The Alphabetic Abbreviation Process in Persian Scientific Texts

Fatemeh Yazdani
Department of Linguistics, Central Tehran Branch, Islamic Azad University, Tehran, Iran

Foroogh Kazemi
Department of Linguistics, Central Tehran Branch, Islamic Azad University, Tehran, Iran

Abstract—Abbreviating has a long history. Lots of new services and products are distributed in the markets daily due to the wide use of languages all over the world. Certainly, labeling these services and merchandise is quite an interesting subject in the domain of language. Moreover, regarding the increasing number of organizations and developments in science and technology, the need for accelerating the process of communication especially linguistic communication is a must. This study examines the process of alphabetic abbreviation and the extent of its application in some Persian scientific journals. The research data were gathered from The Pulse of Economy, Air Industry and Future Trade magazines and were analyzed by the analytical research method. Among the mentioned magazines, the Air Industry had the highest number of alphabetic abbreviations; Future Trade was in the second place and The Pulse of Economy was the third in rank. In this article, the use of alphabetic abbreviation was analyzed in Persian and English languages separately and it was determined that 99 percent of abbreviated words were used in English scientific journals whereas in Persian journals this number was only 1 percent. This indicates the high role of borrowing in Persian texts. An explanation for the higher use of abbreviated elements is the scientific genre, which indicates that message transmission in the least possible time period is the main characteristic of any human communication, especially scientific communication.

Index Terms—morphology, word formation process, alphabetic abbreviation, Persian magazines, scientific texts

I. INTRODUCTION

Language is a social exchange instrument through which people interchange their thoughts, their experiences and their ideas. In fact, language is an instrument through which the members of a society can interact and the result of which is a social unit. Mass media plays a significant role in this regard. By taking the tools of mass media into account, we see that the press is the most important element in the formation of the public mind. This can be the consequence of different reasons—excluding its long-lasting effect, its availability at any time and place, its cheap price and the variety of its content, which can address different groups and classes of people; The result is an effect on the public mind on a vast scale. For a long time, due to the cultural interactions in human societies, and because of scientific developments, as well as the emergence of new phenomena and other considerable factors, new concepts have been entering the domain of language; In consequence, the need for word formation became an important issue.

Nowadays, with the arrival of new scientific explorations, the need for forming new words is felt more than before. In this paper, we have tried to study the alphabetic abbreviation in scientific journals from a linguistic point of view. One of the major processes in word formation is abbreviating. Alphabetic abbreviation is a type of various kinds of abbreviations, in which several letters or a set of letters replace a word or a long phrase. In general, we can consider the principle of economy or minimum effort in language as the main cause of abbreviation. This principle was proposed for the first time by Zipf in the field of communications science and linguistics (Zipf, 1606). According to this principle, human beings try to transmit the highest rate of information in the least possible amount of time. If we consider this principle in the domain of the structure of words and language, we can claim that some processes of word formation are done as a result of the principle of minimum effort. The main research question of this article is to determine the condition and the application of alphabetic abbreviation in Persian press especially the different scientific journals. The data collection process is done through library research. The research is done by an analytic method. This paper examines alphabetic abbreviation in The Pulse of Economy, Air Industry and Trade magazines. Three issues are selected from each of these magazines, so the sample size is 9 issues of Persian scientific journals.

II. REVIEW OF THE RELEVANT STUDIES

Taghva and Gilberth (1999) encountered a problem called acronym when they were designing a post processing system for output text information from optical character recognition tools. Abbreviations were the words which were not mentioned in lexicons, but were widely used in texts. A large part of the mentioned site was busy with recognizing
the wrong words. Because of the emergence of this problem, Gilbert developed a system for tracking these types of words. This system operated in four phases: the preparation phase, the input filter, chunking the remained input into words, and the function of tracking algorithm for the abbreviated words.

In other research, Schwartz and Hearst (2003) tried to develop a pattern for taking out information related to different words which challenged researchers with a big problem. By taking the high frequency of abbreviated words in scientific texts into account, the two researchers focused on finding these words. They proposed a simple algorithm which could quickly find abbreviations in biomedical texts. Their work was in fact, extracting words as well as their definitions in texts. Although they welcomed machine learning approaches in computational linguistics, they believe that in this specific field, a simple algorithm is more appropriate. In their idea, the simplicity factor of the algorithm means that it can get us to the aim as quickly as possible; For example, by using a simple algorithm, we can process thousand abstracts in one second and extract abbreviations from them.

