the Persia. The majority of the words in Classical Persian are rooted in Old Persian, Pahlavi and Avesta and also many other regional languages. Classical Persian for centuries was the the main vehicle of culture, literature and politics in an area extending beyond the

limits of the Iran, from modern Uzbekistan and the Indian Subcontinent in the east to the Caucasus and Anatolia in modern Turkey and until twentieth century was the only language used in textbooks and studied seriously. *Contemporary or Modern Persian*, now the official language of the three sovereign states of Iran, Afghanistan and the Republic of Tajikistan started by the influence and the contribution of Russian, French and English and many other languages.

The use of colloquial language in serious literature began around the turn of the last century and received a forceful social impetus with the advent of the Constitutional Revolution in 1906 [Keyvan]. Although Classical Persian still holds its status of high prestige and is still being studied at every level of education, the literary style of the last few decades has moved closer than ever to Contemporary or Modern Persian and the colloquial idiom and has been instrumental in making it gain currency. [Keyvan].

The language used in Ketab Kuche consists of Classical and Contemporary Persian with the influences of several regional dialects.

Several linguistic factors makes the language a difficult case for Natural language Processing:

Complexities of Perso-Arabic scripts

Directionality :

Written right to left, but numbers are written left to write

Letter features :

Arabic letters don't have one shape. Some join to adjacent letters. Some letters have up to 4 different shapes.

Orthographic variations

- The use of invisible characters, different standards for Arabic and Persian characters, several letters and the same pronunciation, ambiguity of Arabic suffix and Persian words.
- In Persian short vowels are not written and capitalization does not exist make the transliteration to be complex [Megerdoomian].

Due to the number of different dialects and variations of Persian language used in Ketab Kuche besides the grammatical ambiguities and complexities of the language, automatic syntax analysis of the text does not seem practical. Therefore the thesis will focus on the automatic lexical analysis of the text.

Several projects could be mentioned working on the lexical analysis of Persian texts, among them three projects hold by University of Tehran and Shahid Beheshti (FarsNet), Iranian High Council of Information (National Persian Lexicon: dadegan e melli), the University of Princeton (PersiaNet) seems could cover a database of around 20000 Persian words in thesaurus structures based on WordNet definition. Two of the mentioned projects produced open-source databases and applications.

WordNet official website defines the project as follow:

> *WordNet® is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. The resulting network of meaningfully related words and concepts can be navigated with the browser.* [Wordnet].

FarsNet describes:

> Nowadays WordNet is developed for more than 40 languages around the world. EuroWordNet, BalkaNet, AsiaNet and WordNets for Dutch, Italian, Spanish, German, French, Czech and Estonian are among them. Unfortunately some languages such as Persian (Farsi) lack such a semantic resource for use in NLP works [Shamsfard].

The semantic structure and the development of WordNet make it a practical option for automatic semantic tagging of Ketab Kuche in lexical level. The thesis will use the database of WordNet in FarsNet because the National Persian Lexicon remained less developed due to sudden stop of the project for three years by the reason of political changes in government.

1.4 About Ontology Development

An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an Ontology is a systematic account of Existence. For AI systems, what “exists” is that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge. Thus, in the context of AI, we can describe the ontology of a program by defining a set of representational terms. In such an ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names mean, and formal axioms that constrain the interpretation and well-formed use of these terms [Gruber].

According to W3C standard the Semantic Web is a vision for the future of the Web in which information is given explicit meaning, making it easier for machines to automatically process and integrate information available on the Web. The Semantic Web will build on XML's ability to define customized tagging schemes and RDF's flexible approach in representing data. RDF or Resource Description Framework is a standard model for data interchange on the Web. It consists of resources (nodes), and property/value pairs that describes the resource. A node can be any object pointed to by a URI, properties define the attributes of the node, and values can be either atomic values for the attribute, or other nodes. The first level above RDF required for the Semantic Web is an ontology language what can formally describe the meaning of terminology used in Web documents. If machines are expected to perform useful reasoning tasks on these documents, the language must go beyond the basic semantics of RDF Schema. The OWL Web Ontology Language is designed for use by applications that need to process the content of information instead of just presenting information to humans. OWL facilitates greater machine interpretability of Web content than that supported by

XML, RDF, and RDF Schema (RDF-S) by providing additional vocabulary along with a formal semantics.[4]

Wikipedia and the birth of semantic wiki systems opened a new research field that involves both the ontology development systems and the social media networks. Wikis as collaborative content management systems got to be popular by the success of WikiPedia project and they were used for example as knowledge management or community websites. Adding an underlying model of knowledge described in wiki pages (Ontologies for example), merges the the collaborative content management with power of the Semantic Web. After 2005 when semantic wikis were implemented seriously, many of such systems are being purposed, modeled, investigated and used.

1.5 The Digitalization Team

After several attempt during the life-time of Ahmad Shamlou and after his death to register a legal foundation for his works in Iran, the ministry of culture and the ministry of interior Iran both denied to give such permission. Therefore a group of individuals which were mentioned in his publication contracts decided to run a voluntary based underground organization. One of the main project of the organization was to digitalize and freely publish all the works of Ahmad Shamlou according to the digital publishing contract in the context of the Official Website of Ahmad Shamlou. The technical group who were the founders of the website mostly left Iran, one after another to continue their academic career. The author of the thesis , as the contract holder and the director of the digitalization project started to organize the digitalization team by the means of the website and open calls for collaboration. By the year 2008 around 23 books of Shamlou were published online. In the June 2008 two of Shamlou's son announced an auction for their father's belonging. Luckily Ketab Kuche's

4 http://www.w3.org/TR/owl-features/

manuscripts were not among the materials put for the auction; However it was probable that they try to do so. Therefore digitalization of Ketab Kuche came to a priority among 90 titles of the books written by Shamlou. The work was not realized until May 2009 when a group of ten individual volunteers were trained by the thesis author to pursue the project. The equipments including hard drives, scanners, computers were bought by the help of some individual donators. In Nov 2010, The Official Website of Ahmad Shamlou was filtered by Iranian government for not accepting the supervision of the ministry of culture on its contents [5]. However this filtering did not influence much on the number of the website visitors due to the common use of anti-filter softwares inside Iran. In December 2010, the first version of Ketab Kuche website was launched and the Voice Of America television made an interview with author about the process and future of the project [6]. The open-call for contribution published on the Official Website of Ahmad Shamlou and the interview encouraged around 70 individuals to join the project; however most of them left the work because of political or personal matters. The open-call is supposed to remain on the official website until the end of the project and every year some new members join the work. the minimum team members are enlisted below:

Project Manager and Coordinator: 1

Web Programmers: 2

Ontology Experts: 1

Scan and indexing group: 1

Typists : 2

General Editors : 1

Semantic Editor : 3

Documentation group: 1

Due to lack of the resources, political restrictions in Ahmadi-Nejad period in Iran, experts immigrations and distance management the project did not continue as expected. In the year 2010 the estimation for online publishing of all the work was around five years and in 2014 the estimation is

5 http://www.bbc.co.uk/persian/arts/2010/11/101119_l17_golshiri_shamlou_website_filter.shtml
6 https://www.youtube.com/watch?v=fsD5rJOofqU

about three more years. In the context of the project, the author of thesis worked mostly as the project manager but also helped in the process of web programming, ontology development and semantic editing.

Chapter 2: From Manuscripts to bits,

From bits to Paragraphs

27th July, 1998 is the day, Ahmad Shamlou, an internationally acclaimed Iranian writer embraces free digital publishing as the first Iranian writer who does so; however he did not recognize by the time that how difficult it can be. In 27th July, 1998 it seemed free digital publishing is what exactly made for a writer with social-anarchist background who suffered for years from massive state censorship. The contract he signed at that day defines a group of three young Iranian computer scientists as the publishers and a group of two, including his attorney and his wife as the project supervisors [see Appendix I]. Governmental changes in Iran few years before his death (24th July, 2000), permit some of his works to appear in paper and rapidly his works turn to be bestsellers in the whole country and therefore a major source of capital for his family and the traditional publishers. On one hand, he was prepared for the dangers of his political fights but on the other hand he never guessed his posterities

are unable to do so. Traditional publishers and his sons charge a battle against the digital and free publishing in Iranian newspapers for more than 10 years after his death; The main reason lies in the fact that free web publishing would challenge their income. The political fear brings other restrictions to the speed and the purposes of digital publishing. As a result, in the early years of the project, two of digital publishers left the project, exhausted; and one remained to continue the battle. In the very beginning of digitalization project, the three young digital publishers was forming part of Sharif University's Computing Center, a research center carrying "Persian Digital Library", and later "The FarsiWeb", a project which provided the first Iranian Unicode Standard and several tools and programs for information exchange between pervious Persian keyboard layouts and the Unicode.[7]

Even though, Ahmad Shamlou was among the the first generation of Iranian writers who acknowledges computer in his house, but the use of it for him remains as a type-writer. Most of his manuscripts are in paper, typed or handwritten. Few of Ketab Kuche entries were stored on his computer hard drive. Even a FoxPro program written by two Iranian programmers in order to help him in organizing Ketab Kuche entries, did not encourage him to avoid papers; Any comparison between the same entries on FoxPro database and the handwritings demonstrates that the paper-based manuscripts are more updated and completed.

Free digital publishing for him was an idealogical choice, and not an aesthetic shift. His posterities were also conservative in this sense; but they were not sharing the same ideological perspective and therefore for them Digital publishing could serve Print publishing just as an advertisement tool. However, despite the fact that Iran has been listed consistently among the bottom countries in violation of freedom of the press; digital publishing and blogging are the among the few ways to speak freely and read freely for the young generations of Iranians.

7 http://www.farsiweb.info/report/stat1.html

2.1 Materials and the processes

The corpus of Ketab Kuche can be divided in three different categories:

1- *Manuscript Papers:* including about 80000 encyclopedic entries:

These entries are of two types: machine-typed entries and handwritings. (FoxPro entries were not updated and therefore were not of any use.)

2- *Already published volumes*: including 14 volumes; each volumes of about 600 pages.

3- *External references:* many of the manuscripts papers include references to other books in order to be cited inside each entry; therefore these references form part of the developing content.

All of those categories were subject to digitalization, because the print publisher avoided to provide any access to the files of the perviously published volumes in order to put pause on the project.

After investigating in several digitalization guidelines provided by university libraries (ex. UMass Amherst Libraries, Arizona State Library and so on) and museums (ex. Canadian Museum of Civilization, National Museum of Australia and so on), the digitalization team came across the following definitions:

File Type	Published Books	Handwritings	Cited Works
Master Files	**Resolution:** 200 dpi	**Resolution:** 600 dpi	**Resolution:** 200 dpi
	Format: Uncompressed TIFF	**Format:** Uncompressed TIFF	**Format:** Uncompressed TIFF
Access Files	**Resolution:** 200 dpi	**Resolution:** 300 dpi	**Resolution:** 200 dpi
	Format: 8 bit grayscale	**Format:** JPEG 8-10 on a 1/10 scale (high)	**Format:** JPEG 8-10 on a 1/10 scale (high)
Thumbnails	Not Used	**Resolution:** 72 dpi	**Resolution:** 72 dpi
		Format: 4 bit grayscale, 8 bit color	**Format:** 4 bit grayscale, 8 bit color

Table 1: Digitalization Standard

The scanned files were stored according to following rules:

Published Books:

- Master File Name: KucheTIF_<Alphabet>_<Volume>_<PageNo.>_Date
- Access File Name: KucheJPG_<Alphabet>_<Volume>_<PageNo.>_Date

  For example an access file for the page 345 of the volume 1 corresponding to letter "ch" in Persian alphabet and scanned in 2009/01/11 has the following name: *KucheJPG_A_01_345_20090111.jpg*

Handwritings:

- Master File Name: HNDTIF_<3 Beginning letters of the Entry>_<No. Entry>_Date
- Access File Name: HNDJPG_<3 Beginning letters of the Entry>_<No. Entry>_Date
- Thumbnails: HNDTMB_<3 Beginning letters of the Entry>_<No. Entry>_Date

  For example the master file for the entry manuscript no. 21076, corresponding to word "zakhm" in Persian and scanned in 2010/10/22 has the following name: HNDTIF_*ZAK*_21076_*20101022.tif*

Cited Works:

- Master File Name: CITTIF_<5 Letters of the Book Name>_<No. Entry>_<Page No.>_Date
- Access File Name: CITJPG_<5 Letters of the Book Name>_<No. Entry>_<Page No.>_Date
- Thumbnails: CITTMB_<5 Letters of the Book Name>_<No. Entry>_<Page No.>_Date

  For example the thumbnail file for the entry manuscript no. 1376, cited a book called "DastAn-hA-ye Shegeft" and scanned in 2010/07/29 has the following name: CITTMB_*DASTA_01376_20100729.jpg*

Directory structure of the stored files follows the same orders; for example thumbnail files of the cited books are being stored in the following path : KetabKuche/CitedWork/CITTMB.

The digitalization team also contacted several Iranian companies working in the field of OCR softwares in order to speed up the digitalization process for the published volumes but the effort was not successful because the fonts used in published volumes were old-type-faces and were not supported by the recent softwares and as a consequence the price for a customization was more than the retype cost for all the materials; besides there were no guarantee for the accuracy of the customized output.

In Apr 2010 a group of individuals started a project by the permission of The Official Website of

Ahmad Shamlou to publish freely all the content of a literary and political journal Shamlou was publishing in 1980-1981[8]. The strategy they took was to use a Mediawiki-based website, publish scan of each page and ask volunteers to collaborate in a crowdsourcing manner for typing and editing the essays. The project successfully finished in Nov 2012. In the process of digitalization project of Ketab Kuche, the digitalization team had the example of this project and also performed a test running from Dec 2010 to Feb 2011 in a test subdomain to evaluate the speed and the accuracy of the volunteers. Therefore the crowdsourcing strategy in this stage seemed impractical in the case of Ketab Kuche. For the following reasons:

- The whole content of the journal is approximately equivalent to six volumes of Ketab Kuche and it took more than two years to publish them online. Clearly such strategy is more time consuming than a concentrated typing strategy. Even though it is cheaper.
- The essays in the journal do not contain inter-linking and complicated formats and mark-ups, while the content of Ketab Kuche necessitate absolute care and accuracy.

As a consequence, the digitalization team organized a group of two typists in order to transform scanned materials to word processable data.

Due to nationalization process of 80's in Iran during the war time and the lack of Persian support in most of early word processors, Iranian typists were trained for a national DOS-based word processor called Zarnegar. Windows operating systems in their primary versions didn't support Persian and therefore several software patches have been developed to include Persian language; but the variety of those patches did not correspond to any Keyboard standard, specially Unicode. Therefore the team first managed to train the typists and then to make sure they are using the same standard on their keyboards. A Ketab-Kuche desktop environment was developed to guarantee the accuracy of digitalization process. The environment architecture will be described in the following section.

8 http://irpress.org/index.php?title=صفحهٔ_اصلی

2.2 Data Structure

The digitalization procedure from scanned materials to machine-processable texts is a delicate procedure. One must be careful to do not lose any significant information. Data on scanned material are not meaningful unless being interpreted precisely. Each Item, that is to say, each encyclopedic entry, carries a structure. Such structure can be describe in several ways, one of which is a database. Such database structure must corresponds to data-structure of scanned materials.

In Ketab Kuche, each entry has a "title". The title is a phrase, an expression or a word.

For each entry, it assumes two distinct levels of categories:

1- Each entry is inherited from words in the dictionary and actually any word in the dictionary is the entrance to different use of it within folklore. For example the entry "Ab Atash Kardan" (meaning "To fire the water"), is inherited from "Ab" (Water), "Atash" (Fire) and "Kardan" (to do). Therefore one must be able to find this entry using each of these words, separately.

2- Each entry, represent an activity or a common way of expression in folklife. For example the entry "Ab Atash Kardan" is a manner the infinitive compounds are used and interpreted in folklore, "Ab Avardan" refers to a proverb and "No-Kardan-e Ab be mAh" refers to a ritual .

It seems Shamlou assumes a ritual and proverb ("life" and "language") in the same semantic order while one refers to an activity and another to an expression. Rituals for example involve more complex logical constructions while proverbs are more facile to be expressed in terms of logics and such approach in categorization can bring more complexities and confusions. Therefore we decided to place them in distinct categories.

Consequently we can define following category types for entries:

1- Dictionary vocabularies.

