

Case Study

Open Access

Transformation in the Quality of Lexicography with the Entry of International Standards (ISO's Documents): A Case Study of English, Persian and Arabic Dictionaries

Masoumeh Hajizadeh ^{1*}

1. Faculty member, Academy of Persian Language and Literature, Tehran, Iran

Abstract

In this article, by examining three dictionaries - the Comprehensive dictionary of the Persian Language; Oxford Dictionary; Al-Munjid Dictionary - from three authentic languages, English, Persian, and Arabic, we find that the ways of presenting entries, definitions, and overall information provided to users are different from each other in these dictionaries; this is the challenge facing compilers of print and online dictionaries. Considering that part of the input data of the translation machines represents dictionary of different languages, these different methods of compilation downgrades the performance of the translation machines, besides, they sometimes create insoluble challenges and difficulties for users. Unifying the way of compiling printed and online dictionaries with the help of ISO's documents can be a solution to this problem. In this article, we will specifically discuss the function and necessity of using the ISO's documents in the field of lexicography, which is currently being compiled and edited in a working group "TC 37/SC 2/WG 9" and is known as "ISO 1951:2007 Presentation representation of entries in dictionaries / revision".

Keywords Lexicography, machine translation, ISO's document

Introduction

Lexicography, as a scientific field that is directly related to the audience and users, has always been dependent on technology, although the process of aligning it with today's technology is still slow. Today, online dictionaries for different languages have become popular all over the world. Most users have a copy of the online dictionary on their smartphones that they can immediately refer to when needed. The evolution in the field of lexicography has been started for a long time and is following its way, and many studies have been done to improve the trans-

fer of information in this field to the user with the help of technology, but it is not very fast compared to other fields of science and technology has not fully entered this category. And still, the ways of compiling and presenting the data of this field are done in traditional ways, but experts in the field of lexicography are bound to accept this evolution and align with the global trend in this field. Considering that the general trend of human knowledge has always been towards improving the quality of human life, the use of media and computer science in this field, although it may have difficulties at the beginning, in the



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

* Correspondence:
ms.hajizadeh78@gmail.com

end the result is beneficial. Following these developments, the world of translation is rapidly changing, and this growing industry needs solutions for the optimal use of machine translation, as well as ways to carry out translation projects.

We can look at the evolution of lexicography from two perspectives: firstly, the impact of technology on the compiling and writing the dictionaries, and secondly, the impact of technology on the using dictionaries. In other words, there is a lexicographer on one side and a user of dictionary on the other. Since that translation machines need real data from different languages to improve their performance, different dictionaries should be considered as input data for these machines.

In this article, by examining three dictionaries - the comprehensive Dictionary of the Persian language; Oxford Dictionary; Al-Munjid Dictionary - we will show that the methods of providing entries and information provided to users are quite different in these dictionaries. The different methods will decrease the performance of translation machines, and in addition, they sometimes create unsolvable challenges and difficulties for them. One of the solutions to solve this problem is to standardize the way of compiling printed and online dictionaries in different languages. One of the goals of the International Standard Organization (ISO) is to unify and facilitate the exchange of goods and technical knowledge. The ISO's documents can be compared with linguistic universals in linguistic analysis and the symbols of the IPA table in phonological analyses; as they are necessary tools of linguistic analysis, ISO's documents also have a similar function in compiling dictionaries in different languages. In the field of language and dictionary compiling, we can deal with four groups of ISO's documents: the first and most important documents are related to the ISO's documents for people who compile dictionaries; second, the documents related to translators (people who translate texts for others and deal directly with dictionaries); third the document is specific to people who entrust a text to someone else for translation (indirectly dealing with dictionaries); fourth the document is related to the people who designed translation machines and their input is the output of dictionary authors.

In this article, we will specifically discuss the function and necessity of using the document of the first case, i.e. lexicography, which is currently being compiled and edited under the title of "ISO 1951:2007 Presentation representation of entries in dictionaries" in working group "37/SC 2/WG 9". This article is compiled in three general parts. In the first part, the compilation method of each of the Persian, English, and Arabic dictionaries - the Comprehensive Dictionary of the Persian Language; Oxford

Dictionary; Al-Munjid Dictionary - is briefly explained. In the second part, the challenges that the mentioned editing methods arouse will be clarified separately. In the third part, the efficiency of the document "ISO 1951:2007 Presentation representation of entries in dictionaries" will be mentioned in the standardization of dictionary compilation.

2. The methods of compiling dictionary

For the present study, the way of compiling dictionary in three languages, Persian, English and Arabic, is examined. For this purpose, Persian dictionary with the title "A Comprehensive Dictionary of the Persian Language"; The English language dictionary with the title "Oxford Advanced Learner's Dictionary" and the Arabic language dictionary with the title "Al-Munjid fi al-loghatah al-arabiah al-mo'aserah" have been selected. In the following, these three dictionaries are introduced in three separate sections, and the information that each of these dictionaries provides to the user, as well as the arrangement of this information, is introduced.

