

# Development of thesauri in Iran

Masoumeh Bagheri

---

*The need for Persian thesauri became apparent during the late 1960s with the advent of documentation centres in Iran. The first Persian controlled vocabulary was published by IRANDOC in 1977. Other centres worked on translations of existing thesauri, but it was soon realised that these efforts did not meet the needs of the centres. After the Islamic revolution in 1979, the foundation of new centres intensified the need for Persian thesauri, especially in the fields of history and government documents. Also, during the Iran–Iraq war, Iranian research centres produced reports in scientific and technical fields, both to support military requirements and to meet society's needs. In order to provide a comprehensive thesaurus, the Council of Scientific Research of Iran approved a project for the compilation of such a work. Nowadays, 12 Persian thesauri are available and others are being prepared, based on the literary corpus and conformity with characteristics of Iranian culture.*

---

The development of documentation centres in Iran in the late 1960s made it necessary to create thesauri of terms in Persian (the national and official language of Iran). The foundation in 1968 of the Iranian Documentation Centre (IRANDOC), which in turn provided the first indexing service in Iran, created the need for a controlled vocabulary of Persian literature. To meet this need, IRANDOC published a number of articles in the *IRANDOC Technical Journal*, as well as guidelines and a manual for the construction of a thesaurus. In 1977, it published a select list of descriptors used by indexers working in IRANDOC. Thus, in the absence of a regular thesaurus in Persian, the centre published the first Persian controlled vocabulary.

Gradually, more documentation centres were established, and started providing controlled vocabularies in Persian by translating existing English and French thesauri. Iranian documentation centres used these for some time, but it was soon realized that these did not satisfy the needs of the centres. In fact, these thesauri failed to provide a perfect controlled vocabulary in Persian because they were simply translations of the original sources and did not include proper descriptors for expressing the particular concepts of Iranian culture. Clearly, it was still necessary to compile a suitable thesaurus.

After the victory of the Islamic Revolution in 1979, the existing information centres had to improve their services to meet the increasing information needs of users. Meanwhile, the foundation of new centres to provide for the storage and retrieval of huge amounts of documents inherited from the collapsed regime intensified the need for a Persian thesaurus, especially in the fields of history and government documents. Also, during the eight-year war between Iran and Iraq (1980–88), Iranian research centres produced many reports in scientific and technical fields both to support military requirements and to meet national needs. These resulted in the establishment of new information centres and services, and the development of existing ones.

Meanwhile, various research centres in Islamic studies exploited computerized information storage and retrieval systems. This necessitated post-coordinate indexing and a device to make this possible, namely a 'thesaurus'. To meet this requirement, these centres had to compile Persian thesauri in related fields.

## The advent of controlled vocabularies in Persian

The first Persian controlled vocabulary was an outcome of the indexing process carried out at IRANDOC, who also published it. Preferred terms to be used in indexing were selected through a discussion among indexers, information workers, subject specialists and linguists. During the process of indexing, this vocabulary underwent some changes, and descriptors were revised from time to time. After several years of experience in applying the vocabulary, IRANDOC published *An alphabetical list of descriptors accepted by the Iranian Documentation Centre* (IRANDOC, 1977). The second and third editions of this were published in 1978 and 1988. It was not in fact a standard thesaurus with a systematic organization of hierarchical, and associative relations. Rather, it was an effort to provide a controlled vocabulary in Persian that could be used by Iranian documentation centres as an indexing language.

## Translation of existing thesauri into Persian

Although IRANDOC was the first documentation centre in Iran to introduce indexing services, the first Persian thesaurus was brought out in 1977 by the Plan and Budget Organization. It was a Persian version of *Macrothesaurus: a basic list of economic and social development terms*, published in 1972 in English by the Organization for Economic Cooperation and Development (OECD). Only a limited number

of copies of the document were published in Persian. It consisted of two parts: alphabetical and classified. In the alphabetical part each descriptor was accompanied by related terms, as is usual in most thesauri (Viet, 1972).

The next step was taken by the then Asian Cultural Documentation Center for UNESCO. In order to provide a controlled vocabulary to be used in the indexing process, the centre had to ascertain the appropriate descriptors in the fields of culture, arts and other related subjects from social sciences and the humanities. For this purpose, the centre translated three monographs into Persian:

- *Guidelines for indexing cultural documents and developing a thesaurus on cultural development in Asian countries*, by D. Saintville. The translated version was published in 1977 (Saintville, 1977).
- *Subjects descriptors to be used in ISORID*, by the International Federation for Documentation (FID). This was originally published in English and French. A Persian translation was published in 1978 (FID, 1978).
- *Cultural development thesaurus*, prepared by Jean Viet for the Council for Cultural Cooperation of the Council of Europe, originally published in English and French languages. A Persian version was published in 1979 (Viet, 1979).