2- Folk syntactic structures, including: Curse, Proverb, Infinitive compounds, Phrasal compounds and so on.

3- Folk life, including: Ballads, Tales, Games, Rituals. Dreams, Beliefs and so on.

The textual body of each entry contains “descriptions”, “usage examples”, “internal references”, “external references”, “drawings” and “footnotes”.

In the case of the handwritten manuscripts, external references, drawing, and footnotes remain incomplete and must be cited from other references mentioned within the text.

The above mentioned information provide us a perspective to implement a database for the first steps of data entry.

2.3 Database development

Several platforms exist for database development but in the case of Ketab Kuche we decided to make it as simple as possible because the data entry was supposed to be distributed and the staff were nonspecialized typists, unfamiliar with complicated interfaces and programs. In the case of Iran, Windows is the most common OS and everybody has a Microsoft Office installed on his Windows; Therefore we used Microsoft Office Access databases. The following diagram shows the Database Structure created for the purpose of Data Entry :

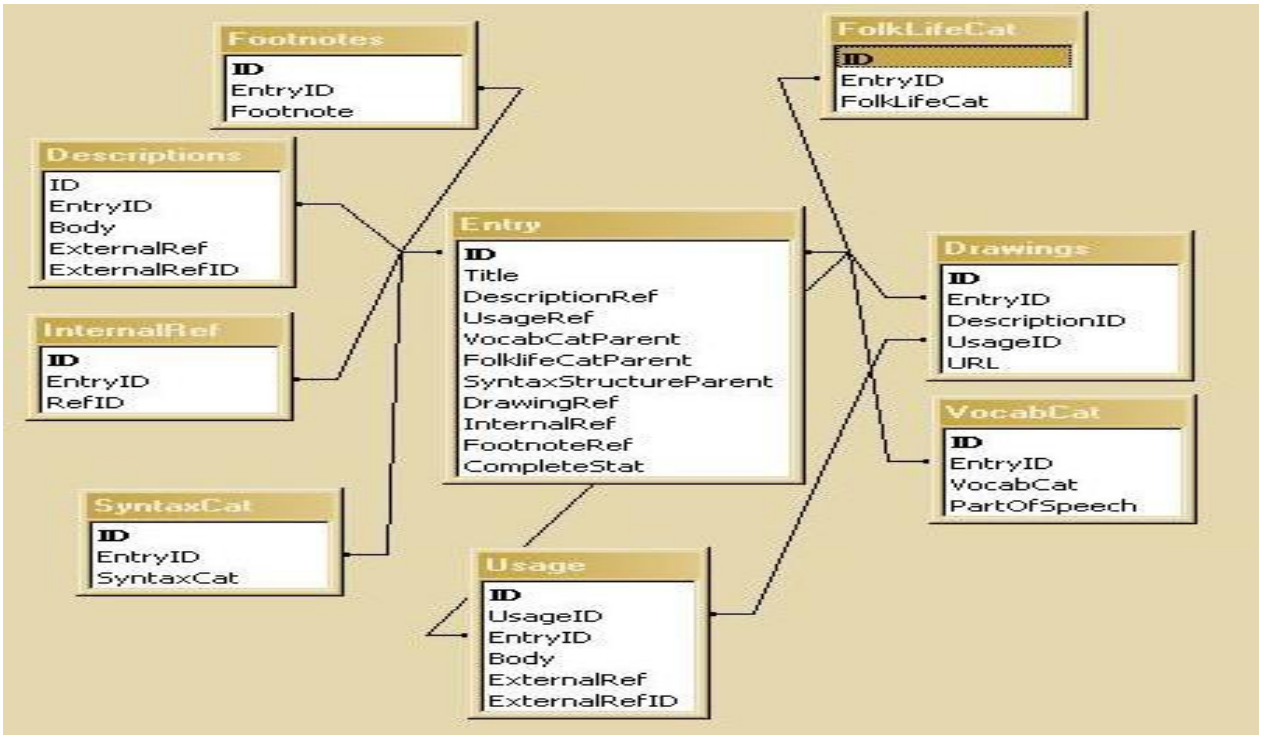


Figure 1: Data Entry Database Structure

The structure above assumes that for each Entry there could be many Vocabulary Categories, Usage Descriptions, Syntax Categories, Descriptions, FolkLife Categories, Drawings assigned to either a usage and/or a description, Footnotes, Internal References and External References.

The interface designed for this database was simple and user-friendly. Microsoft Access provides plenty tools for queries and views.

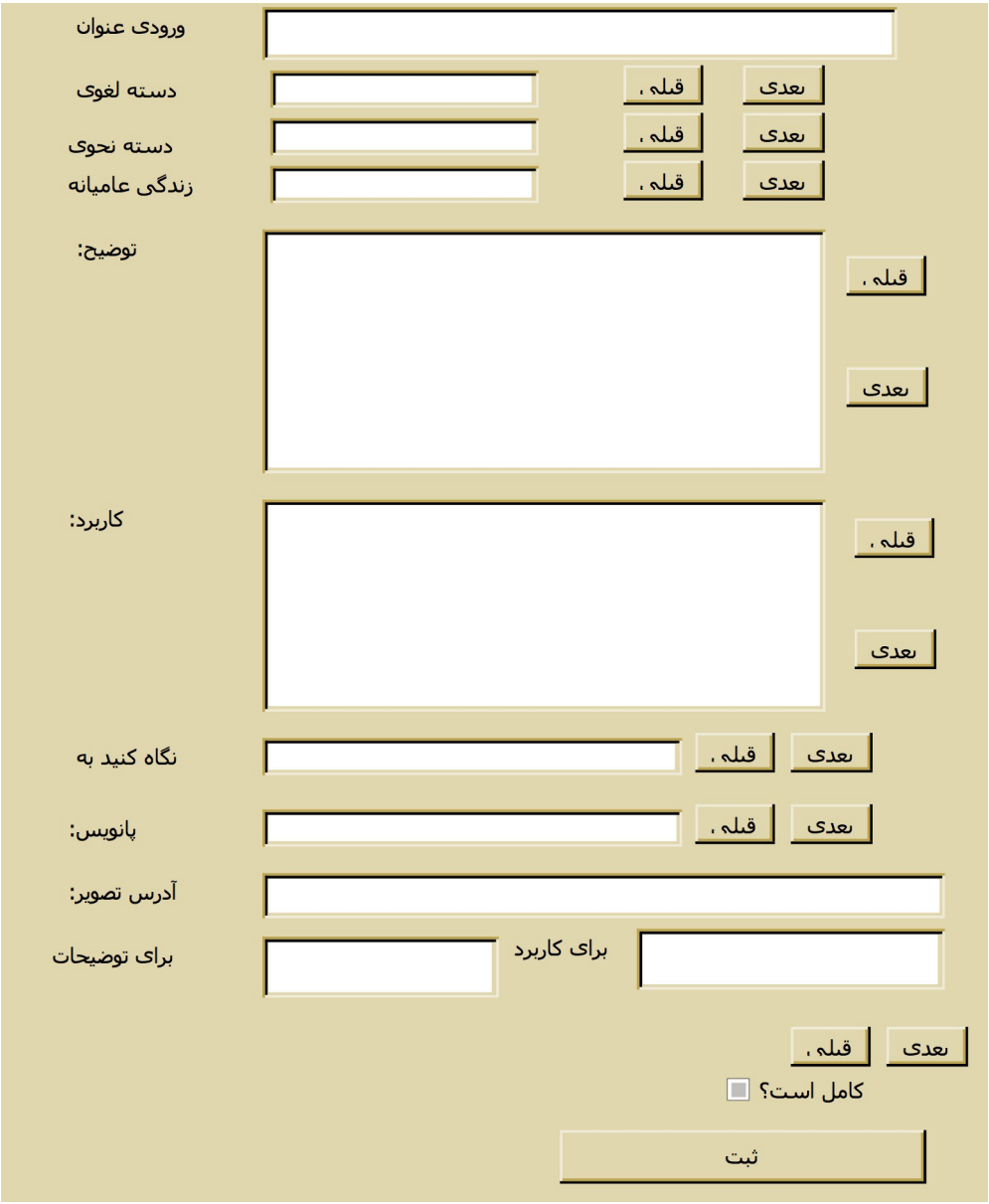


Figure 2 Data Entry Form

Microsoft Access fails on large amount of data; therefore the data entry staff were obliged to empty data rows in all the tables but in "FolklifeCat", "SyntaxCat" and "VocabCat". They had an alphabetical table of Entries with their updated "ID"s and whenever necessary they would use it for "InternalRef"s.

Data were collected weekly from the staff and an editor used three VBA modules on the Microsoft

Access databases in order to *Merge*, *Stem* and *Export* the data:

- Merge function makes a unique database and update indexical keys inside each table.
- The intended web-platform for presenting the Encyclopedia to the public audience were MediaWiki; a free and open-source wiki software tested on WikiPedia and many other wiki platforms with strong emphasis on multilingualism and internationalization. Therefore the export module provides a migration between the Microsoft Access database and the MediaWiki mysql database. The export module provides a CSV file which can be imported into MediaWiki's database by the use of CSVLoader, an AutoWikiBrowser plug-in that allows creating and updating articles using CSV data files.

The following snapshot shows the result of export module for an entry on MediaWiki platform:

Figure 3 : MediaWiki based website, ketab-kuche.com

2.4 Natural language processing in data entry

In the filed of NLP, Persian is a challenging language: Its Morphology is complex and the grammatical rules are complicated. One of the techniques that can increase the speed of information

retrieval in data-inquiry is storing "Dictionary Vocabulary Category" entries according to the roots of the words by using a Stemmer software. In the essay representing "Bon, the first Persian Stemmer" the authors describes:

> "In Natural languages, we can find limited words that are syntactic' roots of the other words. In an Indo-European language like Persian, a typical word contains a stem (root) which refers to central idea or meaning, and certain affixes have been added to this stem to modify the meaning and/or fit the word for its syntactic role.
>
> Stemming is a widely used method of word standardization designed to allow the matching of morphologically related terms. If, for example, a searcher enters the term stemming as part of a query, it is likely that he or she will also be interested in such variants as stemmed and stem. Stemmers are softwares that extract stems of word automatically.
>
> In natural language processing and other fields such as information retrieval (IR), Stemmers play an important role. In IR using stemmed words instead of the original words, could increase the level of the exhaustivity of indexing, and could contribute as much as 15 percent to increasing overall performance. Also stemming reduce the size of indexing files. Since a single stem typically corresponds to several full terms, by storing stems instead of terms, compression factors of over 50 percent can be achieved" [Tashakori, 487].

Despite the importance of Stemmers in NLP and IR, it is still difficult to find any Open-Source and effective Stemmer for Persian. The only working Open-Source Stemmer was provided by John Dehdari, the professor of linguistics in the Ohio State University. The script called PerStem is written in Perl using regular expressions substitutions to separate inflectional morphemes, and optionally remove affixes. The efficiency of the stemmer checked on the corpus extracted from Hamshahri Newspaper, was approximately 73%. [Jadidinejad] The perl script of PerStem was called through a VBA module but the result on Ketab Kuche entries didn't show more than 40% of efficiency. Clearly because the perl script was written for standard Persian and not for compound and complicated non-standard expressions. However, the team decided to use it in order to provide a sample stemming

data for the editing staff.

2.5 Digitalization Pipeline

According to the pervious sections of this chapter, the digitalization process of the project pursues the following steps:

Agent	Procedure	Expecting Results
Scanning Staff (two person)	Scan/Organize	Preservation of the Paper Materials
Typists (two persons)	Typing the entries using the type framework	Word-processable data
Editors (one person)	Merging	Providing an updated unified database
	Stemming	Providing a complete list of stemmed words for Vocabulary Category
	Charset Standardization	Transforming all the textual data to UTF8 Persian Standard
	Exporting	Importing CVS data to the MediaWiki Online Database in www.ketab-kuche.com

Table 2: Digitalization Pipeline

Chapter 3: From Paragraphs to text:

Basic Ontology Development

Not long ago, but before the digital era and the invention of search engines any comparative study on the use of a simple phrase in the works of Rumi and Ferdwosi, two major Iranian poet, was truly time-wasting. Now, a simple search in their works can give us enough information for such a study. Speaking of such studies we are facing the question of Information Retrieval or IR. In the year 2001 when Berners-Lee, Hendler and Lissila were writing about a dream, the Semantic Web dream, perhaps they could not imagine how many people, in different languages share the same dream. Berners-Lee writes :

> "F*or the Semantic Web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning*" [Berners-Lee].

"Interoperability", or the capability of different technological systems to exchange data, is the core of

such dream; and for that, two systems must share the same vocabulary in order to be able to communicate.

According to World Web Consortium (W3C) :

> *"The Semantic Web is about two things. It is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing."* [9]

Semantic Web uses ontologies and some mark-up languages to model and represent information so that machines and humans could use them co-operationally. In Philosophy, Ontology is a study of being in general and it deals with the basic categories of being, becoming, existence or reality and their relationship. In computer science, an Ontology is defined as a formal and explicit specification of a shared vocabulary or conceptualization which can be used to model a domain of knowledge [10]. Usually a graph structure is used to represent Ontologies. The graph consists of :

> *" 1. a set of concepts (vertices in a graph),*
>
> *2. a set of relationships connecting concepts (directed edges in a graph), and*
>
> *3. a set of instances assigned to a particular concept (data records assigned to concepts or relations)." [Caliusco]*

Although within philosophy ontologies are used to describe world, classify and categorize it, in Natural language processing the objective of ontologies are to model lexical and domain knowledge and in Semantic Web they are supposed to provide semantics for web resources. [11]

There are several ontology languages for encoding and representing ontologies. One of which is

9 http://www.w3.org/standards/semanticweb/data
10 http://tomgruber.org/writing/ontolingua-kaj-1993.pdf
11 http://www.ida.liu.se/~janma/SemWeb/Slides/ontologies1.pdf

OWL, Web Ontology Language, that uses formal semantics and RDF/XML-based representations for the Semantic Web and is endorsed by W3C, World Wide Web Consortium.

RDF defines a set of triples as the core structure of its abstract syntax. each triple consists of a subject, a predicate and an object. An RDF graph is a set of these triples connecting a Subject to an Object by the predicate. An RDF graph can have three types of nodes: IRIs, literals and blank nodes. IRIs are internationalized resource identifier and are defined as a generalization of uniform resource identifier or URI, a string of characters used to identify a name of a resource. Literals are language-tagged strings consisting a lexical form and a datatype IRI. Literals are used for values like strings, numbers and dates. In RDF graphs, any IRI or literals, denote some resources. Resources can be anything like documents, physical things, concepts and so on. Any RDF triple represents a RDF statement which explains some relationship, indicated by the predicate, holds between the resources denoted by the subject and object. RDF is not limited to describe just internet-based resources and URIs in RDF can be absolutely anything. [12]

An OWL ontology consists of a sequence of annotations, axioms, and facts. An Ontology can have a name and annotations are being used to record authorship and other associated information and also references to other ontologies. Facts contain either information about a particular individual, in the form of classes that the individual belongs to plus properties and values of that individual; or are used to make individual identifiers be the same or pairwise distinct. Axioms are used to provide information about classes and properties.[13]

Tools available for developing ontologies are also diverse. One of which is Protégé, an ontology IDE developed by Stanford University. Generally IDEs or interactive development environment provide an editor for source-codes, automation tools to compile and run the source-code and a debugger. Protégé offers an interactive graphical interface. It is open-source and there are many plug-ins available to extend its standard functionality. In Ketab Kuche project we used Protégé for ontology development.

12 http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/
13 http://www.w3.org/TR/owl-semantics/syntax.html

3.1 Mapping an Encyclopedia to a Thesaurus

Ketab Kuche, by definition is an Encyclopedia. That is to say, it contains encyclopedic entries and their definitions. Each encyclopedic entry has inter-references to other related entries within the encyclopedia. In a very common sense an Encyclopedia is a dictionary in many volumes, providing detailed information on a subject and often is organized alphabetically. Ahmad Shamlou, while looking for a structure to organize his articles and entries on Persian Folklore, followed the advices of a friend to use a thesaurus-like structure. Thesauruses in contrast to dictionaries which list words in alphabetical orders, provide grouping of words according to similarities of their meaning and contain synonyms and sometimes antonyms. However Shamlou failed to structure the work as a thesaurus. From the concept of thesaurus, he borrowed indexing and organizing entries under simple "nouns, verbs, adverbs and adjectives" and usage categories, representing language-uses or a folklife activities. That is to say, an entry like "az dast dAdan" (literary : to give from hands), meaning "to lose", is indexed under both "dast" (hand: a noun) and dAdan (give: a verb).