2-1 The method of compilation of the Persian Language dictionary (A Comprehensive Dictionary of the Persian Language)

In the comprehensive dictionary of the Persian Language, we are faced with two types of entries: main entry and sub-entry. In this dictionary, the most common spelling and the written form of an entry are chosen as the main entry. Other spelling forms are entered separately and referred to the original form. The entries in this dictionary are arranged in alphabetical order. A main entry can have several sub-entries; however, sometimes the main entry does not have a sub-entry. In the macro structure of each entry, all the semantic distinctions are listed first, and then the sub-entries are listed in alphabetical order. In this dictionary, nouns are entered in the singular form, except for Arabic plurals; verbs are also entered as infinitives.

As mentioned, each entry begins with the written form of the headword. After that there is the pronunciation of the headword, which is the most common pronunciation among Persian speakers. If a word has several common pronunciations, all of them are mentioned in order; however, the old pronunciations of a word will appear in the supplementary information section. Due to the fact that the place of the stress is fixed according to the part of speech of the word in Persian Language, therefore the sign of stress is not used. After the pronunciation of each entry, there is a section called supplementary information. Supplementary information comes between two brackets, and its first part includes information such as: short form; spoken form; homophone; informal form; ab-

abbreviated form; other spelling ways; plural; comparative form; superlative form; present form of the verb; imperative form of the verb; etymological information. The second part of the supplementary information is placed after the etymological information with the title of additional information, which includes information such as: specific phonetic information, historical-geographical information, information related to the approved words of The Academy of Persian Language and Literature, and so on [1].

After additional information, different senses of an entry will be presented separately. But before the senses, the part of speech of the entry is mentioned in parentheses. Sometimes an entry has several parts of speech, each of them comes in a separate section, and different senses of each part of speech come separately below it. If an entry has several parts of speech, numbers in words are used to distinguish different parts of speech. Different labels are used to inform the user about the specialized field of the entry, such as: mathematics, philosophy, botany, astronomy, zoology, etc. Different styles such as spoken, formal, colloquial and the like are mentioned here. Additional explanations related to one or more senses also come in parentheses before the main definition. If the entry is used with preposition(s), all of them are shown in parentheses. For each sense, a descriptive or referential definition is provided first. If the entry has a synonym or antonym, or if it needs to be compared with another entry, it will be placed at the end of the definition. Then, its examples turn; they will be arranged in chronological order. Since the comprehensive dictionary of the Persian Language is a historical dictionary and the corpus to which the dictionary is compiled according, includes old texts from the fourth lunar century, therefore, the examples related to each sense is given in chronological order, and this historical order of evidence is a criterion for numbering semantic distinctions. For each example, information such as the year in which the text was published, the author's name and the page number of the related book or magazine are recorded in parentheses at the end [1].

Referential definitions are also very common in this dictionary, and one of the methods of definition is to refer to a similar definition in another entry. In this dictionary, several different ways are used for referring the meanings; a reference within an entry used by a hand sign (□) or a straight arrow (←), a reference to a specific sense within another entry by an arrow sign, or using the given word in boldface, which is usually used to indicate synonyms. Also, if entry A is referenced to entry B, to show this two-way relationship, at the end of the definition of entry A, entry B is shown in boldface [1].

After the various separations of a main entry and finishing

the information related to the main entry, the sub-entries come in alphabetical order and each of them is semantically separated like the main entry; the necessary examples come after each separation as well. Lexical units larger than words are selected as sub-entries. Sub-entries can be composed of two to several words that have a high frequency of occurrence. Sub-entries are selected according to the rules and principles specified by the compilers of this dictionary. For example, the frequency of occurrence of two or more words together is one of the criteria for choosing a combination as a sub-entry.

In the Comprehensive Dictionary of the Persian Language, in addition to the different spelling forms that are recorded as entries, other varieties of an entry are also recorded. Different written forms of a word that have changed over time are known as different varieties of the given entry. The difference between varieties and different spellings is that different spellings of the same entry have the same pronunciation although different varieties of the same entry differ both in terms of pronunciation and written form. Different varieties of an entry are entered separately and different senses are semantically separated, but its definitions are referred to the main entry. Short form, colloquial, homophones, spoken forms and abbreviated forms are also recorded in different entries separately and referred to the main entry. At the end of some entries, there is a section called other compounds or references. Considering that the headword for the sub-entries is selected according to its first word of the group, so if the headword is not the first word of the phrase, the desired phrase comes under the relevant headword, and to maintain the connection, it is referred to as other compounds here.