In 1996, IRANDOC published the first edition of the *NAMA thesaurus: science and technology policy information exchange system*, based on *SPINES thesaurus of UNESCO*. (NAMA was a mere translation of SPINES into Persian.) The second edition of NAMA was published in 1998, in two volumes. It has an alphabetical part, permuted index, and Persian–English, and English–Persian glossaries. The equivalent of every descriptor has been given in English. NAMA contains about 11,000 preferred and non-preferred terms drawn from Persian texts, several glossaries and subject dictionaries.

Similarly the Cultural Research Bureau decided to translate into Persian two UNESCO publications: the *UNESCO Thesaurus* (Aitcheson, 1996), and the *International thesaurus of cultural development: sub-Saharan Africa* (UNESCO, 1998). The first of these was published in Persian in 1996 and contains only three parts out of the 24 in the original work. It meets the need for a controlled vocabulary in subjects like philosophy, religion, culture, arts, communication policy, communication technology, library, archives, documentation and information science. The Persian version has three sections: classified, alphabetical by Persian alphabet, and alphabetical by English alphabet. In the classified section, descriptors are accompanied by ‘Used for’ references that refer to non-preferred synonyms, ‘Scope notes’, and qualifiers wherever explanation has been considered necessary. In the alphabetical section, each descriptor carries a notation which indicates the location of the descriptor in the classified section.

The second work referred to above is not a mere translation of the original thesaurus. Wherever necessary, terms referring to African culture in the original source have been adapted to the Iranian culture. The work has two sections: classified and alphabetical. The Persian version has broadly followed the sequence of the classified order of the original thesaurus. Under each main subject heading, related descriptors have been given in an alphabetical order in English. In this part, relations between synonyms have been indicated by means of ‘Use’ and ‘Used for’ references. Hierarchical and associative relations have been shown through standard symbols (BT/NT and RT/RT). It also contains ‘Scope notes’ under vague descriptions. This assists the user to apply the thesaurus properly. However, the alphabetical section contains Persian–English and English–Persian glossaries only. In this part, each descriptor refers to its location in the classified section.

## Provision of thesauri in Persian

In order to encourage the compilation of thesauri in Persian, IRANDOC published a number of monographs in the 1970s. The first was published under the title *Rules and principles of thesaurus construction* (Hendesi Afshar, 1971). The second was a translation of the *Guidelines for the establishment and development of monolingual thesauri* by Austin, and Dale (1976). In spite of these efforts, the first original Persian thesaurus did not come into existence until several years later, and was produced by the Islamic Revolution’s Cultural Documentation Centre (IRcDO).

In 1982, along with a few other centres, the Asian Cultural Documentation Center for UNESCO was merged with the IRcDO, which continued the activities of the erstwhile centres that had been merged with it. Within a decade, the centre had indexed about 170,000 articles and collected more than 8,000 descriptors. In 1991, IRcDO established a committee for the compilation of a Persian thesaurus in the humanities. While compiling the thesaurus, the committee considered the existing descriptors used at the centre, pertinent thesauri and glossaries available in the field, and the opinions of various Iranian subject specialists. Eventually, the first edition of the *Persian cultural thesaurus*, popularly



Masoumeh Bagheri (left) at the ANZSI conference 2005  
Reproduced by permission of Jane Purton.

known as ASFA, was published in 1996 (IReCDO, 1996). This edition covers fields such as education, communications, economics, sociology, geography, psychology, and library and information science. ASFA contains about 3,400 terms (descriptors and non-descriptors). Where a descriptor pertains to several fields of knowledge, a scope note is added. Also if a descriptor admits of more than one meaning, a qualifier is given in a bracket to explain the intended meaning. ASFA consists of four parts:

1. A classified section, containing descriptors. Under each descriptor, non-preferred synonyms are shown by means of 'Used for' references.
2. An alphabetical section. This contains both descriptors and non-descriptors. Descriptors are given in bold print, while non-descriptors appear in normal print. Non-descriptor terms refer to descriptors through 'Use' references. The reciprocal reference ('Used for') is also provided under the appropriate descriptor. Under each descriptor, all usual relationships with other descriptors are shown by standard symbols: 'BT' and 'NT' for hierarchical relations, and 'RT' for associative relations.
3. A hierarchical section. In this section each group of descriptors is arranged under a top term. It is similar to the classified section, but has no 'Used for' references.
4. A permuted (i.e. rotating) index, listing compound descriptors alphabetically.

All descriptors carry a notation to determine the location of each term in the classified part of the thesaurus.

IReCDO intends to publish the next edition of ASFA in the near future. In addition to the subjects already covered, the next edition will include literature, language, linguistics, religion, politics, culture and arts.