As a consequence, the entries in Ketab Kuche, represent taxonomies due to their classification into ordered categories, but the book is far from being a thesaurus; even though entries includes information about synonymies. The process of making a basic ontology for Ketab Kuche consists of transforming the book into a thesaurus with hierarchical information and the next step will be the ontology development. International Organization for Standardization (ISO) defines a thesaurus as "the vocabulary of a controlled indexing language, formally organized in order to make explicit the a priori relations between concepts (for example 'broader' and 'narrower')." Therefore a thesaurus offers Equivalence, Hierarchical and Associative relationships between lexical units.

The entries in Ketab Kuche, demonstrate equivalence and associative relationships but they lack hierarchical relationships.

In this sense Princeton University's WordNet by definition creates a combination of dictionary and thesaurus. In WordNet words are grouped into sets of synonyms called synsets which is a weak notion of synonymy. The above mentioned 'broader/narrower' relationship in WordNet is described by super-subordinate relation(hyperonymy, hyponymy or ISA relation) between synsets. It's the most common relation among synsets and is a transitive relation. Verbs and Nouns are described hierarchically. Adjectives are organized in terms of antonymy and there are only few adverbs in WordNet. It neglects prepositions, determiners and other function words. [WordNet]

WordNet includes the following semantic relations:

> *"Synonymy is WordNet's basic relation, because WordNet uses sets of synonyms (synsets) to repre- sent word senses. Synonymy (syn same, onyma name) is a symmetric relation between word forms.*
>
> *Antonymy (opposing-name) is also a symmetric semantic relation between word forms, especially important in organizing the meanings of adjectives and adverbs.*
>
> *Hyponymy (sub-name) and its inverse, hypernymy (super-name), are transitive relations between synsets. Because there is usually only one hyper- nym, this semantic relation organizes the mean- ings of nouns into a hierarchical structure.*
>
> *Metonymy (part-name) and its inverse, holonymy (whole-name), are complex semantic relations. WordNet distinguishes component parts, substantive parts, and member parts.*
>
> *Troponymy (manner-name) is for verbs what hyponymy is for nouns, although the resulting hierarchies are much shallower.*
>
> *Entailment relations between verbs are also coded in WordNet." [Miller]*

There are several project running to adopt WordNet structure in Persian. For example PersiaNet is strictly on a volunteer basis and developed a web-based lexicographer's interface. There is no publicly published output of the project available yet. There is also an effort for Automatic Persian WordNet Construction trying to use a Persian WordNet constructor which consists of Word Translator, Related Word Extractor, Synset Extractor and Synset Selector. [Montazery]. No published result of this

project is also available. The only working Persian WordNet is held by Shahid Beheshti University of Tehran under supervision of Dr. Mehrnoush Shamsfard. An open-source version of FarsNet 1.0 is available for public and FarsNet 2.0 is available upon request for academic use.

FarsNet project uses two Persian monolingual dictionaries, two annotated databases of Persian texts, three bilingual English-Persian dictionaries with more than 200000 words, one Persian Thesaurus and one dictionary of Persian synonyms and antonyms. Farsnet 2.0 could represent a database of 30.000 entries organized in about 20.000 synsets.[14]

Using FarsNet we could organize automatically parts of Ketab-Kuche database according to WordNet class hierarchy:

Synset
AdjectiveSynset
AdjectiveSatelliteSynset
AdverbSynset
NounSynset
VerbSynset

WordSense
AdjectiveWordSense
AdjectiveSatelliteWordSense
AdverbWordSense
NounWordSense
VerbWordSense
Word
Collocation

Problems arise not only because of the limits of FarsNet in number of synsets, but also because of non-homogeneity of entries in Ketab Kuche: Some entries are simple words, some compound words and some are expressions (ex. Proverbs or lines of a poem) and sometimes an expression have inconsistent meanings, for example, a compound verb like "Ab bordan" (literary : To carry water), means "to cause thirst" and "to cost" and "to have some hidden intention" and specially in such

14 http://mailman.uib.no/public/corpora/2013-March/017772.html

cases, human-interaction is necessary.

To solve this problem we had to exclude entries representing an expression, or a line of poem from being listed in synsets. Compound words is to be revised manually. We used Visdic software for dictionary editing written by Tomas Pavelek which was used and tested before in Balkanet and FarNet projects.

3.2 From Thesaurus to Ontology

Even though W3C offers an OWL and also a RDF representation of WordNet but one must be careful in using WordNet as an ontology. The class structure of WordNet imported in Protégé demonstrate the following relationships:

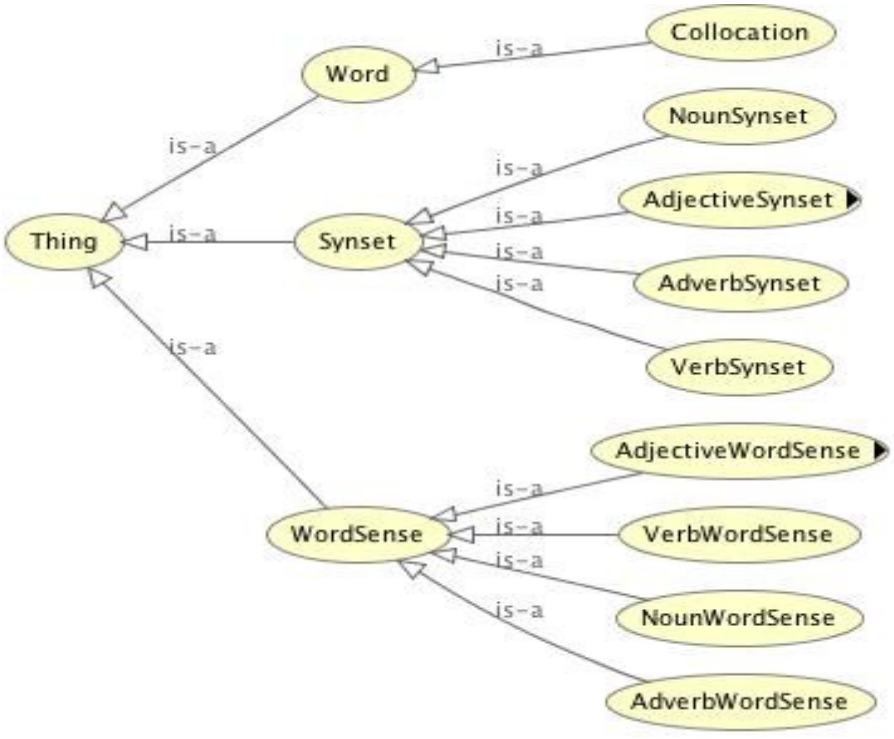


Figure 4: WordNet Schema

A study done by Aldo Gangemi, Roberto Navigli and Paola Velardi working on OntoWordNet project describes the problem:

> "WordNet is serviceable as an ontology (in the sense of a theory expressed in some logical language) if some of its lexical links are interpreted according to a formal semantics that tells us something about the way we use a lexical item in some context for some purpose. In other words, we need a formal specification of the conceptualizations that are expressed by means of WordNet' s synsets . A formal specification requires a clear semantics for the primitives used to export WordNet information into an ontology, and a methodology that explains how WordNet information can be bootstrapped, mapped, refined, and modularized" [Gangemi, 2003].

In another paper [Gangemi, 2003] they count WordNet's ontological problems naming different confusions:

- Confusing Concepts and Individuals
- Confusing Object Level and Metalevel
- Heterogeneous Levels of Generality

Therefore according to the study three conditions must be fulfilled in order to proceed to an ontology for WordNet: "

- *Logical commitment: WordNet synsets must be transformed into logical types, with a formal semantics for lexical relations. The WordNet lexicon is also separated from the logical namespace.*
- *Ontological commitment. WordNet is transformed into a general-purpose ontology library, with explicit categorial criteria, based on formal ontological distinctions.*
- *Contextual commitment. WordNet is modularized according to knowledge- oriented domains of interest. The modules constitute a partial order*
- *Semiotic commitment. WordNet lexicon is linked to text-oriented (or speech act- oriented) domains of interest, with lexical items ordered by preference, frequency, combinatorial relevance, etc" [Gangemi, 2003].*

To solve the problem they started to map WordNet into DOLCE a top-level ontology or foundation ontology. But the ontology suggested by them are highly complicated. Other upper ontologies as well shows the type of complexities. Since Ketab Kuche contains a large amount of uncertain entities and the development of FarsNet is incomplete and corresponds just to 5% of its English equivalent, using

a top-level ontology appears very risky and involves more human-interactions, and therefore more time-resources would be used.

As a consequence, mapping WordNet concepts to a light-weighted ontology that does not carry the WordNet problems and conceptual confusions can give us the possibility to resolve this problem at this stage. W3C suggests SKOS as a standard model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, 'folksonomies', other types of controlled vocabulary, and also concept schemes embedded in glossaries and terminologies.[15]

Studies have been done on the possibilities of mapping between WordNet and SKOS; for example Mark van Assem in his thesis on "Converting and Integrating Vocabularies for the Semantic Web" suggests an approach for such a conversion. W3C also refers to John M. Linebarger of Sandia National Laboratories's results on the conversion of WordNet 2.0 to SKOS and describes:

> "The central class of SKOS is **skos:Concept**. Its instances are connected using the **skos:broader/skos:narrower** properties. To each concept one can attach exactly one skos:prefLabel and zero or more **skos:altLabels**.
>
> The term "mapping" can have two meanings in this context. In the first meaning, the schema of WordNet (i.e. its classes and properties) is mapped to the SKOS classes and properties using **rdfs:subClassOf**, **rdfs:subPropertyOf**, **owl:equivalentClass** and **owl:equivalentProperty**. This is only possible without loss of information if the WordNet schema is equal to or is a strict specialization of SKOS. In the second meaning, a set of rules is specified that converts WordNet into instances of the SKOS schema. This is a more flexible approach and allows for more complex mappings (mappings other than property/class equalities and strict specialization).
>
> A first choice concerns what WordNet class(es) to map to **skos:Concept.**"[16]

15 http://www.w3.org/2004/02/skos/
16 http://www.w3.org/wiki/SkosDev/DataZone

SKOS also can be used in other upper ontology systems like CYC. Therefore, upon improvement of FarsNet and Ketab Kuche data entry, it would be possible to merge SKOS-based ontology of Ketab Kuche into a sophisticated top-level ontology.

The general schema of SKOS as an ontology is simple:

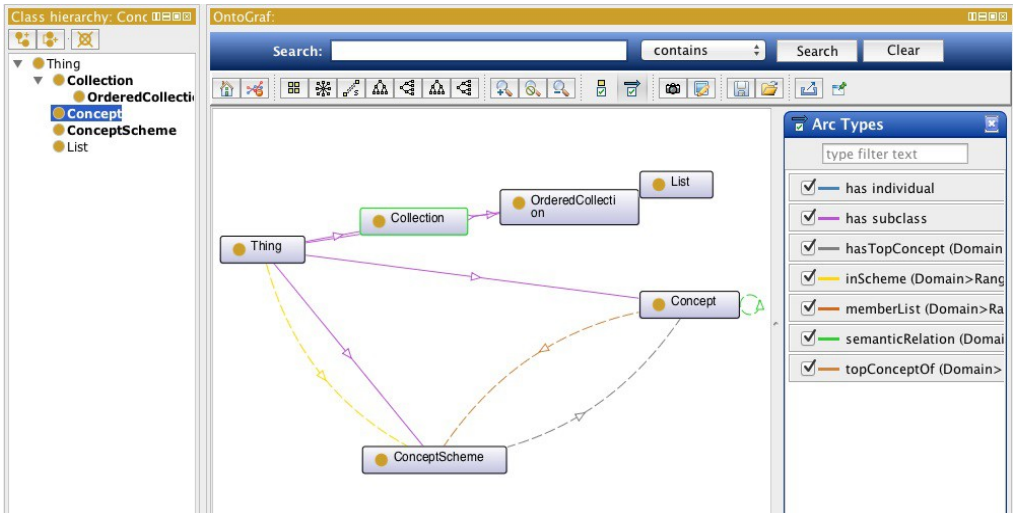


Figure 5: SKOS core classes

3.3 The Basic Ontology of Ketab Kuche

Even though SKOS representation of WordNet reduces many confusions, still it is not sufficient. For example SKOS's **prefLabel**, **altLabel** and **hiddenLabel** do not distinguish between syntactic preference (like canonicalForm etc) and pragmatic preference, that is whether the term is preferred for terminological reasons.[17] And this is the reason we take lemon project and its model into account.

lemon is a RDF model for representing lexical information relative to ontologies developed in the Monnet project to be a standard for sharing lexical information on the semantic web.

According to **Lemon cookbook**, the model consists of a core path defined as:

> "**Ontology Entity:** The ontology entity that describes the meaning of the concept in a language-independent manner

17 http://lemon-model.net/lemon-cookbook/node5.html

Lexical Sense: This object is used to attach all meaning-dependent properties of the word or term.

Lexical Entry: This represents the word or term itself.

Lexical Form: This object is used to describe a single form (e.g., plural, perfect, etc.) or an entry

Written Representation: The actual string that the lexical entry is realized as."[18]

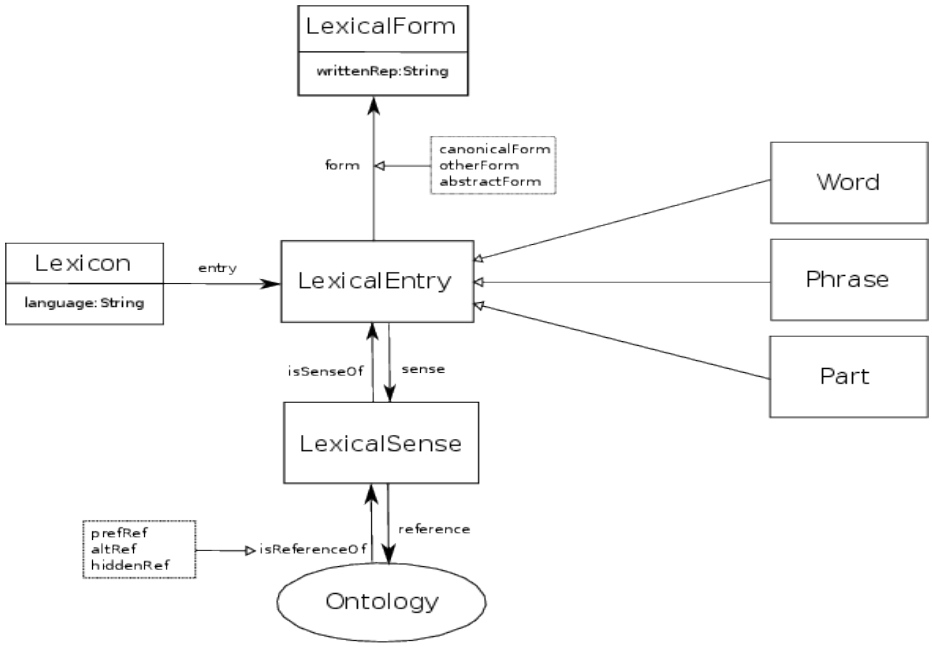


Figure 6 : Lemon Core

Lemon model is being used by DbPedia, Eurosentiment, Dbnary and is referred widely in ontology research field. DbPedia can be considered as a successful application of lemon model. It uses Wikipedia's categorization and its entry Infoboxes to improve WordNet's definitions and reduce ambiguities. Unfortunately Persian Wikipedia covers approximately 9% of English Wikipedia and it lacks structural qualities of the English version; besides most of the entries are rough translations of the English articles. Therefore experiments like Dbpedia are not applicable in a realistic approach to the condition of Persian language in semantic web technologies.