Dannels (2005) states that because there are a lot of documents in the virtual world, finding the intended information would be very difficult for a checker. Computational linguists have tried to speed up the process of searching for and finding information by designing computer programs. Dannels tried to test the recognition of abbreviations in Swedish language medical texts. He has done it in a rule based manner and has compared it with different modes of machine learning. The findings of this research show that if the computer program have enough syntactic information and the pre-text processing get performed, the program can recognize the abbreviations used in medical texts of Swedish websites. Due to the wide use of abbreviations for naming equipments, methods, illnesses or drugs in science and medical sciences, this work can help a lot the researchers in medical fields to find information as quickly as possible.

Reza Golifamian (2008) has done a research in Persian and English abbreviations and the extent of their use of internet websites. In this research, it is claimed that there is a higher tendency to use abbreviations in developed societies in which life speed is higher and the people tend to apply the economy of language. To assess this hypothesis, some websites in both Persian and English were checked for the six fields of news, government, sciences, finance, sports and culture, and their abbreviations were counted. The findings of this research show that English has a higher rate of usage of abbreviations than Persian.

Playfoot, Izura and Tree (2013) in their Neuropsychologia article state that abbreviations are seen in different forms in English words like BBC, DVD, and HIV, which are pronounced literally. Some other abbreviations like NASA and AWOL are pronounced in general and irregular ways. Some limited studies have indicated that alphabetic abbreviation and initialism are accumulated along the main route of mental lexicon and there is a tremendous similarity between them, which is related to semantic processing. As a result, in general terms, semantic disorders are assessed regarding different types of abbreviations and the level of our understanding of each text is calculated in relation to the rate of our development regarding the level of words.

Piers, et al, (2017) state that medical faults are caused as a result of misunderstanding of abbreviations. According to the rules of selling drugs in drugstores, using abbreviations and badges on drug packages is not permitted and their application on drug packages is harmful. They have also stated that the meaning of abbreviations should be mentioned completely in case of their application on drug packages. They could identify 828 types of abbreviations and written badges and also designed a special software for checking abbreviations and written badges on drug packages.

III. THEORETICAL FRAMEWORK

A. Morphology

“Morphology is the study of the internal structure of words. Somewhat paradoxically, morphology is both the oldest and one of the youngest subdisciplines of grammar or morphology is the study of the combination of morphemes to yield words. Morphology analysis typically consists of the identification of parts of words, or, more technically constituents of words. We can say that the word nuts consists of two constituents: the element nut and the element s” (Haselmath & Sims, 2010). Morphology is a branch of linguistics which studies the internal structure of words and the relationships governing them. In other words, The study of the construction of a word is called morphology. It is the study of the form and the way that words are formed in a language. Each word is pronounced and conveys meaning in different ways. Moreover, each word is used in specific linguistic and social contexts that seem commonplace and obvious to the speakers. But, this concept is not quite clear and simple for researchers in language and they have to pay attention to different phonological, morphological, and syntactic field and the interaction of these fields together in order to be able to describe a word, its structure and its use in an orderly and systematic manner (Shaghagh, 2012, p. 9).

B. Word Formation Process

New words are formed by the processes of compounding, derivation, semantic expansion, word repetition, and so on. Some examples of word formation are mentioned here: 1) Acronyms: by putting the initial letters of words beside each other acronyms are formed. 2) Alphabetic abbreviations: it is read alphabetically-letter by letter. 3). Coinced words: are the newly formed words with no origins. 4) Clipping: is written in short forms, but read completely. 5) Shortening: in this case, we cut some parts of the word in order to make it short. 6. Blends: two words are blended in a way that the first part of the word is combined with the second part of another word in order to make a new word. 7) Generified words: using a proper name in a general sense. 8. Borrowing: getting words from another language directly or indirectly.
is called borrowing. 9. Conversion: in this case, the grammatical category of word is changed without formal, structure representation. 10. Derivation: this is a common process, i.e. adding a derivational affixs to a lexical base (Akmajian, et al, 2010).