2-2 The method of compilation of English dictionary (Oxford Learner's Dictionary)

The Oxford Learner's Dictionary is chosen for this research because it is the most comprehensive dictionary of the English Language. The Dictionary aims to cover the full spectrum of English language usage, from formal to slang, as it has evolved over time. The Oxford Learner's Dictionary has a highly organized structure and most entries contain information on all of the following: spelling, pronunciation, derivation, meaning, and usage. The usage of each word, meaning, or idiom in the Dictionary is documented through comprehensive examples drawn principally from quotations taken from printed texts of the present and the past. Main entries give users the most comprehensive information about the most recent, or main form of a word, phrase, combination, abbreviation, letter of the alphabet. The main entry is made up of two major parts, a generic or 'headword' section and 'sense'

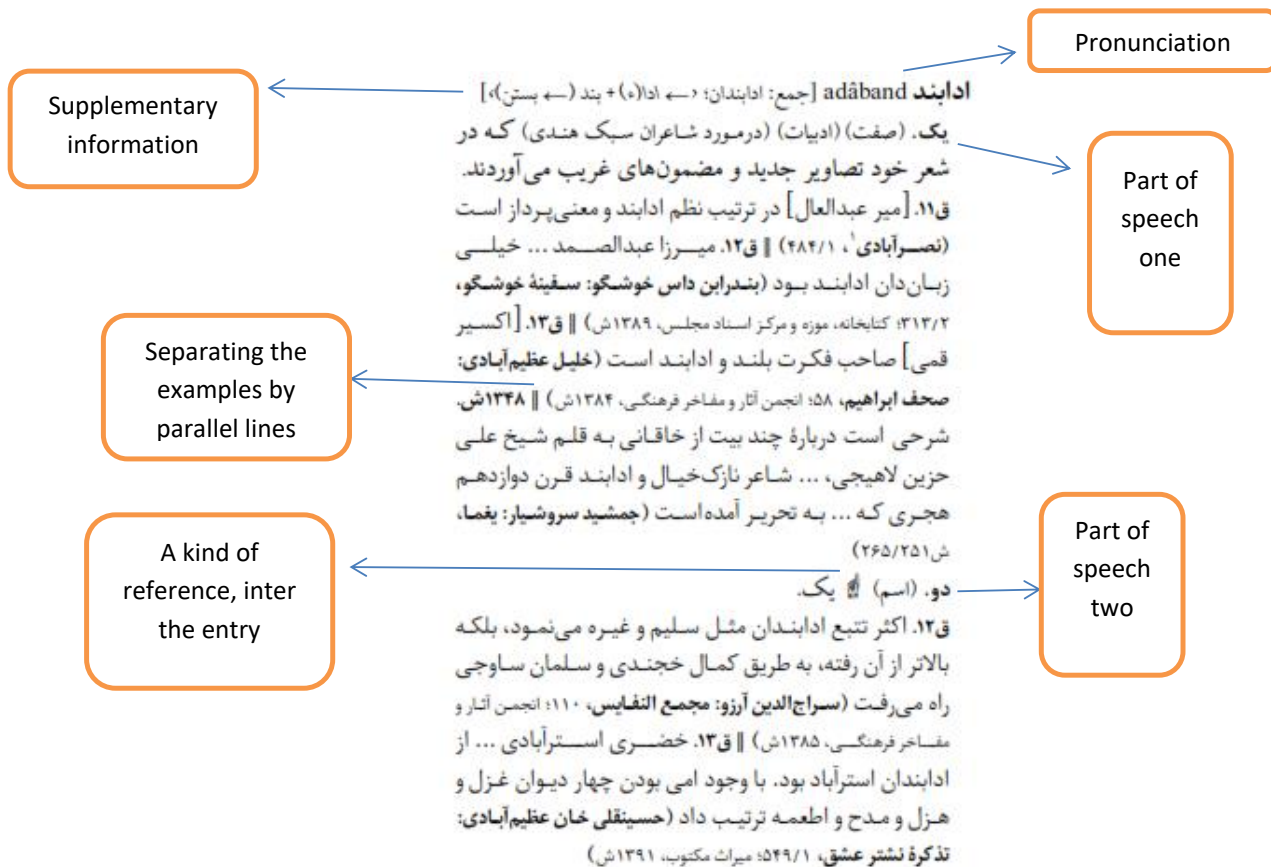


Fig 1. A sample entry of the Comprehensive Dictionary of Persian Language

section. Compounds, special uses and derivative are other parts which can be found in some entries [3].

The headword is the subject of a dictionary entry. The headword part focuses exclusively on general information about a word - its spelling, pronunciation, and grammatical forms, its origin, evolution, and other information that applies to the entire entry and not necessarily to a particular sense or meaning within it. The headword constitutes the first part of the entry. The headword shows the most common spelling of the word. Sometimes two alternatives are given. The first after headword is a kind of symbol named status symbols which are visual guides that alert the dictionary user to particular characteristics of a word, meaning, or usage. For example, a status symbol will indicate whether a term is a foreign word that hasn't been fully naturalized either in form or pronunciation. Status symbols appear before the sense number (or before a part of the meaning). The status symbols used in the headword section are a dagger, which indicates obsolete status, and vertical parallel lines, which indicate non-naturalized terms. Next is pronunciation; it is a kind of feature that shows the standard pronunciation of the headword. The pronunciation is represented by the sym-

bols of the International Phonetic Alphabet [3].