Considering the fact that FarsNet project uses the English version as its main reference and applies the English-Persian translation from bilingual dictionaries would speak of a condition in Iranian society, called "Translatory Condition" By Ahmad Shamlou. One the main intention of Shamlou in collection Ketab Kuche was to fight against this condition. In the last years of his life, he retranslated

18 http://lemon-model.net/lemon-cookbook/node3.html

The Silent Don by Mikhail Sholokhov into Persian, a work which was translated years ago by a very acclaimed Iranian translator and was widely read.[19] The reason he does so, lies in his conception of "naturalness" within the language. He proves that a natural translation of The Silent Don is possible if we avoid using the cliches inside bilingual dictionaries and instead try to apply the colloquial language and other norms used in the streets(for example the normalized translation of the word "waiting" to Persian is "entezAr", derived from Arabic, the colloquial phrase with the same meaning is "cheshm be rAhi", literally means "to have an eye on the road"). FarsNet and Persian Wikipedia establish the translatory condition and that's another reason why Ketab Kuche necessitate more manual work, human interaction and a distributed approach.

One of the main conceptions in the development of lemon model resides in an application-oriented approach to Semantic lexicons, as is mentioned by Paul Buitelaar, "An ontology-based semantic lexicon would leave the semantics to the ontology, focusing instead on providing domain-specific terms and object descriptions in the ontology." [Buitelaar] That is to say most of the semantic lexicons utilizes general purpose over-representations and basically it is not necessary to define semantics within lexicons.

Ketab Kuche, according to the extent of its subject involves several domain-specific ontologies and lemon model provides us the possibility to establish semantic links between a term in lexicon and the domain-specific ontology of the term. Therefore, instead of converting FarsNet to SKOS we decided to use lemon model in our base semantic lexicon. From a lemon model it's not difficult to drive a SKOS model, as is described by lemon cookbook:

> *"lemon is designed to subsume most of the features of SKOS (Miles and Bechhofer, 2009), in particular the ability to state preferred labels and represent soft semantic relations. lemon uses the sub-properties of lexicalForm and isReferenceOf to more precisely capture the same semantics as SKOS's prefLabel, altLabel and hiddenLabel. The conversion is as follows:"*[20]

19 http://shamlou.org/?p=363

20 http://lemon-model.net/lemon-cookbook/node53.html

	Canonical Form	Other Form	Abstract Form
Preferred Reference of	prefLabel	altLabel	hiddenLabel
Alternative Reference of	altLabel	altLabel	hiddenLabel
Hidden Reference of	hiddenLabel	hiddenLabel	hiddenLabel

Table 3: SKOS to lemon conversion rules

An RDF representation of lemon model was used as the basic Ketab Kuche ontology[21]. The OWLViz visualization of the ontology is presented below:

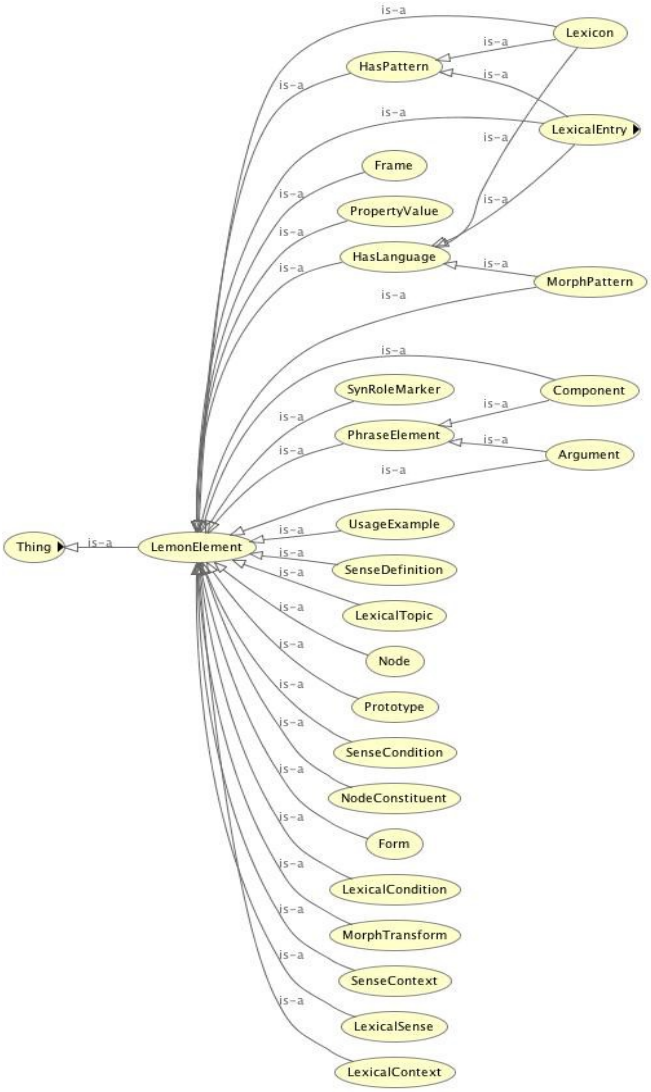


Figure 7: Lemon Ontology visualization

21 lemon-model.net/lemon.rdf

Chapter 4: Developing an extended Ontology

For Wittgenstein the limits of his language meant the limits of his world and for Shamlou commitment to the world and to the language were complementaries. In the essay about the translation of The Silent Don, he claims that half of the writers' commitment is the commitment they have to their language.[22] Ketab Kuche and its intended structure were supposed to reflect the mentioned commitment. That's why he tries to index Persian Folklore on the axis of the Persian lexicon. That is to say, if the limits of our languages define the limits of world, we might be able to index the world through the language and for him Folklore was more alive than the formal literature because it is originated from lived lives. Therefore Ketab Kuche reflects to several aspects of folklives: it contain games and their rules, it contains songs and lyrics, it contains stories about fictional or historical characters, or descriptions about some agricultural tools and so on.

The previous chapter tries to concentrate on the first design layer of the encyclopedia, that is to say on the lexicon and this chapter will focus on the content of each entry and their variations. Distinguishing such variations gives us the possibility to categorize different Domains involving in the

22 http://shamlou.org/?p=363

Ontology development. Next sections will be discussing on the Ontology development process and the aspects of extending Ketab Kuche.

4.1 Domains of Knowledge in Ketab Kuche

Appendix III represents some of the Ketab Kuche entries' variations. The categories assigned by Shamlou to each entry were supposed to classify the lexicon level of the encyclopedia. However even in that level, the category assignments are inconsistent, confusing and paradoxical. One of major problem occurs when there is no clear distinction between a lexical entry and its instances, for example a concept describing "expectation" and a riddle using the same concept are in the same lexical level in the lexicon, although the riddle must be an instance indicating to the use of the concept; therefore, classification of the entries into *classes of concepts* and their *instances* within the lexicon is inevitable.

Another problem of the book, as is mentioned before in Chapter 2, is in the level of the categorization of the lexical entries based on their "use of the language" or "use in the folklife". The book does not see any ontological difference between them. However they must be separated into distinct category levels.

Besides all, the book's category assignment does not reflect to most of the information each entry contains and just indicates to the significant aspect the author sees, neglecting others; for example an entry about a proverb is assigned to the category "proverbs" but it contains information about historical characters, architecture and sometimes even grammatical instruction on how to stem a verb in a regional dialect.

In order to solve these problems, we need to categorize the major domains of knowledge in the book.

Of course any text is source of several information comprehensible for human agent but we can always point out some major domains of the knowledge as the common uses of a text.

By analyzing the content of Ketab Kuche, at first we came out to the following list of categories:

1. Geographical:

Villages, Cities, Provinces and Countries. Hills and Mountains. Rivers and Valleys. Forests. Deserts and so on.

2. Material Life:

- Economy: Hunting and Fishing. Agriculture. Livestock. Jobs. Castes and Classes. And so on.
- Settlement: Houses. Building. Architecture. Bridges. Fountains. And so on.
- Tools: Furnitures. Cars. Animals and so on.
- Foods and Drinks and Cloths and so on.
- Joy and Free time.

3. Subjective Life:

- Language: Proverbs. Jokes and so on.
- Literature: Poems. Songs. Tales and so on.
- Ethics: Prayers. Curses and so on.
- Science: Disease and Remedies. Herbs. Astronomy and so on.
- Arts: Artifacts. Dances. Musics and so on.
- Mysteries: Black Magic. Dreams and Interpretations and so on.
- Sacred Materials: Places. Trees and so on.
- Ceremonies: Rain Prayers. Weddings. New Year and so on.
- Rituals: Religious special rituals like Ramadan and so on.
- Secret Life: Porn. Violence and so on.
- Social Life: Birthdays. Games. Weddings. Funeral and so on.

The categories enlisted here are extracts of different type of the content in the book, but of course the book is not categorized according to them and it covers few of them. Of course the categorization above must be standardized and must follow a common standard in order to be able to communicate with other applications in the subject. One of the must common classification for cultural material is OCM, Outline of Cultural Materials, widely used in social science. The Yale University's OCM, founded in 1947 within the context of HRAF is one the most used OCM in the area. As is described by them:

> *"The Outline of Cultural Materials, first developed by G.P. Murdock in the 1940s, is an ethnographic classification system on human behavior, social life and customs, material culture, and human-ecological environments. In the past this indexing system was used in the paper and microfiche versions of the Collection of Ethnography and now the OCM subject thesaurus serves the eHRAF World Cultures and eHRAF Archaeology databases."*[23]

Applying their coding to cultural materials included in Ketab Kuche, can work as bridge to other cultural and folk project across the world. The process of inserting OCM tags to Ketab Kuche entries can best be done by a mixture of automatic tagging and crowdsource editing of assigned tags. This needs providing a semantic mapping between the lexical dictionary and the OCM which is not studied yet by the project.

4.2 The Process of Ontology Development

There are several methodologies in Ontology Development, namely TOVE, Enterprise Model Approach, METHONTOLOGY, KBSI IDEF5, Ontolingua, CommonKADS and KACTUS, PLINIUS, ONIONS and many others.

The following model suggests different entities involving in the process of Ontology Development:

23 http://hraf.yale.edu/online-databases/ehraf-world-cultures/outline-of-cultural-materials/

Lexicon	Language	World	Cultural Materials
FarsNet, Ketab Kuche Lexical Entries	Encyclopedic Entry Ontologies: MediaWiki Ontologies	Folk Object Ontologies	OCM Lexicon: Definitions, Codes

Table 4: Entities interconnection in Ontology Development

Lemon model represents the mechanism of how these Ontologies work together, by separating Lexicon from Ontologies. The process for Folk Object Ontologies necessitates consulting Domain Fields experts on each subject. In some cases, there are Open-Source Ontologies available, but they must be customized for the project.

The following flowchart demonstrates the ontology development process in each domain field:

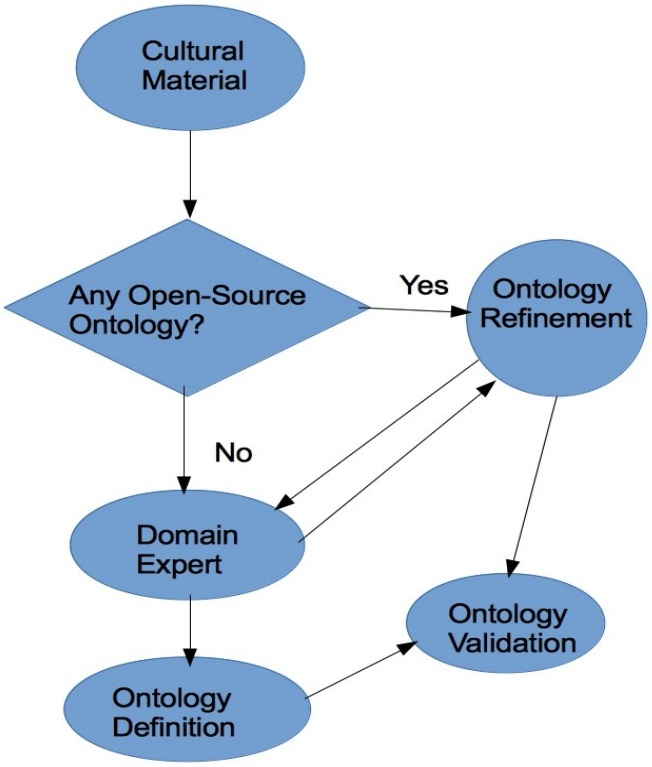


Figure 8 : Ontology Development Process

In this context, both process of Ontology refinement and Ontology Definition produce a preliminary

ontology which contains proto-concepts i.e. initial descriptions of kinds, relations and properties. The process of Ontology Validation tests the proto-concepts. The test process is a deductive validation procedure because the ontology structures are "instantiated" with actual data, and therefore it is possible to compare the result of the instantiation with the ontology structure. [Jones]

A very good example of an open-source Ontology is the Music Ontology that involves four other essential ontologies: "

- **FOAF**, a vocabulary for describing people, groups of people and organizations.
- **The Event Ontology**, a vocabulary for describing events
- **The Timeline Ontology**, a vocabulary for describing time intervals and instants on multiple (possibly related) timelines, e.g. an audio signal's timeline.
- **The FRBR ontology**, a vocabulary for describing works, expressions, manifestations and items and their relationships."[24]

The approach which is taken in Ketab Kuche extended ontology development is very similar to the one of "Semantic Kalevala"; but with one important difference: in the portal "CultureSampo—Finnish Culture on the Semantic Web" the design of the ontology is based on interpreting the text as a series of events because Kalevala is an epic and its narration is based on events and consequences, while in Ketab Kuche the narration of each entry is different from others and there is no other way by treating each text as a collection of cultural materials. In Semantic Kalevala "the text to be rendered is annotated using ontologies"[Hyvonen] and that's exactly one of the main goals of Ketab Kuche project.

The entities represented in Table 4, lead us to a model of three layer Ontologies:

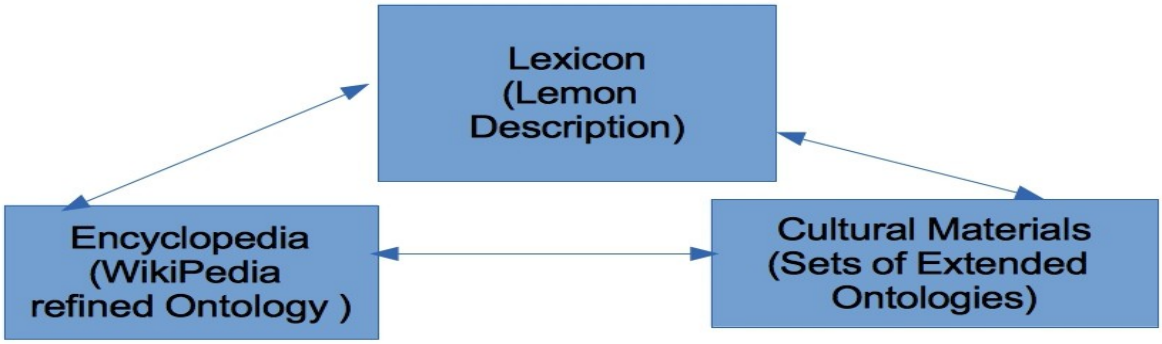


Figure 9 : Layers in Ontology Development

24 http://musicontology.com/docs/faq.html

4.3 Collaborative Ontology Deployment

The discussions in the pervious chapters were centralized on a concentrated approach to Ontology Development. However projects with the scale of Ketab Kuche mainly use collaborative approaches; as an example we can mention to FinnONTO (Finnish thesaurus and ontology service) or InPho (Indiana Philosophy Ontology).