C. Abbreviation

The speakers of a language try to abbreviate the spoken and written forms of the language in order to economize on energy and the needed space for expressing the necessary materials (Shaghaghi, 2012, p. 108).

D. Alphabetic Abbreviations

The process of alphabetic abbreviation is one of the ways of shortening the long expressions in a language. The first letters of the combining words of that expression are used as a sign for each word. In this case, each letter is pronounced one by one and separately. By using the process of abbreviation, we can write the long names of organizations and institutions on plates or the limited spaces like the door of cars (Shaghaghi, 2012: 108). The characteristic of these alphabetic abbreviations (or initials) is that each of their letters is individually pronounced (Akmajian, et al, 2010). Here are a few well-known examples:

- IT (information technology)
- ATM (automatic teller machine)
- AFC (Asian Football Confederation)

E. Scientific Texts

A scientific text is an article written specifically to explain or explore a scientific idea. Scientific texts are often found in science journals or textbooks (Ensar & Eyyüp, 2016). One of the main reasons for the development of every country is the level of its production and the effect of its scientific publications (Osare, 2003). Today, the number of articles and the published scientific magazines is a main index of growth and the scientific credit of a country at an international level. Scientific articles are one of the most important tools for industry and technology because they can establish a link between scientific centers and researchers on the one hand and transmit science from the university and research centers to the industry and production sectors on the other hand. In fact, scientific magazines are a link between university and industry. It is quite clear that scientific articles which are considered as the main source of scientific information have a special place in all the scientific works (Nowruzi & Alimohammadi, 2006).

IV. DATA ANALYSIS

1. The analysis of alphabetic abbreviations from the Pulse of Economy magazine

By examining 3 issues of the Pulse of Economy Magazine, 60 pages were studied for each issue and several different alphabetic abbreviations were taken out. Some of the alphabetic abbreviations related to current economic subjects are mentioned several times in this magazine. For example, on page 42 of issue 16, the following samples are mentioned.

Sample 1: dar doran tahrim faaliyat arz y bank keshavarzi seh barabar shodeh bud zira tanha dar sal gozashteh noh milyard dolar goshayesh elsı tavasot bank keshavarzi anjam shod

(During the economic sanctions, the activities of agricultural bank tripled because only in last year 9 billion dollars LC were opened by agricultural bank).

As we can see, LC is an abbreviation which is read letter by letter, is an initialism, and is pronounced in Persian as /elsı/. It is also the abbreviation of two words: letter and credit.

Sample 2: bank biemo aedgham technology ba hes lameseh ensan anech dar avalin lahzeh vurud be shoab bank biemo be nazar miresad tedad kam karkonan ast

In issue 17 of this magazine, on page 40 we read, (BMO bank, the combination of technology with tactile sensation in human beings, the low number of employees is what takes the attention in the first glance when entering the BMO bank).

In the above sample, BMO is an alphabetic abbreviation. It is pronounced as /biemo/ in Persian language. The letter B stands for bank, M stands for Montreal, and O stands for online. This abbreviation is read letter by letter.

Sample 3: bebinid dar faseleh dah sal gozashteh chand keshvar be ozviyat faal sazman dabelyutio dar amadand dar hamin ejlas akhir nayrobi afghanestan ham ozv faal dabelyutio shod

On page 33 of this monthly magazine we read, (see, in the past 10 years, how many countries were accepted as active members of WTO? At the recent Nairobi conference, Afghanistan was also accepted as an active member of WTO).

The abbreviated word WTO in the above sentence is composed of three terms of world, trade, and organization. This sample is also alphabetic, i.e. it is pronounced letter by letter. It is pronounced as /dabelyutio/ which means world trade organization.

Sample 4: hala va dar in sharayet elam shodeh sherkati keh mojry enhesary hagh pakhsh mosabeghat eyefsi ast

On page 38 of this magazine we read, (now and in this condition, it is announced that the company which is the exclusive proprietor of the right for broadcasting AFC races).

AFC is pronounced in Persian as /eyefsi/ and is an alphabetic abbreviation which is read alphabetically-each letter is pronounced separately. This term is composed of three words- Asian, football, and federation and means the Asian football federation.
The abbreviations extracted from these 3 issues are presented below:

- FDA, MIT, NYU, MRI, EBCL, JFC, AT, ISIC, ABS, EBD, AMC, SME, PR, DNW, POH, CKD, BMW, NFC, GT.