Entries usually present British English pronunciation as well as American English pronunciations, and those of other regions where the use of the term is significant.

The other information is about part of speech that explains the headword's function in a sentence; that is, whether it is used as a verb, noun, adjective, adverb, article, preposition or other part of speech. Some entries have several parts of speech. In these cases the information about each of these aspects of the headword are collected together in one entry but in separated parts.

The homonym in Oxford Dictionary is shown by an index known as homonym number that is an index number, used for reference purposes, to distinguish two or more headwords with the same spelling and part of speech, but different derivations.

The other information related to entry is shown by a label which is a term, usually displayed in italics, which gives some information, usually in abbreviated form, on the context in which that term is used. For instance, a label will give a term's regional origin (e.g. the United States, Australia), the subject area from which it derives (e.g. Biology, Chemistry, and Music), its grammatical function

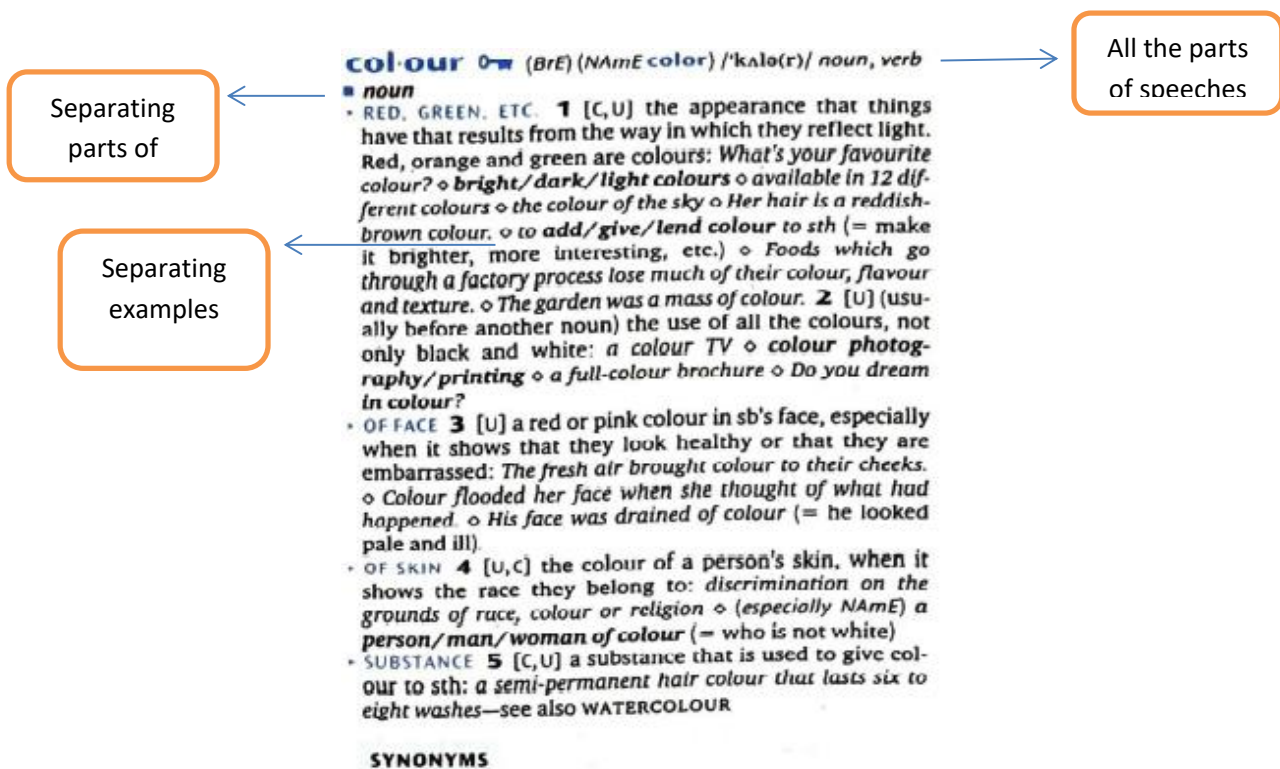


Fig 2. A sample entry of Oxford Dictionary

(e.g. plural, collective) and the type of meaning assigned to a word in a particular context (e.g. figurative, specific). The other labels are: figurative, formal, ironic, offensive, old-fashioned and so on.

Variant forms are the alternative spellings in which a word has been found over the centuries. Centuries of use are denoted by the first two digits of their years, e.g. 17 denotes the century 1700-1799. Old and Middle English spellings are denoted OE and ME respectively (sometimes with the prefixes 'e' (= 'early') and 'l' (= 'late')). The other part is related to etymology that is the derivation of a word. The etymology section explains how a word became part of the English language. The etymology part may also give detailed information on the meanings, history, or aspects of grammar that are relevant to a word.