The case of InPho is more close to one of Ketab Kuche, it is an effort in producing ontologies out of an encyclopedia. However the domain of philosophy seems more conceptualized than the Folklore; in another hand the articles forming the Stanford Encyclopedia of Philosophy (SEP) are well-formed texts following the regulations of academic essays. Ketab Kuche in this sense does not have any of these characteristics. The SEP is a "dynamic reference work" with expert authors who are enjoined to revise their articles on a periodic basis. To meet the needs of the SEP, InPho have chosen to focus their efforts on building and maintaining a "dynamic formal ontology", which they have dubbed the "Indiana Philosophy Ontology" (InPhO).[25]

The dynamism of InPho metadata engine is described in the image below:

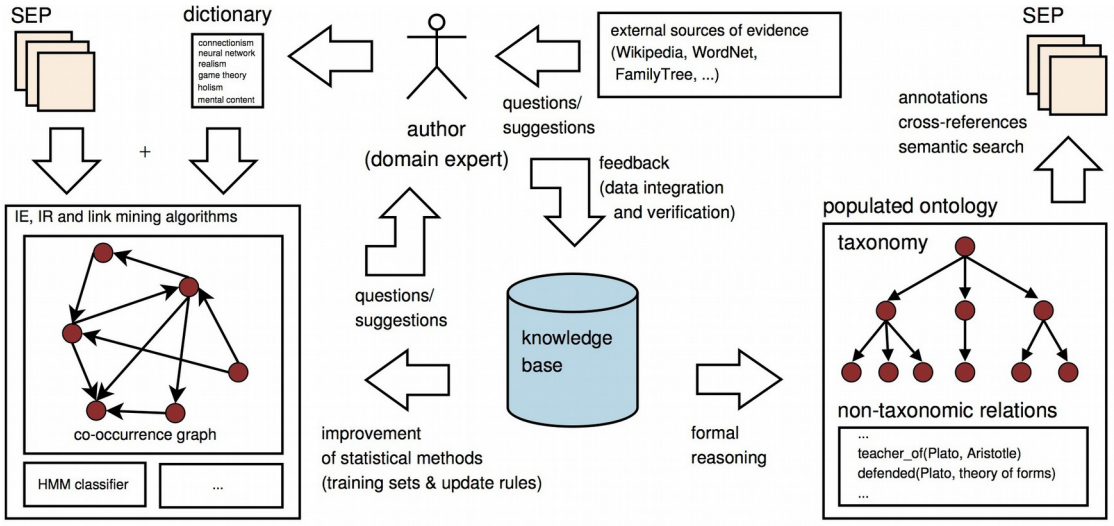


Figure 10: InPho meta-content generating engine

25 https://inpho.cogs.indiana.edu

As clearly is shown in the diagram, the dynamism is highly related to the set of their internal and external sources as well as the domain experts. That is to say, the ontology development in InPho is based on the collaboration between the system(and its internal or external resources), entry authors and the domain experts"[Niepert]

FinnONTO is a completely different case as is described by them:

"National Semantic Web Ontology Project in Finland (FinnONTO) defines its goal as follow:

> "The ambitious goal of FinnONTO is to lay a foundation for a national metadata, ontology, ontology service, and linked data framework in Finland, and demonstrate its usefulness in practical applications. In our vision, a conceptual semantic infrastructure is needed for the semantic web in the same way as roads are needed for traffic and transportation, power plants and electrical networks are needed for energy supply, or GSM standards and networks are needed for mobile phones and wireless communication. A solid, commonly agreed open infrastructure would make it much easier and cheaper for public organizations and companies to create interoperable intelligent contents and services on the coming semantic web. The infrastructure should be open source and its central components be maintained by the public sector order to guarantee wide usage and interoperability across different application domains and users. " [26]

FinnONTO provides a national level semantic web infrastructure and is based on centralized ontology services. The General Finnish Ontology YSO, is the heart of FinnONTO project. Also several other ontologies from different organizations e.g., various museums were involved in deploying the ontology.

Ketab Kuche project suffers from four main deficiencies that force the project to pursue a concentrated approach in ontology development:

> 1- The lack of National Standards and National Ontologies: Museums, National Institute of Cultural Heritage and many other organization mainly are acting as a propaganda service. For example in Dec. 2013 the director of National Institute of Cultural Heritage announced that in Nov. 2014 they will open the first Iranian

26 http://www.seco.tkk.fi/projects/finnonto/

standard museum without mentioning "what does he mean by a standard museum?"[27]. Therefore most of the cultural projects, like Ketab Kuche, can just rely on individual resources and not the governmental or institutional.

2- Most of the resources for Persian Folklore are not digitalized yet. The stories, the songs, historical characters, books and many other highly-important ontological references are not available online.

3- Ketab Kuche is an incomplete work. Shamlou died before finishing the work as a whole. Because of his political and health situation, the work never turned to fulfill the purposes of an encyclopedia.

4- Shamlou was not an expert in several domains of the knowledge within the folklore. Ketab Kuche remains as a massive effort of an individual.

In Dec 2010 the objective of the project was to perform a mixture of FinnONTO and InPhO project in ontology development as is mentioned in the project first public announcement[28]. That is to say a collaborative, dynamic and distributed approach to Ontology development. However in 2013 the above mentioned reasons forced the project to acquire an alternative concentrated approach.

27 http://danakhabar.com/fa/news/1166399/استاندارد-کردن-موزه%E2%80%8Cهامهم%E2%80%8C -ترین-برنامه-سازمان-اندازی-نخستین-موزه-استاندارد-کشور-تا-مهر-۹۳%E2%80%8Cمیراث-فرهنگیراه

28 http://shamlou.org/?p=525

Chapter 5 : Conclusion

The study intended to evaluate several strategies in digitalization and semantic representation of Ketab Kuche, an incomplete massive encyclopedia of Folklore written by late Ahmad Shamlou. The encyclopedia provides a proper case study on the difficulties and the realities of digitalization in developing countries like Iran. Many of the sophisticated systems for ontology development can not be realized when the standards and other dependent resources are missing.

Project like Ketab-Kuche in any other context would be pursued by means of a foundation and governmental or institutional fundings and resources. In the year 2011 the author of the thesis was authorized by the supervision committee for the works of Ahmad Shamlou to found a foundation in order to organize and publish his works outside of Iran. The foundation did not established because of the condition the author was living in exile within several countries from Finland to Czech Republic to Spain and now in Mexico and therefore the project was stopped for a period between Dec 2011 until Feb 2013. A population of nearly six million Iranians live outside of Iran and this

community can be considered of a great potential for such works. Establishing the mentioned foundation as an international foundation is planned for the year 2015. Hopefully it would be possible to acquire institutional and grass-root fundings for the work and therefore to speed up the project. One of the main benefits of such foundation is to remove the dependency the project has to the individuals.

The most basic component for ontology development is a lexical dictionary and the study provides a mechanism to manipulate and improve such a lexicon. Besides, it purposes a system for a fundamental ontology as the core of the system and introduces a model to extend it based on different involving domains of the knowledge.

Iran is a multicultural country and several languages live there. Ketab Kuche was intended to cover Persian folklore in term of Persian language. However it fails in the case of Persian speaking communities in other countries like Tajikistan, Afghanistan, Uzbekistan and India. There is no doubt that example of Ketab Kuche, if successful, can be used and extended in other Persian speaking countries.

One of the major objective of the study is to provide such a framework which could be useful for other regions and language communities of Iran as well as Persian speaking communities in other countries. To do so, the framework will be examined on a bottom-up approach, first in the north of Iran, in Mazandaran. A work-manual is provided and a group of volunteers are being trained to write an encyclopedia from the sketch on the basis of the ontology; meanwhile another team in Tehran pursues the publication of the existing Ketab Kuche and consults domain experts in specific domains of the folklore.

References

- Berners-Lee, T., Hendler, J. and Lissila, O. 2001. "The Semantic Web".Scientific American. 2001. http://www.scientificamerican.com/article/the-semantic-web/ (accessed 30 July 2014).
- Buitelaar, Paul. "Ontology-based semantic lexicons: mapping between terms and object descriptions". Ontology and the Lexicon: A Natural Language Processing Perspective. Cambridge University Press. 2010.
- Caliusco, M.L., Stegmayer, G. "Semantic Web Technologies and Artificial Neural Networks for Intelligent Web Knowledge Source Discovery: Chapter 2". Emergent Web Intelligence: Advanced Semantic Technologies. Springer. 2010.
- Gangemi, Aldo , Nicola Guarino, Claudio Masolo and Alessandro Oltramari. "Interfacing WordNet with DOLCE: towards OntoWordNet". Ontology and the Lexicon: A Natural Language Processing Perspective. Cambridge University Press. 2010.
- Gangemi, Aldo , Roberto Navigli and Paola Velardi. "The OntoWordNet Project: extension and axiomatization of conceptual relations in WordNet". On the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ···: Volume 1. Springer. 2003.
- Gruber, Thomas R. "A Translation Approach to Portable Ontology Specifications". Knowledge Acquisition, 5(2):199-220. Stanford University. 1993. http://lisas.de/~david/brunel/references/Gruber_ontolingua-kaj-1993.pdf (accessed 30 July 2014).
- Hyvonen, Eero, Joeli Takala, Olli Alm, Tuukka Ruotsalo, Eetu Makela. "Semantic Kalevala: Accessing Cultural Content Through Semantically Annotated Stories". Semantic Computing Research Group (SeCo). 2007. http://www.seco.tkk.fi/publications/2007/hyvonen-et-al-kalevala-2007.pdf (accessed 30 July 2014).

- Jadidinejad, A. H., Faribroz Mohammadi and John Dehdari. "Evaluation of Perstem: A Simple and Efficient Stemming Algorithm for Persian". Multilingual Information Access Evaluation I : Text Retrieval Experiments. Springer. 2010.
- Jones, Dean, Trevor Bench-Capon and Pepijn Visser. "Methodologies For Ontology Development". http://cgi.csc.liv.ac.uk/~tbc/publications/itknows.pdf (accessed 30 July 2014).
- Keyvan, F., Habib Borjian, Manuchehr Kasheff and Christiane Fellbaum."Developing PersiaNet: The Persian Wordnet". 3rd International WordNet Conference. Republic of Korea. 2006. http://semanticweb.kaist.ac.kr/conference/gwc/pdf2006/52.pdf (accessed 30 July 2014).
- Magoulick, Mary."FOLKORE INFORMATION SHEET: Folklore and Literature". Studies in Folklore. Georgia College. 2009. http://www.faculty.de.gcsu.edu/~mmagouli/folkloreinfo.htm (accessed 30 July 2014)
- Marzolph, Ulrich. "FOLKLORE STUDIES i. OF PERSIA". Encyclopædia Iranica. Center for Iranian Studies. Columbia University. http://www.iranicaonline.org/articles/folklore-i. December 15, 1999. (accessed 30 July 2014)
- Megerdoomian, Karine. "Finite-State Morphological Analysis of Persian". In Proceedings of the Workshop on Computational Approaches to Arabic Script-based Languages. Coling 2004, University of Geneva. August 28, 2004. http://www.zoorna.org/papers/Coling04workshop-PersianFSM.pdf (accessed 30 July 2014).
- Miller, G. A. (1995). WordNet: a lexical database for English. Communications of the ACM, 38(11), 39-41.
- Montazery, M. and Heshaam Faili. "Automatic Persian WordNet Construction". COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Association for Computational Linguistics. Stroudsburg. 2010.
- Niepert, Mathias, Cameron Buckner and Colin Allen. "A Dynamic Ontology for a Dynamic Reference Work". JCDL '07 Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries. ACM New York. 2007.

https://inpho.cogs.indiana.edu/papers/2007NiepertBucknerAllen.pdf (accessed 30 July 2014).

- Shamsfard, Mehrnoush, Akbar Hesabi, Hakimeh Fadaei and others. "Semi Automatic Development of FarsNet; The Persian WordNet". Global Wordnet Conference. India. 2010. http://www.cfilt.iitb.ac.in/gwc2010/pdfs/57_FarsNet__Shamsfard.pdf (accessed 30 July 2014).
- Tashakori, M., Meybodi, M. and F. Oroumchian. "Bon: First Persian Stemmer". Lecture Notes on Information and Communication Technology, LNCS 2510. Springer Verlag. 2002. http://ceit.aut.ac.ir/~meybodi/paper/Tashakori-Meybodi-Orumchian-BON1.pdf (accessed 30 July 2014).
- Volker, Johanna, Peter Haase and Pascal Hitzler. "Learning Expressive Ontologies". Ontology Learning and Population: Bridging the Gap Between Text and Knowledge. IOS Press. Amsterdam. 2008.
- WordNet."About WordNet." WordNet. Princeton University. 2010. http://wordnet.princeton.edu (accessed 30 July 2014).

Acknowledgements

Raine Koskimaa for all his care and patience; Giovanna Di Rosario for her encouragements; All of Jyväskylä university; Aksari Pashai and Aida Shamlou for their advices; The team member of the official website of Ahmad Shamlou; The Folklore group of Bamdad-e-Sari; Lucia Paprckova for bearing my sorrows; My mother and my father for staying strong in my exile.

Appendix I: Documents

(Digital Publishing contract)

قرارداد

مقدمه

این قرارداد منعقد می‌شود بین آقای عسکری پاشایی فرزند محمد باقر، متولد ۱۳۱۸، به شناسنامه‌ی ۶۵۳ صادره از ساری، به نشانی تهران، خیابان شهید کلاهدوز، خیابان شهید اصغر نعمتی، کوچه‌ی هشتم، فلکه‌ی مفتح، پلاک ۲۴، طبقه‌ی دوم، که در این قرارداد «وکیل» خوانده می‌شود، به وکالت رسمی، به شماره ۵۲۱۳۰۶، از دفترخانه‌ی اسناد رسمی ۳۰۸ حوزه‌ی تهران از سوی آقای احمد شاملو فرزند حیدر، دارنده‌ی شناسنامه‌ی شماره‌ی ۵۵۵۶۵ صادره از بخش ۲ تهران، ساکن شماره‌ی ۵۵۵ نسترن شرقی در شهرک دهکده در فردیس کرج، و با توجه به اختیارات مندرج در یادداشت مورخه‌ی ۵ / ۲/ ۷۶ آقای احمد شاملو، که از این پس در این قرارداد «مؤلف» نامیده می‌شود از یک طرف، و آقایان سید محسن عمادی فرزند سید حسین، به شماره‌ی شناسنامه‌ی ۵۸۲ صادره از ساری، متولد ۱۳۵۵، به نشانی ساری، خیابان امیر مازندرانی، خیابان فیضیه، خیابان علمیه، کوچه‌ی مقابل مغازه‌ی مهدوی، شماره‌ی ۸۵، و روزبه پورنادر فرزند محمد حسن، به شناسنامه‌ی شماره‌ی ۷۰۴، صادره از تهران، متولد ۱۳۵۸، به نشانی تهران، خیابان ستارخان، نرسیده به خیابان نصرت، خیابان گل‌بهی، شماره‌ی ۵۴، و حسین مسرّت مشهدی فرزند محمد جواد به شناسنامه‌ی شماره‌ی ۱۰۸۰ صادره از مشهد، متولد ۱۳۵۶، به نشانی مشهد، بلوار عبدالمطلب، کوچه‌ی ۱۲، شماره‌ی ۴۷، که از این پس «ناشران» نامیده می‌شوند.

موضوع قرارداد

واگذاری حق web publishing یا نشر اینترنتی تمام آثار آقای احمد شاملو است، اعم از شعر، مقاله، داستان، ترجمه‌ها، مجموعه‌های کتاب کوچه و فرهنگ کوچه، و هرگونه عکس، صدا، فیلم، و نیز هرگونه آثار و نشانه از ایشان ــ خلاصه تمام آثار «مؤلف» که تا تاریخ امضای این قرارداد چاپ شده یا پس از این چاپ خواهد شد و یا پس از این تولید یا ترجمه یا گردآوری خواهد شد، در سایت ثبت‌شده (registered) با نامِ دامنه‌ی (domain) احمد شاملو، یا شاملو یا بامداد یا کوچه، یا هر نام دیگری.

ماده ۱. مدت این قرارداد دائمی و انحصاری «ناشران» است.

ماده ۲. «مؤلف» برای پرداخت هزینه‌های ضروری نگهداری و هزینه‌های دیگر سایت، و هزینه‌های برنامه‌نویسی و تایپ آثار و درستی و امانت امر برای انتقال به شبکه، نظارت بر فرایند نشر و اجرای طرح، و نیز چاپ و نشر «آثار» به شکل CD در داخل کشور ایران و خارج

از ایران، آیدا شاملو (ریتا آتانث سرکیسیان) فرزند آشوت دارنده‌ی شناسنامه‌ی ۲۸۴ صادره از بخش یک کرمانشاه، ساکن شماره‌ی ۵۵۵ نسترن شرقی در شهرک دهکده در فردیس کرج، و آقای عسکری پاشایی فرزند محمد باقر دارنده‌ی شناسنامه‌ی ۶۵۳ صادره از ساری به نشانی تهران، خیابان شهید کلاهدوز (دولت سابق) خیابان شهید اصغر نعمتی، کوچه‌ی ۸، فلکه‌ی مفتح، شماره ۲۴، ط ۲، را از سوی خود «سرپرستان» مادام‌العمر (عمر سرپرستان) معرفی می‌نماید. کسب موافقت کتبی «سرپرستان» و ذکر نام آنان به‌عنوان سرپرست برای «ناشران» الزامی است.