In general, 60 pages of each issue of the Pulse of Economy magazine were studied and 23 unique abbreviations were found in these three magazines. The frequency of these abbreviations was 85. Here, EBCL had the highest number of repetitions with 18 times of repetition in the data, and LC was in the second place with 17 times of repetition. Moreover, WTO was in the third place with a frequency of 15.

2- The analysis of alphabetic abbreviations extracted from the magazine of Air Industry

By considering 3 issues of Air Industry quarterly, 60 pages were covered for each issue and several alphabetic abbreviations were collected. Some of the most repeated alphabetic abbreviations in this magazine are presented below; For example, in issue 292 page 59 of this quarterly, the following text is presented.

Sample 1: anjoman eimeny tarabary ostoraliya eytiesbi darbareh beh ruzavary taein manategh jostojuy zir abi miguyad guruh danesh defaey va fanavary ditiesji baray nemunayee masir ehtemali virayeh manategh jostoju etelaat mojood dar moerd parvaz ra be surat koli baznegari kardeh ast

(The Australian association of transport security, ATSB, stands about updating the positioning of underwater search areas: the group of defense, science and technology of DTSG has completely revised the existing data about flight for making samples of the probable editing of search areas).

In the above sample, ATSB and DTSG are examples of abbreviations in which each letter is pronounced separately. ASTB stands for four words respectively - Australian, transport, safety and bureau. DTSG is the abbreviation of four words- department of defense, templates, science, and guide. They are pronounced as /eytiesbi/ and /ditiesji/ respectively.

ATS, is used in place of the Australian association of transport security and DTSG is used as a substitute for the department of science and defensive guidance.

Sample 2: bartarin tajhizkonandeh sal jabejaei zamini khavarmiyaneh jayezeh casb va car havaee aitipi

In issue 293 of this quarterly page 38, we read (the highest equiper of the land transport in themiddle east, the prize of air entrepreneurship ITP, 2007).

In Persian, it is pronounced as /aitipi/. ITP is an alphabetic abbreviation. I is the abbreviation of international, T stands for trade, and P is an abbreviation for policy. It means the policy of international business.

This alphabetic abbreviation is also read letter by letter.

Sample 3: elat in mozu niz gamhaeist keh nasa ba mosharekat barkhi marakez sanati va be onvan bakhshi az barnameh havanavardi dustar mohtist iarey khahad bardasht iarey dar sal 2015 va pas az yek doreh sheshsal keh bar ruy toseeh va azmayesh hasht fanavari dustar mohtist zist motemarkez bud

On page 56 of this quarterly, (The reason for this matter is also the steps that NASA will take with the partnership of some industrial centers as a part of the program for pro-environmental aviation-ERA. ERA, in 2015 and after a 6 year period of performance, focused on the development and a test of 8 pro-environmental technologies).

ERA is another instance which is used in the above mentioned text. In Persian, it is pronounced as /iarey/. This abbreviation is composed of three words, i.e. environment, relevant, and activity and means pro-environmental spaceshipnship.

The other extracted alphabetic abbreviations are as follows:

- NCO, INCO, FASC, CG, ROE, CAOC, PPL, CPL, AR, KDC, STA, LCA, ATR, SOC, NDT, IECS, ADS, VIP, EOS, ERT, EORD, AEW, DJ, UV, OMS, XT, DXB, PTY, DSCA, AI, WSO, IFR, VFR, MATC, ULD, CRU, AACS, EMS, DAO, QMS, QAT, UN, KPI, SOP, IMS, CSR, ESS, DQAP, DSV, DQA, SEAD, CMU, DSCA, DSSS, FHSS, DMSS, PRM, ONR, TTNT, UAS.

From the total pages studied, 65 unique alphabetic abbreviations were found in these magazines with a frequency of 139. For example, DAO and ATR had the highest frequency with a repetition of 9 in the data, and EOTS was in the second place with a repetition of 6.

3- The analysis of the alphabetic abbreviations extracted from the magazine of Future Trade

By analyzing 3 issues of financial weekly Future Trade (60 pages from each) several alphabetic abbreviations were recognized and studied. Some of the abbreviations related to the current economic and financial themes of the country, which were repeatedly mentioned in this magazine, were analyzed. For instance, in issue 178 of this weekly magazine, page 56, the following text is mentioned. For example, in issue 178 of this weekly magazine, page 56, the following text is mentioned.