Rahadust similarly compiled the first Persian controlled vocabulary in medicine, by adopting *Medical Subject Headings (MeSH)* as a pattern for the structure of her work. To meet the needs of specialized libraries in the field of medical sciences, the Iran University of Medical Sciences published *Persian Medical Subject Headings* in 1993. It contains about 1,500 subject headings arranged in alphabetical and classified orders. Several common subheadings based on form, language, geography, etc. make them more specific. Each subject heading is accompanied by its equivalent in English. Subsequently, the compiler developed the *Persian Medical Thesaurus* on the basis of her earlier work, and the National Library of Islamic Republic of Iran published it in 1998 (Rahadust, 1998). It has more than 7,500 preferred and non-preferred terms. This thesaurus consists of two parts: an alphabetical part and a permuted index. An English-Persian glossary is also provided. Preferred terms (descriptors) are referred from non-preferred synonymous by means of 'Use' references, wherever necessary. Hierarchical and associative relations have been provided to help the reader make better use of the thesaurus. The contents of this controlled vocabulary have been drawn from the literary corpus on the terms actually available in the collection of several Iranian medical libraries, thus enhancing its practical utility and realism (Khorramshahi, 1994).

Following these developments, several research centres started taking an interest in Islamic studies, which in turn created a need for a special thesaurus in Islamic subjects. The Research Centre for Islamic Studies, founded in 1989 in Qom, began to index significant texts of Islamic literature. Since no suitable thesaurus in the subject was available, however, the centre had to stop the indexing process after only a short time. To meet its requirements, the centre recognized the need for a comprehensive thesaurus of Islamic subjects. The first effort of the centre was a general thesaurus, with an emphasis on Islamic sciences. It published only a limited number of copies of this work in 1994. It contained about 4,000 descriptors and 2,500 non-descriptors, arranged alphabetically. Synonyms were related by means of 'Use', and the reciprocal 'Used for' references. 'Scope notes' provided definition wherever needed.

Since it was not possible to compile a comprehensive specialized thesaurus immediately, the centre decided to prepare and publish the thesaurus in parts, and has since published five parts of the specialized thesaurus in Islamic subjects. The first was the *Qoranic science thesaurus*. This contained about 1,300 descriptors, arranged in alphabetical and systematic orders. In order to enhance the utility and the efficiency of the thesaurus, about 2,800 synonyms referring to these descriptors through 'Use' references were added to the work. The second work, the *Glossary of Islamic philosophy*, was published in 1997. It contained about 23,000 terms arranged alphabetically, although synonyms as well as hierarchical and associative relationships are shown in the alphabetical section of the thesaurus. The centre intends to publish the systematic and hierarchical sections of the work in due course. The third part, the *Glossary of Islamic legal principles*, containing about 5,000 terms, was published in 1999. In addition to these, the centre has compiled two more parts of the comprehensive thesaurus in Islamic sciences. These are the *Glossary of Islamic theology* and the *Glossary of logic*, both published in 2004. It is expected that the centre will publish the other parts in the near future (Centre of Research in Islamic Studies 1997; 1999; 2004a and b).

In order to provide a comprehensive thesaurus of Persian literature, the Council of Scientific Research of Iran has approved a project for the compilation of a Persian comprehensive thesaurus covering: the arts, humanities, medical sciences, and science and technology (Beheshti, 1998). In this project IRANDOC is responsible for compiling the section on science and technology through a thorough revision of NAMA.

Recently IRANDOC has published the *Engineering thesaurus*, the *Thesaurus of chemistry* and the *Thesaurus of sociology* in electronic and print forms. These thesauri consist of alphabetical and classified parts, permuted indexes, and English-Persian glossaries (IRANDOC 2003; 2004a and b). It is expected that thesauri in physics, geology, zoology and other fields will be published in the near future.

## Conclusion

A glance at the growth and development of vocabulary control in Persian will make it clear that the process can be divided into three distinct stages:

- preparation of lists of Persian terms selected by documentation centres while indexing
- translation into Persian of thesauri available in English and French
- compilation of Persian thesauri using corpus-based terms derived from the terminology found in Persian documents, indexed by various centres in Iran.

Nowadays, at least 12 Persian thesauri are available within Iranian documentation centres and others are being prepared. In addition, some information centres use their own thesauri or lists of descriptors, which have not yet been published. For example, the Atomic Energy Organization, and the National Iranian Oil Company follow a particular list of accepted descriptors. The strong point of these thesauri is their base in the literary corpus and conformity with the characteristics of Persian language and Iranian culture, which were not found in the translated versions of thesauri from foreign languages. However, the need for further revision and development of these thesauri is becoming increasingly evident.

