



University of
Sistan and Baluchestan



Uppsala University

A Fuzzy Analysis of Derivational Prefixes Productivity in Persian: A Corpus-based Study

Seyyed Zohreh Aftabi¹, Abbas Ali Ahangar², Hasan Mishmast Nehi³

¹ Department of English Language and Literature, Faculty of Literature and Humanities, University of Sistan and Baluchestan, Zahedan, Iran, Email: s_zohre_aftabi@yahoo.com

² Professor, English Language and Literature Department, Faculty of Literature and Humanities, University of Sistan and Baluchestan, Zahedan, Iran, Email: ahangar@english.usb.ac.ir

³ Professor, Department of Mathematics and its Applications, Faculty of Mathematics, Statistics and Computer Science, University of Sistan and Baluchestan, Zahedan, Iran, Email: hmnehi@hamoon.usb.ac.ir

Abstract

This study utilizes a fuzzy approach to explore the productivity of Persian derivational prefixes based on Baayen's three measures of productivity, namely realized, potential and expanding productivity for Persian prefixes extracted from Bijankhan's Persian corpus. To this end, 2.6 million words in the corpus were thoroughly examined. The results were then categorized using the fuzzy set theory. Moreover, to validate the authenticity of the concept of hapax legomena, two measures, i.e., potential and expanding productivity measures were recalculated considering the total number of once, twice and three times occurred words coined by each prefix in the corpus. The results are almost the same in both methods suggesting that prefixes *por-*, *bi-*, and *na-* were and are the most productive prefixes in Persian. Besides, the reliable outcomes of applying fuzzy set theory to the productivity phenomena indicates the availability of this theory for dealing with other vague linguistic concepts for which there are discrepancies among linguists.

Keywords: Productivity, derivational prefixes, Baayen's productivity measures, fuzzy set theory, Bijankhan's Persian Corpus

1. Introduction

The study of morphological productivity, especially derivational prefixes, offers profound insights into the dynamic nature of language. Bauer (2001, p. 21) refers to productivity as “the property of an affix to be used to coin new complex words”. Over time, various researchers (e.g. Aronoff, 1976; Anshen & Aronoff, 1988; Baayen & Lieber, 1991; Baayen, 1992, 1993, 1994, 2009; Renouf 1996, among others) have proposed quantitative measures to evaluate affix productivity. While these methods illuminate different facets of productivity, they all have their own limitations. One basic method equates productivity with the number of words coined by a particular affix. However, this approach only reflects the degree of generalization and past productivity, failing to indicate an affix’s potential to form new words (Bauer, 2001, p. 144). Early attempts to quantify affix productivity include Aronoff’s (1976, p. 36) suggestion to calculate the ratio of actual words formed by a word-formation pattern to the potential words that could be created. Nevertheless, as Bauer (2001), along with Baayen and Lieber (1991), observed, this measure is impractical because counting all potential words and actual types in a corpus is unfeasible. To address these challenges, Baayen (1992, 1993, 1994, 2009) introduced three measures of productivity: realized, potential, and expanding productivity. These measures, grounded in the concept of hapax legomena (words that occur only once in a corpus), have become pivotal in morphological productivity research.

Beyond morphology, fuzzy set theory has been widely applied to fields involving vague concepts, such as linguistic hedges, cognitive linguistics, and semantics (Labov 1973; Lakoff 1973; Joyce 1976; Osherson and Smith 1981, 1982; Kamp and Partee 1995; Lasersohn, 1999, among others). However, its application in morphology, especially in Persian, remains underexplored. This study aims to fill this gap by conducting a fuzzy analysis of the productivity of Persian derivational prefixes using a corpus-based approach. In fact, The research problem addressed in this study is the quantification and analysis of the productivity of Persian derivational prefixes based on three measures of productivity proposed by Baayen (2009) using Bijankhan’s (2004) Persian Corpus.

Despite some research on morphological productivity, there is also a need to understand how prefixes function within Persian, particularly through a fuzzy analytical lens. Previous studies have typically examined a limited number of Persian prefixes and suffixes in a mixed and undifferentiated manner, which has restricted the depth of their findings. In contrast, our study examines a substantial number of prefixes, allowing for a more comprehensive analysis. Moreover, while other studies have often assessed productivity based on a single criterion, our assessment is grounded in the three measures of productivity proposed by Baayen (2009), namely, realized, potential, and expanding productivity. Furthermore, applying Fuzzy logic, introduced by Zadeh in the 1960s, offers a mathematical framework capable of handling ambiguity and gradation, making it particularly suitable for this type of analysis. Thus, this multidimensional approach provides a better understanding of derivational prefixes productivity.

In line with the aforementioned explanations, the present study tries to find answers to the following research questions:

- 1) Which Persian derivational prefixes demonstrate the highest and lowest levels of realized productivity?
- (2) Which prefixes exhibit the greatest potential productivity in generating new lexical items?
- (3) Which prefixes show the highest levels of expanding productivity within the corpus?
- (4) How can Fuzzy Logic be applied to classify and compare the productivity levels of Persian derivational prefixes based on the three measures mentioned?

To answer these questions, the study employs the three productivity measures proposed by Baayen (2009)—namely, realized productivity, potential productivity, and expanding productivity—as a benchmark for assessing the productivity of Persian derivational prefixes in Bijankhan's (2004) tagged Persian corpus. Subsequently, the productivity values obtained from these measures are analyzed within a fuzzy-set framework and categorized into five productivity levels. Given its ability to model gradience and handle linguistic uncertainty, fuzzy set theory, originally introduced by Zadeh in the 1960s, provides a suitable mathematical framework for capturing the continuum of productivity rather than treating it as a strictly categorical phenomenon. This combined corpus-based and fuzzy analytical approach offers a more comprehensive understanding of the productivity patterns of Persian derivational prefixes. It should be emphasized that the reliability of the concept of hapax-legomena proposed by Baayen is also verified by a new approach that is comparing it with *dilegomena*¹ and *trilegomena*².