Sample 1: majame sherkatha dar iran az do bod mored baresi gharar migirad aval sud naghdi ta ghsim shodeh dipies keh hamvareh az mohtemarin mavared moerd bahs miyan sahamdaran va karshenasan budeh ast

(abbreviations of the companies in Iran are studied from two aspects. First, the dividend cash profitof DPS, which is always a hot debate among shareholders and experts).

The abbreviation DPS, which is used in financial texts, is an abbreviation. In this instance, each letter is read separately. DPS is pronounced as /dipies/ in Persian. D is the abbreviation of department, P is the abbreviation of public, and S is the abbreviation for service. The meaning of this abbreviation is cash profit for each share dividend.)
Sample 2: banabarein safar aghay deyvid lipton moaven aval modir amel sandogh beynolmelali pul ayemef keh hudud do hafteh pish beh iran anjam shod faghat beh manay shoru hamkari dobareh ba sandogh beynolmelali pul ayemef nist

In this magazine, page 28, we read, (so Mr. David Lipton's trip, the first deputy CEO of international monetary fund, IMF, which happened two weeks ago to Iran, does not solely mean a start of cooperations with the international monetary fund, IMF).

IMF is pronounced in Persian as /ayemef/ which refers to three words - international, monetary, and fund. IMF is an alphabetic abbreviation which means the international monetary fund.

Sample 3: ancheh moshakhas ast nesbat mablagh vam beh arzesh vahed maskuni ya haman eltivi dar iran paein ast

In issue 179 of this magazine, page 58, we read (what is clear is that the ratio of the loan amount to the value of each residential unit or LTV is low in Iran).

LTV is formed of three words - loan, to, and valuation, and its translation in Persian is the loan for assessment. It is pronounced as /eltivi/.

Sample 4: anva moharekha eyties va mavad ravangardan jadid enpies va asarat an ruy salamat ensan pardakhteheim

Issue 178 page 34 of this magazine, (different stimulators ATS and new hallucinating drugs NPS and their effects on human health are examined).

They are pronounced as /eyties/ and /enpies/ respectively in Persian. ATS and NPS are alphabetic abbreviations. ATS is the abbreviation of three words- anti, tetanus, and serum. NPS is also composed of three words- new, psychoactive, and substances. The other extracted abbreviations are as follows:


In total, 60 pages of each magazine were studied and it resulted in finding 64 unique alphabetic abbreviations with a frequency of 256 in these three issues. For instance, IMF abbreviation, with a frequency of 60, had the highest number of repetitions in the data, and USSD with a frequency of 30 was in the second place.

Alphabetic abbreviations for the whole corpus

By considering 9 issues- with an equal number of pages (60 pages from each issue) from the magazines of Pulse of Economy, Air Industry, and Future Trade- a number of 152 unique alphabetic abbreviations were extracted. The number of repetitions of these abbreviations is mentioned in the following table:

<table>
<thead>
<tr>
<th>Row</th>
<th>The name of magazine</th>
<th>The number of alphabetic abbreviation</th>
<th>Frequency of repetition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pulse of Economy</td>
<td>23</td>
<td>85</td>
</tr>
<tr>
<td>2</td>
<td>Air Industry</td>
<td>65</td>
<td>139</td>
</tr>
<tr>
<td>3</td>
<td>Future Trade</td>
<td>64</td>
<td>256</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>152</td>
<td>480</td>
</tr>
</tbody>
</table>

Considering the above table, we can see the most alphabetic repetitions in the scientific magazine of the Air Industry, with a frequency of repetition of 139. The Future Trade magazine is in the next place, with a frequency of 256, and the Pulse of Economy magazine is in the third place, with a frequency of 85. In the following figure, the number of abbreviations is shown for each magazine separately:

Figure 1: The frequency distribution of alphabetic abbreviations in different magazines

Another categorization of abbreviations is done in Persian and English languages separately in order to show their applications in these magazines. The number of English and Persian alphabetic abbreviations for English and Persian magazines is mentioned separately in the following table:
As we can see, in the above table the English alphabetic abbreviations printed in Persian scientific magazines have a higher role in comparison to Persian alphabetic abbreviations.