After the headword section there is sense part which explains the headword's meaning or meanings. In this section you can see one or more definitions, arranged chronologically. Some words, especially those that have existed for centuries, have acquired many senses. Because of this, the sense section for some entries is quite extensive. Senses are ordered according to a structure resembling a family tree, so that the development of one sense from another can be plotted. The individual senses are numbered within this structure for ease of reference.

Definitions and etymologies often include words displayed in capitals, and forming hyperlinks. These are cross-references to other entries where additional related information may be found. Sometimes cross-references appear in other contexts. Words are often linked to other words by meaning or origin. Tracing the paths laid by these cross-references is another way of discovering more about words and their interrelationships [3].

The paragraph of illustrative quotations known as quotation block contains a selection of authentic examples of usage illustrating a definition. The quotations document the history of a term from its earliest recorded usage, and can be extremely helpful tools for clarifying grammatical and syntactic aspects of a definition. The following are the main components of each quotation cited within quotation blocks: date of publication; author; title

Compounds which are known as single grammatical units are frequently collected together in a section or group of sections at or near the end of an entry. They are followed by a quotation block in which examples of each compound are presented in alphabetical order of the compound. Some major compounds are entered as headwords in their own right.

Derivatives are words formed from the headword by the addition of a suffix. These are often entered as the final section of an entry. Many derivatives are included

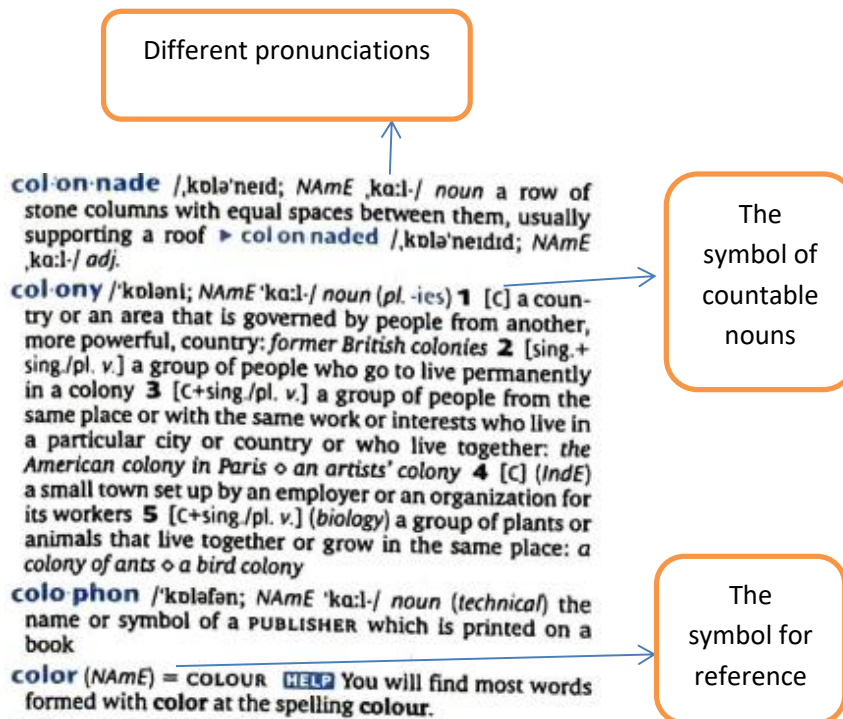


Fig 3. A sample entry of Oxford Dictionary

as headwords in their own right. They are followed by a quotation block illustrating examples the show how they are used.

Letters of the alphabet; initializes, acronyms, abbreviations; affixes and combining forms; proper names; erroneous, spurious, or ghost words are special types of main entries in Oxford Dictionary. In this dictionary many of them are entered as headwords [3].

2-3 The method of compilation of the Arabic language dictionary (Al-Munjid fi al-loghatah al-arabiah al-mo'aserah)

As it was mentioned previously, to show the way of compiling dictionary in the Arabic Language, an Arabic dictionary has been selected named "Al-Munjid fi al-loghatah al-arabiah al-mo'aserah", briefly known as Al-Munjid. The entries of this dictionary are not ordered alphabetically; they are ordered according to the roots of the words. So the main entries are roots of the words; and sub-entries are different derivations of that root. The roots of this dictionary are based on the alphabet and in each entry there are various derivatives related to the given entry along with examples to show usage. In each entry first there is the simple form of the word, and then there are different classes and derivations, names of place and time, nouns of tools, subject nouns, object nouns, adjectives, etc. After each of these parts, there are different meanings and then, if possible, an example to show the

usage of the word.