تبصره: این «سرپرستان» یا وکلا می‌توانند اختیارات خود را به هر کس یا کسان یا هر نهادی که خود صلاح بدانند به طور موقت یا دائم، کتباً تفویض کنند، و نیز همین‌گونه است اختیارات «سرپرست» یا «سرپرستان» گروه دوم.

ماده ۳. این سایت کاملاً غیر انتفاعی است و استفاده از آن نیز برای همه (اعم از خواننده گان و پژوهش‌گران ایرانی و خارجی) رایگان خواهد بود.

ماده ۴. «مؤلف» هیچ‌گونه هزینه‌یی را برای برنامه‌نویسی و راه‌اندازی این سایت متقبل نخواهد شد.

ماده ۵. پرداخت هرگونه هزینه‌ی سایت از سوی دیگران، باید با نظارت «سرپرستان» صورت گیرد و طبعاً پردازنده‌گان هزینه‌ها در صورت انتفاعی شدن این سایت صاحب حق خواهند بود. در هر صورت هیچ ماده‌یی نباید در این قرارداد نافی یا ناقض ماده ۳ باشد.

ماده ۶. اگر به هر دلیلی نگهداری و بقای این سایت ناشران آن را به کسب درآمدی در این سایت وادارد (که این نباید نافی یا ناقض ماده ۳ باشد) «مؤلف» مخالفتی با آن ندارد، مشروط به آن‌که ضرورت آن برای «سرپرستان» مستدلاً قابل قبول باشد. موافقت کتبی ایشان ضروری است.

ماده ۷. این قرارداد قابل انتقال به غیر است با نظر و موافقت کتبی «سرپرستان».

ماده ۸. هرگاه «ناشران» بر اساس برنامه‌های تهیه شده برای شبکه، چاپ و نشر CD (دیسک لیزری) را برای تمام آثار یا بخش‌هایی از آن‌ها در داخل یا خارج کشور ضروری بدانند این امر بلامانع است مشروط به آن‌که حقوق مادی و معنوی «مؤلف» با توافق «سرپرستان» و «ناشران» از نظر نوع نشر و درصد آن رعایت شود. ضروری است که این توافق مکتوب باشد تا حقی از طرفین قرارداد ضایع نشود.

تبصره. ضروری است که در این مرحله از کار سرپرستان صاحب حقوق مادی باشند. طبعاً توافق آنان و «ناشران» ضروری است.

(این‌جانب عسکری پاشایی وکیل جناب شاملو از عنایت ایشان سپاسگزاری می‌کند و ذیل همین تبصره اعلام می‌کند که از هرگونه حقوق مادی موضوع این ماده برای همیشه چشم می‌پوشد.)

ماده ۹. نشر آثار گوناگون تولیدکننده‌گان گوناگون در این سایت درباره‌ی «مؤلف» منوط به اخذ اجازه‌ی کتبی «ناشران» از تولیدکننده‌گان و موافقت «سرپرستان» خواهد بود. هرگونه مسئولیتی که در این ماده پیش آید، به‌عهده‌ی ناشران خواهد بود مگر مواردی که اجازه‌ی «سرپرستان» مکتوب بوده باشد.

ماده ۱۰. «ناشران» مجاز به ایجاد بخشی در نقد آثار «مؤلف» نیز هستند. ضروری است که گروهی داوری در چند و چون ارائه‌ی نقدها را در شبکه بر عهده بگیرند. طبعاً «ناشران» نیز می‌توانند در صورت انتخاب اعضای این گروه هم باشند.

ماده ۱۱. در صورت بازماندن یا انصراف هر یک از «ناشران» از اجرای تعهدات و وظایف و اختیارات خود می‌تواند با توافق اعضای دیگر «ناشران» و نیز «سرپرستان» این تعهدات را به شخص یا اشخاص دیگر یا هر نهادی واگذار کند. ناگفته پیداست که تمام مسئولیت‌ها و حقوق مادی و معنوی «ناشر» مورد نظر به جانشین وی منتقل می‌شود.

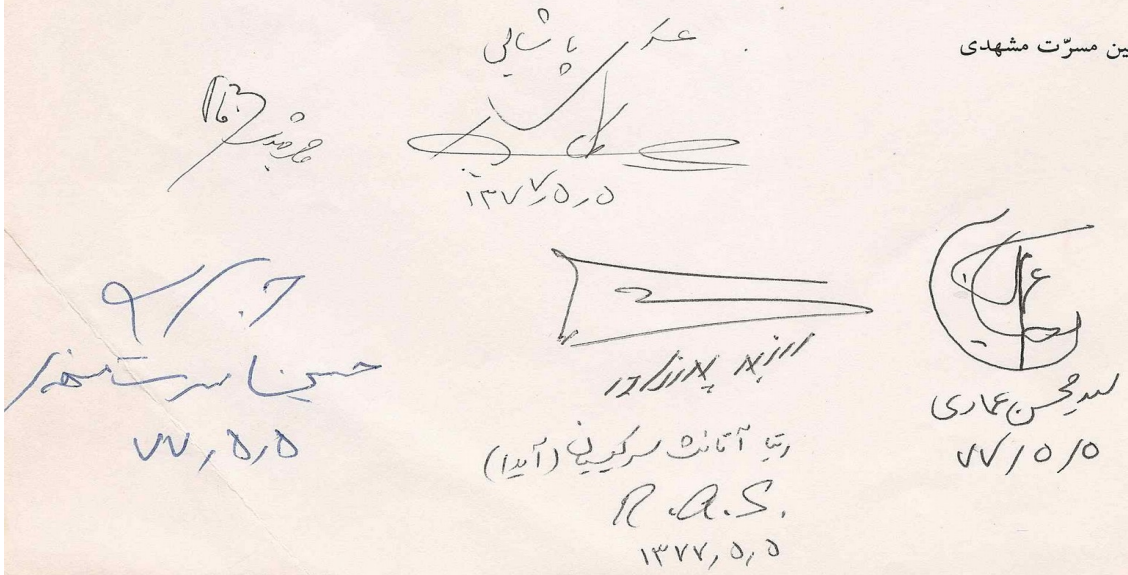
ماده ۱۲. نقض هر یک از مواد و تبصره‌های این قرارداد از سوی طرفین یا اختلاف در تعبیر و تفسیر مواد آن بنا بر ماده ۱۳ حل و فصل خواهد شد.

ماده ۱۳. هرگونه اختلاف در اجرا یا تغییر یا تفسیر مواد و تبصره‌های این قرارداد و یا تعیین خسارات موضوع ماده ۱۲ خواه میان «ناشران» و میان «ناشران» و «سرپرستان» از طریق حکم مرضی‌الطرفین حل و فصل خواهد شد.

ماده ۱۴. این قرارداد در تاریخ پنجم مرداد یکهزار و سیصد و هفتاد و هفت هجری شمسی (۱۳۷۷/۵/۵) در ۱۴ ماده و ۲ تبصره و در ۵ نسخه با اعتبار مساوی به امضای «وکیل» و «ناشران» رسیده است، و هر نسخه‌ی آن نزد یکی از امضاکنندگان آن خواهد ماند.

عسکری پاشایی　　سید محسن عمادی　　روزبه پورنادر

حسین مسرّت مشهدی



Draft translation of some articles

WHEREAS, Ahmad Shamlou ("Author"), an individual who, under Iranian law is identified as the son of Heidar and holder of ID card No. 55565 issued in Tehran, District 2, has authorized Attorney to act on his behalf by virtue of formal power of attorney No. 521306 vested in him and drawn in the Office of the Notary Public No. 308, as well as by an agreement entered into between Author and Attorney dated April 25, 1997;

WHEREAS, Publisher desires to acquire the sole and exclusive electronic publishing rights to the literary work (the "Work") of the Author; and

WHEREAS, Attorney desires to grant said rights to the Publisher

NOW, THEREFORE, the parties, in consideration of the mutual promises and agreements herein made and intending to be legally bound thereby, hereby agree as follows:

1. DEFINITIONS

The following words and phrases when used in this Agreement shall mean:

"Work" shall mean the entire body of Author's literary work including but not limited to all poetry, essays, translations by Author, translations of Author that are controlled by the Supervisory Committee/Guardians of Ahmad Shamlou's Work, journalistic writing, fiction, plays, screenplays, serializations, and all versions of the encyclopedia known as "Ketabe Kuche: The Book of Alleyways", as well as any photographs, audio recordings, videos, films and physical items of any kind related to the Author and/or his Work.

Work shall further be defined as any Work that has been published, produced, translated and/or collected in any medium as of the date of this Agreement, in addition to any Work that is published, produced, translated and/or collected in any medium after the date of this Agreement, as well as any Work that has been, or in the future will be, registered under the domain name of Ahmad Shamlou, Ahmad Shamlu, Ahmad Shamloo, A. Bamdad, Bamdad, Kuche, or any other name or variation thereof.

"Rights" shall be defined as any and all electronic publishing rights.

"Supervisory Committee/Guardians of Ahmad Shamlou's Work" (the "Guardians") shall mean the permanent committee formed by the Author, whose current members are Askari Pashai and Rita Atanes Sarkisian (formerly, Aida Shamlou) daughter of Ashoot, who have permanent supervisory authority over any licenses of the Author's Work.

4. WEBSITE

It was understood and agreed that any website and/or web pages maintained by Publisher which utilize the Work shall be established on a nonprofit basis. The Work on any such website and/or web pages shall be accessible to any person,

including Iranian and non-Iranian researchers, free of charge.

Publisher may retain moneys earned from any website and/or web pages to the extent that such funds are necessary to maintain and operate said website and/or web pages. Any surplus earnings related to the website and/or web pages, after all expenses and a reasonable working capital 'reserve' have been provided for, will be distributed among any advertisers and sponsors of the website and/or web pages.

The Guardians shall have the right to oversee all expenses and/or charges of third parties including sponsors and advertisers related to any website and/or web pages.

Publisher shall maintain accurate books, records and accounts of the operation of any website and/or web pages. Upon reasonable notice to Publisher, Guardians may inspect, audit and copy such books, records and accounts. Publisher shall fully cooperate with Guardians in connection with the same. Publisher shall furnish to Guardians quarterly reports showing the expenses and revenues of any website and/or web pages.

Appendix II : Extracts of two unpublished letters

(Letter I)

۳ اکتبر ۷۷

سیاوش نازنینم.

پس از ماه‌ها چشم انتظاری نامه‌ات رسید. از سلامت خودت و خانم و نورچشمان به راستی خوشحال شدیم. از نحوه نامه چنان برمی‌آید که از سفر بسیار قبراق و سرحال برگشته‌ای و به قول مؤمنان حسابی استخوان سبک کرده‌ای. گیرم نه از گناهان کبیره بلکه از خستگی‌های کثیره. البته کاری که از درگاه حضرت باب اعظم راهی فرموده بودی - و علی‌رغم تشبثات خانمانه‌ی پستخانه‌ی مبارکه‌ی ایتالیا بالاخره عزّ وصول بخشیده بود - به نحوی امکان استخوان سبک کردن آن جوری را هم به ذهن متبادر می‌کرد؛ وگرنه فی‌الواقع چرا نبایست فقط از مقر ذوالجلال آن حضرت کارت ارسالیه باشی، مگر جاتی دنیا قحط بود؟! به هرصورت امیدوارم سفر به ماکی برایت لذتبخش بوده باشد. چیزی در نامه‌ات در این باب ننوشته بودی. نمی‌دانم کجاها رفتی و کجاها نرفتی. سامان چون به لندن می‌آمد برای حضرتت پیغامی گذاشته بود و شماره‌ی تلفن لندنش را نوشته بود که اگر به آنجاها سری زدی خبرش کنی. بعد به من نامه‌ای نوشت که تلفن نکرده‌اند و ظاهراً اصلاً به جزیره‌ی انگلیس پیاده نشده‌اند و ادعای عنین می‌کرد.

باری به قول معروف اگر از احوالات ما خواسته باشید سلامتیم و به دعاگوئی مشغول.

از اتفاقات عجیبه برایت بنویسم: دکتر احسان یارشاطر که معروف حضورت هست. ایشان تابستان را به تهران تشریف فرما شدند و هفته‌ی پیش یا بیشتر بگویم دو هفته پیش برگشت به نیویورک. روز بعدش احمد کریمی که در حال حاضر به واشنگتن نقل مکان کرده با تلفن به من گفت که یارشاطر به او تلفن کرده و شماره‌ی مرا گرفته است که با من کار فوری دارد. همان شب دکتر یارشاطر به من تلفن کرد و تعارفات مفصل، و اظهار تمایل شدید به این که وصف کتاب کوچه را شنیده است و دلش می‌خواهد قسمت‌هایی آماده شده‌ی آن را ببیند و قرار گذاشت که روز دوشنبه (۲۵ سپتامبر، هفته‌ی قبل) نامه‌ای بایشان بنویسم. نامه را عذر خواستم و صبح ظهر دوشنبه را پیش او رفتم. رئیس دپارتمان "تحقیقات ایرانی" است و مرکز کارش در دانشگاه کلمبیا است واقع در مرکز نیویورک. اصرار کرده بود که مجلدات آماده‌ی کتاب را ببینند و من هم آنچه آماده بود بردم و دید. خود را سخت متعجب نشان داد که "واقعاً کار بزرگی است" و دوبار گفت "آن قدر از دیدن این مجلدات حیرت زده شده‌ام که مطلقاً یادم رفت چه چیزها در ذهنم داشتم که بگویم!" - بعد از سوابق کار پرسید و این که چه طور باید آن را پی‌گیری کرد و بقیه‌ی کار در چه مرحله‌ای است و غیره... جریانات تهران و بودجه‌ی مورد تقاضا که بعداً با آن نامه‌ی کذائی خانگی آش را خواندم، و این که چاپ کتاب ظاهراً در تهران مدتی است که شروع شده و

با فرو رفتن "تمدن بزرگ" در خاموشی برق، کار حروفچینی آن هم به توبهٔ اهمال افتاده و مشغول قانون کار انقلابی مملکت شد
همه را بر این گفتم. و صحبت به آنجا کشید که حجم زیاد کار و خستگی و بیماری من وضعی پیش آورده است که دیگر یک تنه نمی توانم این بار
را بکشم، و آن بر وجه هم برای این تقاضا شان بود که کم کم کار را به دست سازمانی متشکل سپرده شود و من به مسائل اصلی کار
بپردازم و امور وقت گیر ترس را که دیگران هم می توانند انجام بدهند به دیگران محول کنم. — به همین نکتهٔ قضیه چسبید و
بی مقدمه گفت "من همین جا ترتیب کار را در دانشگاه کلمبیا می دهم. همین امروز دستور می دهم اتاق و وسائل کار کافی
در اختیارتان بگذارند و مجانا" برای تکمیل نواقص همین دو حرف آ و الف دو نفر ایرانی را که در اینجا استاد یار هستند
و در کار تدوین دانشنامهٔ ایران و اسلام هم سابقه دارند و از سواد و ذوق کافی هم بهره مندند به اختیارتان می گذارم
تا بعد که این دو مجلد یا دو حرف کامل شد و خواستید به حرف بی بپردازید، اگر یک دیگری نیاز داشتید که
دست یارتان باشند مستقیما" از تهران استخدام کنیم". بعد هم، بدون اینکه منتظر جواب من بشود ادامه داد که "چون شما
هنوز برگ اقامت در آمریکا را نگرفته اید و طبق مقررات آمریکا نمی توانید شغل رسمی داشته باشید، من مجانا" از بودجهٔ مرکز
تحقیقات ایرانی ماهی هزار و دویست دلار به شما می پردازم که معطل نمانید تا وضع استخدامتان در دانشگاه کلمبیا روشن بشود!"
به او گفتم که یک ناشر آزاد در تهران مشغول چاپ کتاب است و من نمی توانم بدون مشورت با دوستی که در تهران
این مسائل را اداره می کند هیچ گونه تعهدی به شما یا دانشگاه کلمبیا بدهم. — در کمال حیرت بنده این جواب دادند که
"البته شما هیچ تعهدی ندارید. اگر کتاب چاپ شد که چه بهتر، در غیر آن صورت ما و دانشگاه کلمبیا برای چاپ آن هم به بهترین
صورت ممکن فکری خواهیم کرد!"