In the following figure, the number of Persian and alphabetic abbreviations are shown separately in scientific journals:

![Figure 2: Frequency Distribution of Alphabetic Abbreviations in Persian and English for the Whole Corpus](image)

As we can see in the above figure, the English alphabetic abbreviations, with a rate of 99%, have a higher level of usage in Persian magazines and the Persian abbreviations have a lower usage in scientific magazines.

In this research, alphabetic abbreviations were studied separately in scientific texts along with advertising texts in these magazines, which have a general genre and are not necessarily scientific.

In this way, a comparison was made between the usage of abbreviations in scientific and non-scientific texts, the results of which are presented below. In the following table, the number of usages of alphabetic abbreviations in propaganda and the advertisements in the mentioned scientific journals is given:

![The Percentage of Alphabetic Abbreviations in Advertisements](image)

Regarding the above table, we can see that the number of alphabetic abbreviations found in advertisements is much lower than the alphabetic abbreviations extracted from the text. This is because this propaganda was of a general type and was also not necessarily about the genre text of the article and magazine. The following diagram shows the percentage of the usage of abbreviations in propaganda and advertisements in relation to the whole data:
As we can see, in the above figure, it is quite clear that about 13 percent of alphabetic abbreviations have been used in the advertisements of these magazines and most of the extracted alphabetic abbreviations were used in texts, with a rate of 87%.

V. CONCLUSION

In this paper, the process of word formation of alphabetic abbreviation was studied in some scientific Farsi magazines. The results show that this process has a high rate of usage in the studied Farsi scientific magazines. From a total of 152 extracted words, a number of 23 alphabetic abbreviations with a frequency of repetition of 139 were extracted from the Pulse of Economy magazine, and a number of 64 items with a frequency of 256 were related to the Future Trade magazine. The findings show that most of the abbreviations were in the Air Industry scientific journal. Moreover, in this paper, the application of alphabetic abbreviations was studied separately in both Persian and English languages, and it was determined that 99% of abbreviations in scientific journals were used in English form but Persian abbreviations were only 1%. This shows the high role of borrowing in Farsi texts. Other discussions show the usage of alphabetic abbreviations in the advertising section of the mentioned Farsi scientific journals. They show that only 13% of the gained abbreviations were in advertisements and 87% of these abbreviations were found in the texts of of magazines. The use of abbreviations is a quick and easy way of reading and comprehending linguistic materials. The reason for their application is that we want to transfer information to the addressee as quickly as possible. Due to the expansion of languages all over the world, a large amount of goods and services are distributed in the markets daily. It is quite clear that naming these products and services would be an interesting subject in language. Economic institutions, service companies, and corporations consider naming their products as an appropriate name is a key to their financial success. One of the characteristics of a good name in the world of propaganda is its brevity. This has pushed the researchers in the domain of language, especially those dealing with business, to go into the realm of abbreviation. The findings show that regarding an increase in the number of organizations and the development in science and technology, a requirement for accelerating communication, especially linguistic communication by the use of abbreviations especially alphabetic abbreviation is a necessity. The reason for this is that a choice of the shortened and appropriate form for long expressions speeds up the process of communication and adds to its beauty and frankness. In this way, the efficiency of a language in providing a communicative function (in the case of scientific texts here) is enhanced.

REFERENCES


© 2018 ACADEMY PUBLICATION


Fatemeh Yazdani was born in Iran in 1986. She is BS of Accounting at Islamic Azad University, Amol Center, and MA of linguistics at Islamic Azad University, Central Tehran Branch. Mrs Yazdani has written a book about linguistics.

Foroogh Kazemi was born in Iran in 1974. She is head of linguistics department and associate professor of linguistics in Central Tehran Branch, Islamic Azad University in Iran. She has published 42 articles (cited in google scholar) and 15 books in linguistics. Dr. Kazemi has been advisor, reader and referee of many linguistic theses and judge of articles in journal of linguistics. She is a member of research council in faculty of foreign languages at Central Tehran Branch. She is also a member of editorial board, quarterly of language and literature: Parneyan Kheyal. Her areas of interest are functional linguistics, sociolinguistics, typology and discourse analysis.