The authors of dictionary have also introduced abbreviations to introduce the words; Like "ms" (مص) for infinitive; "m" (م) for female; "sj" (سج) for plural nouns; "mth" (مثنى) for dual; "j" (ج) for plural and "jj" (جج) for plural of plural. In addition, instead of using specialized labels, they have used abbreviations to refer to the field of application of the word; For example: "R" (ر) for mathematics; "Teb" (طب) for medicine and medical sciences; "F" (ف) for physics; "H" (ه) for geometry and the like.

In the dictionaries arranged according to the root of the words, words are placed under their three-letter roots, and the roots are in alphabetical order, it is obvious that to realize and find the three-letter root (main letters) from extra letters (subsidiary), it is necessary to know the structure of the word. In this dictionary, the root of the verb is placed at the beginning and following it, the different forms of the verb, nouns and adjective are almost based on a numerical system. This means that first the simplest words are without extra letters, then words with one extra letter or long vowel, two extra letters, three extra letters, etc., and after that, other derivatives that usually start with the prefix "meem" (ميم) come.

We do not find some verbs such as the present participle of verbs or infinitives with extra letters in dictionaries. In general, it can be said that verbs that have a fixed form are rarely found in dictionaries. In the case of plural or singular forms, the singular form is entered first, and then



Fig 4. A sample entry of Al-Munjid Dictionary

the plural form comes with the sign “ج”(ج). A noun may have one or more plurals (jam’-e mokassar) in different forms. Some plurals (jam’-e mokassar) given in dictionaries may not be common; therefore, it is necessary to choose and use a plural that is common in Arabic language.

In this dictionary for presenting verb and noun and putting them in order, first the verb comes and then the noun or nouns derived from it without letters alef- lâm (ال). In the entries which the headword is a verb, first there are 3-letter verbs and then verbs with extra letters. In 3-letter verbs, the most famous and most used classes come first. The origin of loanwords comes in parentheses and sometimes a speculative root is considered for them. Entries that have two or

more pronunciations sometimes come together and then their different meanings appear. In the Al-Munjid Arabic Dictionary, the pronunciations of the entries are shown by using the diacritic. A large solid circle is used to show the root, which is considered as a main entry, and a small solid circle is used to show sub-entries.

For a better understanding of what was explained above, an image of an entry from Al-Munjid Dictionary is presented below and different parts are shown in the figure.

3. Challenges facing online dictionaries

It is clear that accessing data in online dictionaries is much easier and faster than printed one, and searching for infor-

mation have become much better and more efficient than before. However, the question that is currently aroused among researchers in this field is whether there has really been a change in lexicography or whether the way of presenting a dictionary has merely changed from print to digital one. Some believe that electronic and online versions are the same as printed dictionaries, and the only advantage they have is that they are presented only in the web environment [6]. Even seemingly fully structured online dictionaries have simply copied definitions and there is still no standard approach.

Other development that can be expected following the digitization of the way of compiling a dictionary is the transformation of the traditional independent dictionary, printed or online, into an integrated dictionary. It is a dictionary integrated with other information tools such as e-readers, writing assistants or learning tools [7]. Simple and optimal access; increasing the amount of available data; removing the border between dictionaries, encyclopedias and thesauruses; and integrating several dictionaries in one dictionary are other capabilities that digital dictionaries are expected to have in the future.

Nevertheless, lexicographers are trying different methods; they know that reaching a structure and standard method for online dictionaries is essential [8]. Many researchers are researching new lexicography methods and believe that lexicography with a long history has always depended on modern technology to present its products, although technology may initially affect the production of high-quality products. Technology has an impact on this scientific field from several aspects: production of dictionaries, presentation of dictionaries, application of dictionaries and economic issues. Lexicographers in this era are dealing with a bigger challenge, which is to engage and work with programmers and designers to ensure the quality of the final product [7]. The realization of the online dictionaries as well as the integrated dictionaries discussed above requires solving all the challenges faced by print dictionaries in order to lay a foundation for online versions. The lack of uniformity in compiling dictionaries is one of the challenges facing this field. It is better to achieve this uniformity first in print dictionaries and then apply it in online dictionaries as well. Considering the fact that printed dictionaries are still available and used by the public more than online dictionaries, as well as the fact that online dictionaries are often the same printed versions that are available to the user online, it is better to initially solve the problem of print dictionaries.

It was mentioned earlier that in order to find these challenges in this research, three dictionaries were selected from the three languages of Persian, English and Arabic and the structure of its entries was fully investigated. In the previous section, the structures of the entries and the information they

provide to the user are briefly described. It is worth mentioning that among these three dictionaries, the Arabic Language dictionary has a more complicated situation. Apart from the egregious differences that can be observed in the compilation of this dictionary compared to the other two dictionaries, the lack of uniformity in the way of the compilation of Arabic language dictionaries with each other is quite evident. Feten Baccar Ben [9] has mentioned this issue and considered it a challenge even for the user of Arabic language dictionaries. He believes that there is no uniform way of using Arabic dictionaries and users are forced to know the internal structure of each Arabic dictionary completely; in other words, before using any Arabic language dictionary, user must study the guidelines and the way of using it completely.