گفتم "آخر، شما یا دانشگاه کلمبیا که دفتر کار و ابزار کار [redacted] در اختیار من می گذارید و به من حقوق
می دهید و افرادی را استخدام می کنید که دستیار من باشند، در مقابل چی گیرتان می آید؟ — آنها نذر ندارند که در راه
رضای خدا چنین مخارجی را تحمل کنند" — گفت: "هیچ، ما فقط این کار را می کنیم برای اینکه این فعالیت به علل
مادی متوقف نشود". — شما از چهارشنبهٔ آینده (۳ اکتبر) بیائید دفتر کارتان را تحویل بگیرید، با دستیاران تان آشنا
بشوید و کارتان را شروع کنید!" و سوم اکتبر امروز بود. رفتم به نیویورک و به دانشگاه کلمبیا و دیدم که بله، اتاق
و میز و لوازم آماده است و دو نفر آن جا نشسته اند و انتظار مرا می کشند. تا ساعت ۳ بعد از ظهر ترتیب کار را دادیم
و من برگشتم به نیوجرسی. — در راه که می آمدم یادم افتاد که دو سه ماه قبل (در یکی از روزهایی که پارسا طرف به ایران رفته بود)
خبری درباره من در روزنامهٔ کلمبیان چاپ شده بود مبنی بر این که من در دانشگاه کلمبیا استخدام شده ام و کارهای مطالعه

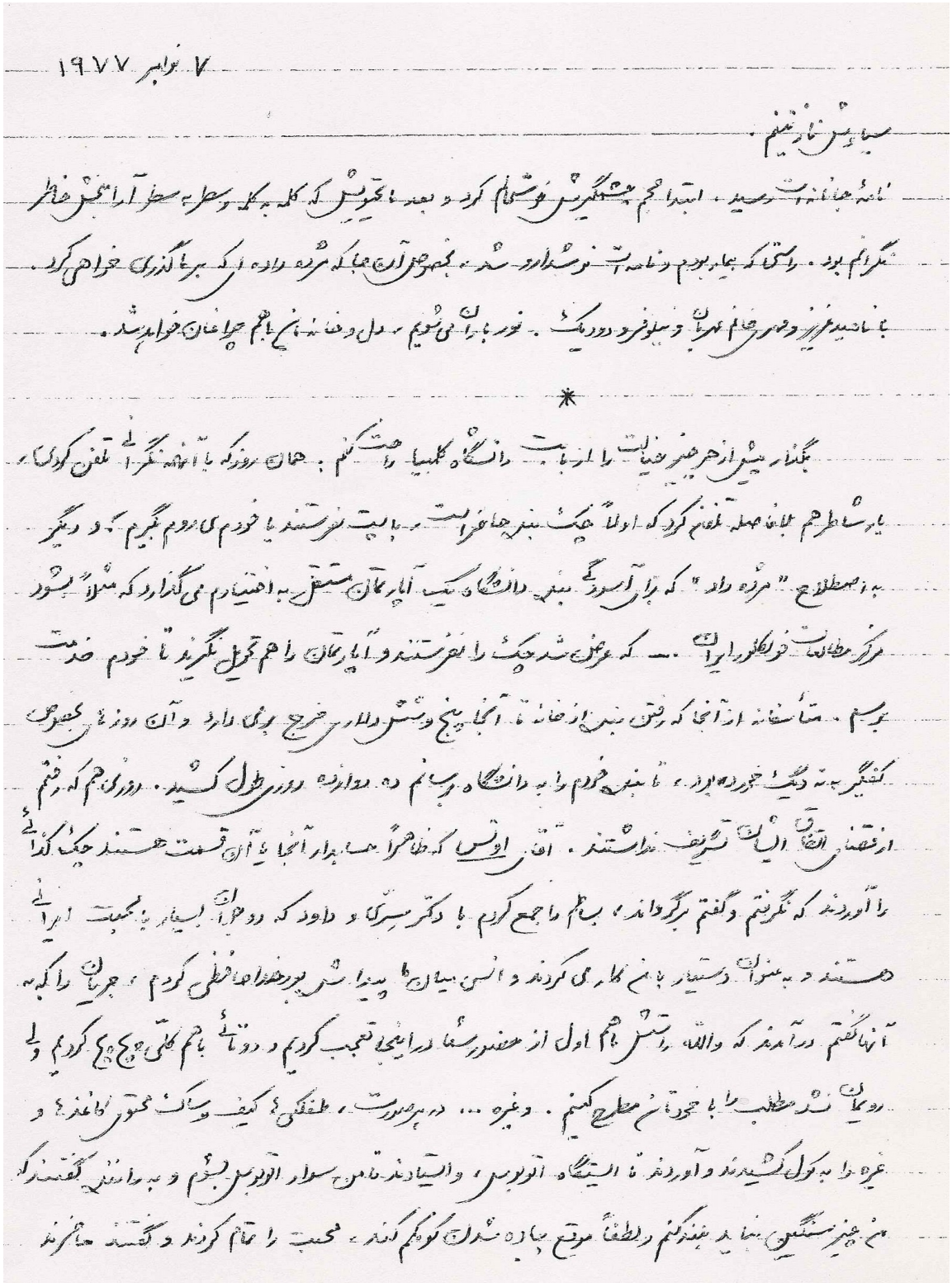
در فرهنگ عامیانهٔ ایران را در آنجا متمرکز کرده ام، و غیره و غیره ... البته ناگفته نگذارم که پادشاهِ در اینجا نمایندهٔ فرهنگ آن شخص مشخص است و بودجه‌ی کلانی در اختیار دارد که با آن دانشگاه‌ها، آمریکایی را می خرد و دامی دارد که به آن اشخاص معروف دکترای افتخاری بدهند و دانشجویان به اصطلاح "منحرف" ایرانی را زیر فشار بگذارند و الخ... البته افسوس نخواهد داشت که بتواند در من کارگر بیفتد. ما اگر با این صندوق سه شاهی، قرار بود از راه دریا برویم تا حالا با دریای سبزِ شرایط چشمگیرتری از راه دریا رفته بودیم. مع ذلک عجالتاً من این کار را به دو دلیل پذیرفته ام. نخست اینکه به هر شاهی پولی که به ام می دهند واقعاً احتیاج دارم. خدا نیل را لعنت کند. اگر لااقل یک جلد این کتاب را درآورده بودم می توانستم به پادشاه پیشنهاد کنم که اگر حیثیتِ دیگر برای این کار رفته است من حاضرم ماهی دو هزار دلار به او بدهم که بیاید و در این کار با من کمک کند! ـ و حالا ناگزیرم با شرائطی که اصلاً برایم روشن نیست پیشنهاد او را بپذیرم، فقط به خاطر اینکه با طناب پوسیدهٔ پول به چاه رفته ام، و فقط به خاطر اینکه وجداناً نخواسته ام با سپردنِ این کار به تراست امیرکبیر که دستگاه سانسور در برابر ناشران کوچکتر از شر حمایت می کند آب به آسیاب دشمن ریخته باشم ... دلیل دوم هم این است که استخدام دو نفر با سواد که سوابقی در امر تألیف و انشای نامه‌ی ایران و اسلام دارند به عنوان دستیار، امتیازی است که من شخصاً با دست و جیب خالی نمی توانسته ام به دست بیاورم و با قبول این امر یک چنین دستیارانی را به دست آورده ام و این می تواند به مقدار زیادی در پیشرفت کاری که عمر و سلامتم را رویش گذاشته ام بی نهایت مؤثر باشد.

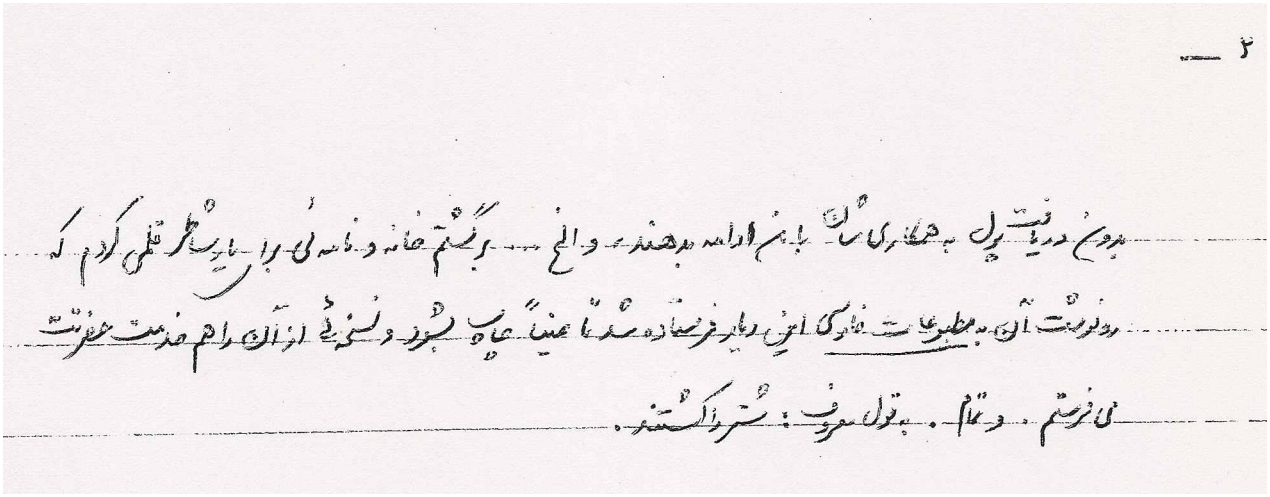
اینها را نوشتم برای آنکه بدانی و برایم بنویسی که چه فکر می کنی. در حال حاضر کار را شروع کرده ام بدون اینکه هیچ گونه تعهدی سپرده باشم. البته اگر دیدم کلک در کار است گرو پدر پول شان و امکانات شان. فکر نمی کنی در تهران به او در این مورد توصیه هائی کرده دستور العمل هائی داده باشند؟

Description

The letter is written in 3 October 1977 by Ahmad Shamlou to A.Pashai, a member of the Supervisory Committee of the electronic publishing contract. Shamlou refers to the invitation of Ehsan Yarshater, the director of Iranica, to work in the context of Iranian Studies department of Columbia University, on Ketab Kuche. In this letter he accepts the offer, but he mentions that he is aware about the relation between Yarshater and the dictatorship of the Shah in Iran. He is confused if Yarshater's invitation was initiated by the Iranian government or not.

Letter II

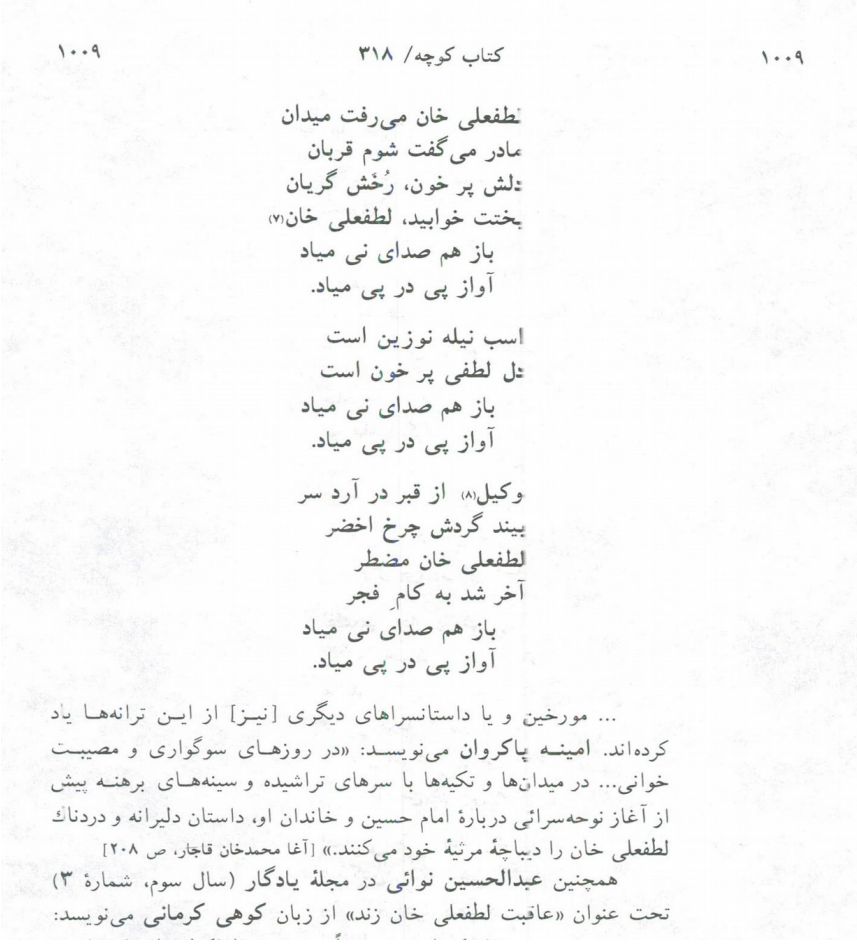




Description

This letter was written in 7 November 1977, almost one month after the pervious one. Here he describes that Yarshater calls him for his salary from the university and he says he will come in person and there is no need to send it by check. By the way, he can not go, because he even does not have five dollars to take the bus. 12 days after the call, he goes there, but not for receiving his salary. He goes to resign. Here he mentions to a political essay he wrote against Yarshater and published in newspapers in meanwhile.

Appendix III: Some variations of the manuscript



۱۰۰۹ کتاب کوچه/ ۳۱۸ ۱۰۰۹

لطفعلی خان می‌رفت میدان
مادر می‌گفت شوم قربان
دلش پر خون، رُخَش گریان
بختت خوابید، لطفعلی خان[۷]
باز هم صدای نی میاد
آواز پی در پی میاد.

اسب نیله نوزین است
دل لطفی پر خون است
باز هم صدای نی میاد
آواز پی در پی میاد.

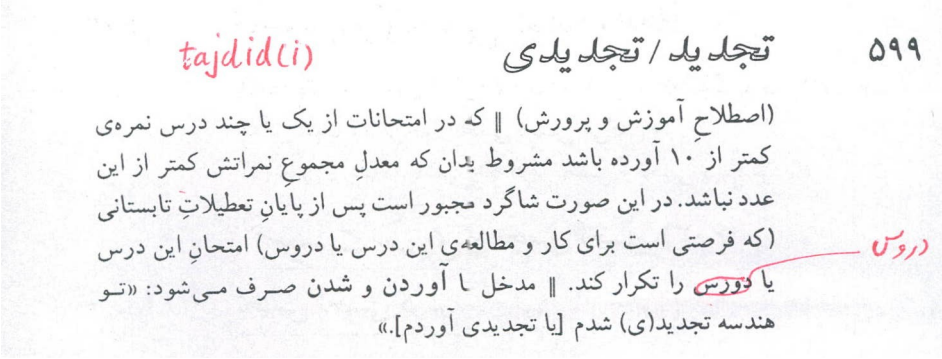
وکیل[۸] از قبر در آرد سر
بیند گردش چرخ اخضر
لطفعلی خان مضطر
آخر شد به کامِ فجر
باز هم صدای نی میاد
آواز پی در پی میاد.

... مورخین و یا داستانسراهای دیگری [نیز] از این ترانه‌ها یاد کرده‌اند. **امینه پاکروان** می‌نویسد: «در روزهای سوگواری و مصیبت خوانی... در میدان‌ها و تکیه‌ها با سرهای تراشیده و سینه‌های برهنه پیش از آغاز نوحه‌سرائی دربارهٔ امام حسین و خاندان او، داستان دلیرانه و دردناک لطفعلی خان را دیباچهٔ مرثیهٔ خود می‌کنند.» [آغا محمدخان قاجار، ص ۲۰۸]

همچنین **عبدالحسین نوائی** در **مجلهٔ یادگار** (سال سوم، شمارهٔ ۳) تحت عنوان «عاقبت لطفعلی خان زند» از زبان **کوهی کرمانی** می‌نویسد:

Description:

An old popular lyrics. The entry describes the political context of the lyrics and refers to some references of it.