## References

- Aitchison, J. (1996) *Culture, communication and information thesaurus extracted from the UNESCO thesaurus*. Translated by Abbas Horri and Parvane Gudarzi as *Estelahname-ye farhang, ertejabat, ettela'at as estelahname-ye UNESCO*. Tehran: Cultural Research Bureau.
- Austin, D. and Dale, P. (1976) *Guidelines for the establishment and development of monolingual thesauri*. Translated by A. Horri as *Rahnamaye tahiye va gostareh-e estelahname-ye yek zabane*. Tehran: IRANDOC.
- Beheshti, M. (1998) *Gozaresh-e tarh-e pazhuheshi-ye estelahname-ye jam-e olum-e paye* (Report on the development of the comprehensive thesaurus in science). Tehran: IRANDOC.
- Centre of Research in Islamic Studies (1997a) *Estelahname-ye falsafe-ye Eslami* (Glossary of Islamic philosophy). Qom: Centre of Research in Islamic Studies.
- Centre of Research in Islamic Studies (1997b) *Estelahname-ye 'olum-e Qorani* (Quranic science thesaurus). Qom: Centre of Research in Islamic Studies.
- Centre of Research in Islamic Studies (1999) *Estelahname-ye osul-e feqh* (Glossary of Islamic legal principles). Qom: Centre of Research in Islamic Studies.
- Centre of Research in Islamic Studies (2004) *Estelahname-ye manteq* (Glossary of Islamic logic). Qom: Centre of Research in Islamic Studies.
- Centre of Research in Islamic Studies (2004) *Estelahname-ye kalam* (Glossary of Islamic theology). Qom: Centre of Research in Islamic Studies.
- FID (1978) *Subject descriptors to be used in ISORID*. Translated by Mohammad Hosein Daneshi and Abd ol-Hosein Azarang as *Estelahname-ye 'elm-e nettle'rasani va documantasion*. Tehran: Asian Cultural Documentation Center for UNESCO.
- Hendesi Afshar, G. (1971) *Qava'ed va qarardadha-ye sakhtan-e estelahname* (Rules and principles of thesaurus construction). Tehran: IRANDOC.
- IRANDOC (1977) *An alphabetical list of descriptors accepted by the Iranian Documentation Centre*. Tehran: IRANDOC.
- IRANDOC (1996) *Estelahname-ye nama: nezam-e mobadele-ye ette-la'at-e 'elmi fanni* (NAMA thesaurus: science and technology policy information exchange system). Tehran: IRANDOC.
- IRANDOC (1997) *An alphabetical list of descriptors accepted by the Iranian Documentation Centre*. Tehran: IRANDOC.
- IRANDOC (2003) *Estelahname-ye shimi* (Thesaurus of chemistry). Tehran: IRANDOC.
- IRANDOC (2004a) *Estelahname-ye fanni mohandesi* (Thesaurus of engineering). Tehran: IRANDOC.
- IRANDOC (2004b) *Estelahname-ye jameashenasi* (Thesaurus of sociology). Tehran: IRANDOC.
- IRECDO (1996) *Estelahname-ye Farhangi-ye Farsi/ASEA* (Persian cultural thesaurus). Tehran: IRECD.
- Khorramshahi, B. (1994) *Sar'onvanha-ye Farsi-ye pezesghi* (Persian medical thesaurus). *Karane*, 1(3-4), 142-4.
- Rahadoust, F. (1998) *Estelahname-ye Pezesghi Farsi* (Persian medical thesaurus). Tehran: National Library of Islamic Republic of Iran.
- Saintville, D. (1977) *Guidelines for the indexing cultural documents and developing a thesaurus on cultural development in Asian countries*. Translated by A. Azarang as *Rahnamaye namaye sazi-ye madarek va gostareh-e estelahname-ye tose'e farhangi dar keshvarhaye Asia*. Tehran: Asian Cultural Documentation Center for UNESCO.
- UNESCO (1998) *International thesaurus of cultural development*. Translated by Abbas Horri and Parvane Gudarzi as *Estelahname-ye beinalmelali-ye tose'e-ye farhangi*. Tehran: Cultural Research Bureau.
- Viet, J. (1972) *Macrothesaurus: a basic list of economic and social development terms*. Translated by H. Abrami as *Estelahname-ye kalan-e sazman-e ta'avon va tose'e-ye eqtesadi*. Tehran: Plan and Budget Organization.
- Viet, J. (1979) *Cultural development thesaurus*. Translated by Naser Pakdaman as *Estelahname-ye tose'e-ye farhangi*. Tehran: Asian Cultural Documentation Center for UNESCO.
- Masoumeh Bagheri works for the Department of Library and Information Science, University of Tehran, Iran. Email: mbagheri@ut.ac.ir

*This article first appeared (with minor variation) in Indexing: Engage, Enlighten, Enrich: Proceedings from the ANZSI Conference, Melbourne, Australia, 17-19 March 2005. [See review p. 63 – Ed.]*

*[It is hoped to publish a companion article by Mansoureh Bagheri (Masoumeh's sister) on the training of indexers in Iran in the October 2006 issue – Ed.]*