The following, it is pointed out some differences that were observed by examining three Persian, Arabic and English dictionaries - the Comprehensive Dictionary of the Persian Language, the Oxford Dictionary, and the Al-Munjid Dictionary.

1. There are different ways to choose an interlocutor in all three dictionaries. In "Al-Munjid" Dictionary, which is a root-based dictionary, the way of choosing the entries is completely different from the other two dictionaries. As explained before, the user must first find the root of the desired word in order to find the headword and then access the desired word through it.
2. The symbols used to represent different information in the entry structure are different in all three dictionaries. Examples of these items are shown in figures 1-4.
3. In both the Oxford Dictionary and the Comprehensive Dictionary, the pronunciation of the entry is shown with the help of IPA symbols; but in Al-Munjid, it shows the pronunciation with the help of diacritic.
4. In the Oxford Dictionary, if an entry has several grammatical identities (parts of speeches), it mentions them in front of the entry after pronunciation and shows the change of parts of speeches with a solid square (■), but the Comprehensive Dictionary of Persian Language uses the number in word to show the border of the beginning of the new part of speech.
5. The Comprehensive Dictionary of Persian Language and the Oxford Dictionary have used several methods to refer to other entries: the Comprehensive Dictionary of Persian Language uses symbols like the arrow sign; hand sign (within an entry); and words in boldface, however, the Oxford Dictionary uses the equal sign (=) or the phrase "refers to...".
6. The Comprehensive Dictionary of the Persian Language lists the prepositions used with a word or a phrase in parentheses and mentions that these are prepositions, but the Oxford Dictionary gives them with the help of several signs and also specifies the place where the prepositions are added with the help of signs.
7. If a word has a different spelling with a different pronunci-

ation, the Comprehensive Dictionary of the Persian Language identifies such cases with the title of other varieties separates their different meanings and refers them to the original form.

8. The Comprehensive Dictionary provides a lot of information at the beginning of the entry in relation to the headword, which is not seen in the other two dictionaries. For example, the information related to the internal structure of the headword. In cases where such information is observed in the other two dictionaries, the way of presenting them is different in all three dictionaries.

9. The selection criteria of entries and main entries are different in all three dictionaries.

10. Each dictionary has a different internal structure for entries; in other words, the different components and parts of the entry are different in all three dictionaries.

There are a number of examples of the many differences that can be seen in the compilation of the three dictionaries in question - the Comprehensive Dictionary of the Persian Language, the Oxford Dictionary and the Al-Munjid Dictionary. Each of these differences, although seemingly unimportant, will be difficult for users and programmers. Therefore, it is necessary to minimize these inconsistencies.

4. The role and effectiveness of document "ISO 1951:2007 Presentation representation of entries in dictionaries/revision"

The International Standard Organization (ISO) is an independent and non-governmental organization that aims to facilitate the exchange of goods, services and technical knowledge at the global level. The development of technical knowledge and the creation of a single language for transferring knowledge, as well as facilitating international trade and raising quality are the main goals of this organization. The documents of this organization are prepared under the title of standard documents in various specialized fields, and committees consisting of experts in each field supervise their compilation and contents. To compile and approve the content of each standard document, various working groups of experts in the desired field are formed around the world, and by holding numerous meetings, a document is finally prepared for the desired field. So, the finalization of the desired document is sent to all standard organizations in that particular field around the world and then it is approved by all members and experts around the world. The structure of the documents is almost the same. Each document is made up of different sections, the ultimate goal of which is uniformity in terms and concepts used in specialized fields around the world.

In the field of language, several working groups are preparing documents for this field. Technical Committee 37, which is known as TC 37 in this organization, is active in the field of language and terminology; this committee is composed of several subgroups and working groups, each of which is spe-

cifically responsible for one of the linguistic areas, including translation and terminology. "Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet (ISO12199: 2000)" and "Translation services Post-editing of machine translation output – requirements (ISO 18587: 2017)ISO 1951:2007" can be mentioned among the documents that have been compiled or are being compiled in the field of language.

Among these documents, which are considered in the present article and are being compiled in the field of lexicography, is the "ISO 1951:2007 Presentation representation of entries in dictionaries" document. This document is about the terms and concepts related to the compilation of the dictionary; this document is being reviewed in the working group "TC37/SC 2/WG 9". In the first part of this document, like other documents of the standard organization, with the title of "terms and definitions", an attempt has been made to provide specific definitions; in that a selection of lexicography terms should be provided.