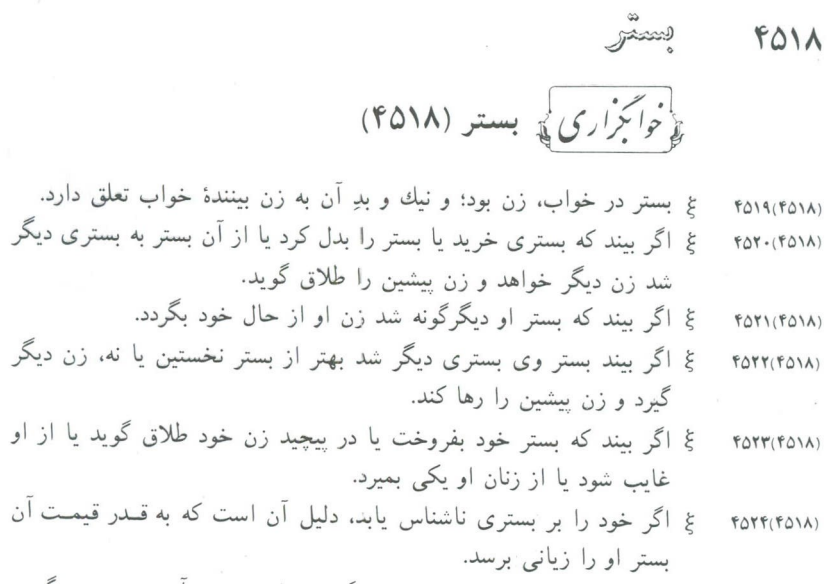
۵۹۹ تجدید / تجدیدی tajdid(i)

(اصطلاحِ آموزش و پرورش) ‖ که در امتحانات از یک یا چند درس نمره‌ی کمتر از ۱۰ آورده باشد مشروط بدان که معدلِ مجموعِ نمراتش کمتر از این عدد نباشد. در این صورت شاگرد مجبور است پس از پایانِ تعطیلاتِ تابستانی (که فرصتی است برای کار و مطالعه‌ی این درس یا دروس) امتحانِ این درس یا دورس را تکرار کند. ‖ مدخل ـا آوردن و شدن صرف می‌شود: «تو هندسه تجدید(ی) شدم [یا تجدیدی آوردم].»

دروس

Description:

Description of a term in modern educational system.



۴۵۱۸ بستر

خوابگزاری بستر (۴۵۱۸)

۴۵۱۹(۴۵۱۸) ع بستر در خواب، زن بود؛ و نیك و بدِ آن به زن بینندهٔ خواب تعلق دارد.

۴۵۲۰(۴۵۱۸) ع اگر بیند که بستری خرید یا بستر را بدل کرد یا از آن بستر به بستری دیگر شد زن دیگر خواهد و زن پیشین را طلاق گوید.

۴۵۲۱(۴۵۱۸) ع اگر بیند که بستر او دیگرگونه شد زن او از حال خود بگردد.

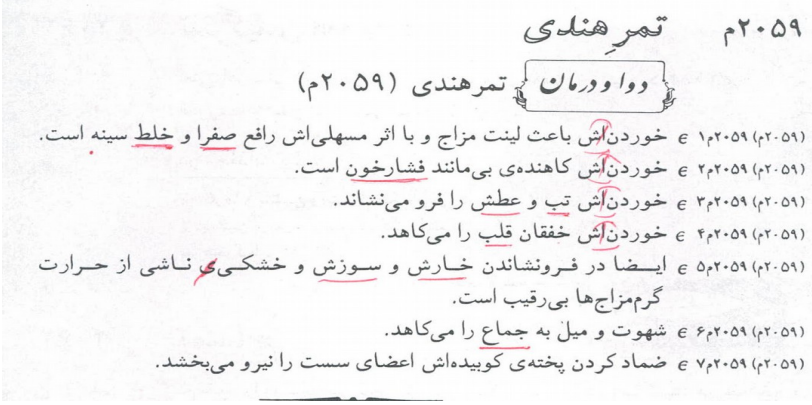
۴۵۲۲(۴۵۱۸) ع اگر بیند بستر وی بستری دیگر شد بهتر از بستر نخستین یا نه، زن دیگر گیرد و زن پیشین را رها کند.

۴۵۲۳(۴۵۱۸) ع اگر بیند که بستر خود بفروخت یا در پیچید زن خود طلاق گوید یا از او غایب شود یا از زنان او یکی بمیرد.

۴۵۲۴(۴۵۱۸) ع اگر خود را بر بستری ناشناس یابد، دلیل آن است که به قدر قیمت آن بستر او را زیانی برسد.

Description:

Several interpretations of the dreams that involve “bed”.



۲۰۵۹م تمرهندی

دوا و درمان تمرهندی (۲۰۵۹م)

(۲۰۵۹م) ۱م۲۰۵۹ ∈ خوردن‌اش باعث لینت مزاج و با اثر مسهلی‌اش رافع صفرا و خلط سینه است.
(۲۰۵۹م) ۲م۲۰۵۹ ∈ خوردن‌اش کاهنده‌ی بی‌مانند فشارخون است.
(۲۰۵۹م) ۳م۲۰۵۹ ∈ خوردن‌اش تب و عطش را فرو می‌نشاند.
(۲۰۵۹م) ۴م۲۰۵۹ ∈ خوردن‌اش خفقان قلب را می‌کاهد.
(۲۰۵۹م) ۵م۲۰۵۹ ∈ ایضا در فرونشاندن خارش و سوزش و خشکی‌ی ناشی از حرارت گرم‌مزاج‌ها بی‌رقیب است.
(۲۰۵۹م) ۶م۲۰۵۹ ∈ شهوت و میل به جماع را می‌کاهد.
(۲۰۵۹م) ۷م۲۰۵۹ ∈ ضماد کردن پخته‌ی کوبیده‌اش اعضای سست را نیرو می‌بخشد.

Description:

Popular medicine. Different ideas existing about eating Tamarind.

۴۵۶۳ بستن

در تداول بَسّن bassan نیز می‌آید. ‖ بند کردن. مقید کردن. ‖ مقابل باز کردن * وا کردن * گشودن.

***

در تداول به اشکال زیر صرف می‌شود:

مضارع:

اخباری

می‌بنده (تش) = می‌بندد (او) — mibande (teš)
می‌بندن (اشون) (دشون) = می‌بندند (آنها) — ——— an
می‌بندین = می‌بندید (شما) — ——— in

التزامی

بنده(تش) = بندد(او) — bebande (teš)
بندن (اشون) (دشون) = بندند (آنها) — ——— an
بندین = بندید (شما) — ——— in

ملموس

داره می‌بنده(تش) = دارد می‌بندد (او)
(با افعال دارم، داری،... بر سر مضارع اخباری)

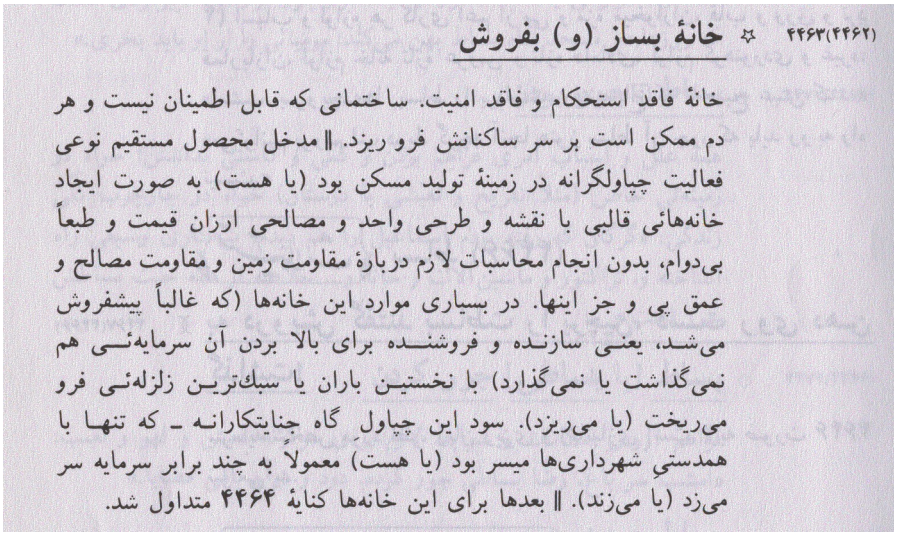
ماضی:

مطلق

بستم و بسّم = بستم (من) — bastam / bassam
بستی و بسّی = بستی (تو) — ——i / ——i
بست و بستش و بسّش = بست (او) — —— (eš) /—(eš)

Description:

Grammar. On conjugating a verb in colloquial language of Tehran.



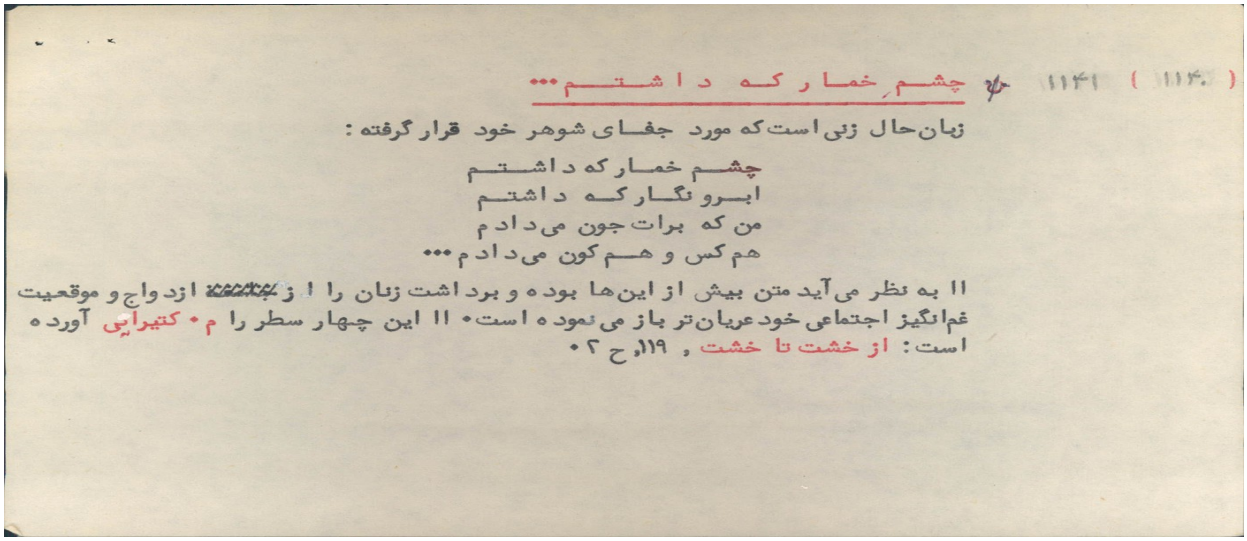
۴۴۶۳(۴۴۶۲) ✡ **خانهٔ بساز (و) بفروش**

خانهٔ فاقد استحکام و فاقد امنیت. ساختمانی که قابل اطمینان نیست و هر دم ممکن است بر سر ساکنانش فرو ریزد. ‖ مدخل محصول مستقیم نوعی فعالیت چپاولگرانه در زمینهٔ تولید مسکن بود (یا هست) به صورت ایجاد خانه‌هائی قالبی با نقشه و طرحی واحد و مصالحی ارزان قیمت و طبعاً بی‌دوام، بدون انجام محاسبات لازم دربارهٔ مقاومت زمین و مقاومت مصالح و عمق پی و جز اینها. در بسیاری موارد این خانه‌ها (که غالباً پیشفروش می‌شد، یعنی سازنده و فروشنده برای بالا بردن آن سرمایه‌ئی هم نمی‌گذاشت یا نمی‌گذارد) با نخستین باران یا سبک‌ترین زلزله‌ئی فرو می‌ریخت (یا می‌ریزد). سود این چپاولِ گاه جنایتکارانه ـ که تنها با همدستیِ شهرداری‌ها میسر بود (یا هست) معمولاً به چند برابر سرمایه سر می‌زد (یا می‌زند). ‖ بعدها برای این خانه‌ها کنایهٔ ۴۴۶۴ متداول شد.

Description:

Description of a slang referring to contemporary rapid approaches in making houses and dwellings.

(Some unpublished entries)



چشم خمار که داشتم ... ۱۱۴۱ (۱۱۴۰)

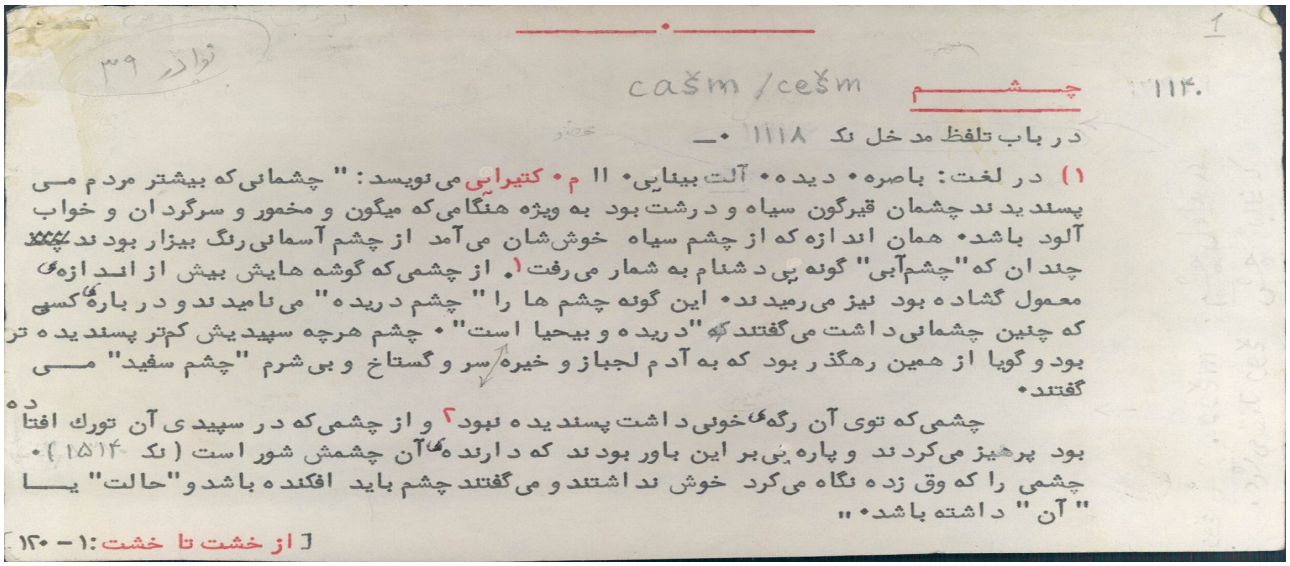
زبان‌حال زنی است که مورد جفای شوهر خود قرار گرفته :

چشم خمار که داشتم
ابرو نگار که داشتم
من که برات جون می‌دادم
هم کس و هم کون می‌دادم ...

!! به نظر می‌آید متن بیش از این‌ها بوده و برداشت زنان را از ازدواج و موقعیت غم‌انگیز اجتماعی خود عریان‌تر باز می‌نموده است. !! این چهار سطر را م. کتیرایی آورده است : از خشت تا خشت , ۱۱۹, ح ۲.

Description:

A slang and some porn verses describing the perspective of women about marriage-life.



۱۱۴۰ چشم cašm /cešm

در باب تلفظ مدخل نک ۱۱۱۸.

۱) در لغت : باصره. دیده. آلت بینایی. !! م. کتیرایی می‌نویسد : " چشمانی که بیشتر مردم می‌پسندیدند چشمان قیرگون سیاه و درشت بود به ویژه هنگامی که میگون و مخمور و سرگردان و خواب آلود باشد. همان اندازه که از چشم سیاه خوش‌شان می‌آمد از چشم آسمانی رنگ بیزار بودند چندان که "چشم‌آبی" گونه‌یی دشنام به شمار می‌رفت. از چشمی که گوشه هایش بیش از اندازه معمول گشاده بود نیز می‌رمیدند. این گونه چشم ها را " چشم دریده " می‌نامیدند و درباره کسی که چنین چشمانی داشت می‌گفتند که "دریده و بیحیا است". چشم هرچه سپیدیش کم‌تر پسندیده‌تر بود و گویا از همین رهگذر بود که به آدم لجباز و خیره‌سر و گستاخ و بی‌شرم "چشم سفید" می‌گفتند.

چشمی که توی آن رگ خونی داشت پسندیده نبود و از چشمی که در سپیدی آن تورك افتاده بود پرهیز می‌کردند و پاره‌یی بر این باور بودند که دارنده آن چشمش شور است (نک ۱۵۱۴).

چشمی را که وق زده نگاه می‌کرد خوش نداشتند و می‌گفتند چشم باید افکنده باشد و "حالت" یا " آن " داشته باشد. "

از خشت تا خشت : ۱ – ۱۲۰

Description:

Beginning of an entry about “eye” and some common stereotypes and beliefs about the forms and colors of the eyes.