With the help of this document, the macro structure of the dictionary, the micro structure, the nature of the main entries, sub-entries, tags, basic forms of the word will be unified with uniform definitions worldwide. The specialized labels required in the dictionary are precisely defined in this document. Senses and meanings and even examples used for different meanings in dictionaries are selected and presented based on specific and uniform principles and rules. In the next part of the document, the internal structure of each entry is described in detail. What are the main headwords (main entries) and how are they selected; what characteristics should the sub-entries have? In another section, signs and symbols are considered in detail to indicate some features in the entries with the help of these signs, for example, the use of white diamonds to indicate dialects that have the same origin. In the final part of the document, there is also information about how to make references and cross-references in entries as well as the way of showing different pronunciations and the like.

The above promises that by compiling the document "ISO 1951:2007 Presentation representation of entries in dictionaries" in the working group "TC37/SC 2/WG 9" and following and modeling it by lexicographers around the world, we will achieve more uniformity in the field of lexicography. By compiling uniform dictionaries, both the challenges faced by users and the challenges of compilers of online dictionaries are removed. In order to use the dictionary in any language, the user no longer needs to know the editing method and signs used in each dictionary in particular; all dictionaries provide information to users in the same way and with the same signs and symbols. The users of dictionaries are not the only ones who easily grasp different dictionaries and use their information; compilers of online dictionaries, followed by machine translation programmers, are two other main groups that will benefit from the uniformity of editing dictionaries.

5. Conclusion

The purpose of this research is to show the effectiveness of the document “ISO 1951:2007: Presentation representation of entries in dictionaries” that is being compiled in working group “TC 37/SC 2/WG 9” in the uniformity of compilation of printed dictionaries and ease of online lexicography, which will ultimately lead to machine translations with more accuracy and reliability. To do this, three dictionaries of Persian, English, and Arabic languages were selected, the Comprehensive Dictionary of the Persian Language, the Oxford Dictionary, and the Al-Munjid Dictionary. By examining the way these three dictionaries are compiled, as well as examining their entries and sub-entries and their internal structure, it was found that there are many differences in the structure of the entries, subentries, definitions, reference methods and finally their compiling method. What was observed by examining these three dictionaries indicates the high efficiency of ISO’s documents in facilitating compiling dictionaries. This research showed that we need a single language in the field of lexicography as well as transliteration; a single and common language among all the compilers of dictionaries around the world; the same language as IPA symbols. The structural diversity of the available dictionary resources made the exchange and the fusion of their data obviously difficult and complex. This is why creating a global language by standardization in lexicography is necessary. By creating and applying a global language in lexicography, moving towards online lexicography will be facilitated. Creating ISO’s documents in lexicography is a way of attaining this consistent language. Specifically, unified methods in micro structure and micro structure of entries; applying the same method in defining the words, cross-referencing, choosing the entries and sub-entries and alike. In this article, only three dictionaries from three languages were examined, if other dictionaries of different languages are also examined, we will realize more and more the inconsistencies in the way a dictionary is compiled. By compiling standard documents and applying them on print and online lexicography all over the world, a big step will be taken on the way of compiling dictionaries; as well as experiencing a big change in this field. It was also mentioned in the introduction that other working groups can be considered for the field of translation and dictionary and their final product can be used in these fields as well.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: Jul. 2023 Accepted: Sep. 2023

Published online: Sep. 2023

DOI: 10.22034/ASAS.2023.410799.1031

Reference

- [1] The Academy of Persian Language and Literature. A Comprehensive Dictionary of Persian Language. Tehran: The Academy of Persian Language and Literature.2016.
- [2] The Academy of Persian Language and Literature. A Comprehensive Dictionary of Persian Language. Tehran: The Academy of Persian Language and Literature.2023.
- [3] Oxford English Dictionary. <https://www.oed.com/>
- [4] Oxford university press. Oxford Advanced Learners Dictionary. 7th. Oxford: Oxford University Press. 2010.
- [5] Maluf, lu'is. Al-Munjid Fi'al-Lagheh Al-Arabiyyah al-Mudhamdane. Beirut : Darol-Mashregh. 2007.
- [6] Granger, Sylviane. “Electronic Lexicography: From challenge to opportunity”. IN Electronic lexicography, by Sylviane Granger and Magali Paquot, 1-11. Oxford: Oxford University Press. 2012.
- [7] Fuertes-Olivera, P. A.; S. Tarp; P. Sepstrup. “New Insights in the Design and Compilation of Digital Bilingual Lexicographical Products: The Case of the Dictionaries ‘Valladolid-Va’”. Lexikos 28 (AFRILEX-reeks/series 28). 2018. <https://www.ajol.info/index.php/lex/article/view/182659>
- [8] Rundell, Michael. “From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions”. Lexikos 25. 2015. <https://www.ajol.info/index.php/lex/article/view/126250>
- [9] Ben Amar, Feten Baccar. et al. “LMF Standardized Model for the Editorial Electronic Dictionaries of Arabic”. Scitepress. 2008. <https://www.scitepress.org/papers/2008/17378/17378.pdf>

Submit your manuscript to Advances in the standards and applied sciences journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open Access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

**Submit your next manuscript at:
journal.standards.ac.ir**