This article is structured as follows: after the introduction section, the second section provides a detailed review of the literature on morphological productivity and linguistic works that utilized fuzzy logic for their analysis. The third section describes the theoretical framework including the measures proposed by Baayen and fuzzy set theory. The fourth section presents the methodology, including the corpus-driven approach and the use of Bijankhan's corpus for data extraction. The results obtained from calculating the productivity degrees of prefixes for all the three measures and their fuzzy categorizations are presented in Section 5. The final section concludes with a summary of the research contributions and potential directions for future studies.

2. Literature Review

The study of derivational prefixes is essential for understanding morphological processes and lexical productivity of a language. In Persian, as in many other languages, prefixes contribute significantly to word formation and semantic variation. Previous research highlighted various aspects of derivational morphology, yet a comprehensive analysis of the productivity of Persian

¹. The authors' terminology for twice occurred words in the corpus.

². The authors' terminology for the three times occurred words in the corpus

prefixes remains underexplored. This section reviews key studies on morphological productivity, focusing on Baayen's productivity measures—realized, potential, and expanding productivity—and their application across linguistic contexts, setting the stage for this study's corpus-based analysis of Persian derivational prefixes.

Gaeta and Ricca (2003) applied Baayen's productivity measures, using hapax legomena to study frequent Italian affixes in La Stampa (a corpus of approximately 75 million tokens). They proposed a variable-corpus procedure to ensure consistency in productivity comparisons by analyzing affixes at equal token counts. Their findings highlighted the impact of derivational cycles, showing that controlling for high inner-cycle derivations minimized the effect on productivity counts. Italian prefixes *ri-* and *in-* were identified as the most productive in their domains.

Badakhshan (2010) focused on compounding in Persian morphology, using Plag's (2003) framework and data from the Institute of Humanities and Cultural Studies in Iran. He concluded that productivity is scalar and inversely related to constraints (morphological, phonological, syntactic, or semantic) imposed on word formation processes.

Faragordi (2010) investigated the frequency, productivity, and semantic scope of Persian prefixes using Bauer's (2004) productivity framework. She categorized prefixes as unproductive (frequency <50), semi-productive (frequency 50–200), and productive (frequency >200). While her categorization was insightful, she did not provide a rationale for the numerical thresholds.

Abbasi (2016) analyzed the structural and pragmatic constraints on derivational affixes in Persian. Using Baayen's (1994) hapax-conditioned productivity measure, she found an inverse relationship between affix productivity and the constraints imposed on affixation. High-productivity affixes often had compound bases, while low-productivity affixes were more likely to attach to simple bases. Abbasi also examined syntactic categories and subcategory changes associated with affixation, showing that fewer syntactic constraints correlated with higher productivity. However, her study was limited to a single productivity measure, whereas our study incorporates all three measures proposed by Baayen (2009).

Abedini, Moloodi, and Khormaei (2019) who considered productivity as a scalar concept tried to rank 19 derivational prefixes of Persian in terms of productivity based on Baayen's (2009) global productivity measure and Rosch's (1977) prototype theory. They made use of Hamshahri Persian corpus and specified the category-preserving / category-changing properties of each prefix as well as the prototypical lexical category of the input for each prefix. At the end, they decided that *qejr-*, *na-*, and *bi-* were the most and *bol-*, *la-* and *forud-* were the least productive derivational prefixes of Persian. It is worth mentioning that just like Abbasi (2016), they only considered one of Baayen's proposed productivity measures. Moreover, their so-called prefix *bol-* does not seem to the authors, at least, to be a prefix in contemporary Persian.

Aftabi, Ahangar and Mishmast Nehi (2021) introduced a fuzzy way of dealing with the vague concept of suffix productivity in Persian. Using Baayen's (2009) three measures of productivity for

each suffix in BijanKhan's corpus, they gave each suffix a degree of membership to different productivity categories. Accordingly, they indicated the possibility of belonging simultaneously to different fuzzy sets for each suffix. The results of their study signified that suffixes *-i* and *-e* in Persian had the highest degree of membership to the most productive set.

Kohanzad, Fallahi and Pahlevanzadeh (2021) studied the productivity of 166 Persian affixes based on only one of Baayen's measures, that is general productivity measurement $P^* = VN(1, c)/h$. In this formula (VN) stands for hapax legomena for each affix and (h) stands for all hapax legomena for all affixes. They also examined the results of their study based on Bauer (2004) about productivity and predictability of words, to show that the more frequent the affixes of words are, the more predictable their meaning will be. Results showed that the theory is accepted in most cases but some examples do not match.

3. Theoretical Frameworks

In the present research, Baayen's (2009) three measures of productivity, that is, realized productivity, potential productivity and expanding productivity are adopted due to their comprehensive and multifaceted approach. These measures have been widely recognized and applied for estimating morphological productivity in corpus-based studies. To this end, these measures were initially computed for all words containing specific prefixes in Bijankhan's (2004) corpus, providing three degrees of productivity for each prefix. Subsequently, fuzzy set theory, introduced by Zadeh (1965), was applied to categorize prefixes according to their degrees of productivity. Fuzzy set theory, renowned for addressing concepts with imprecise boundaries, provided an effective framework for analyzing the productivity of prefixes as well.

3.1. *Measures of Productivity Proposed by Baayen (2009)*

As aforementioned, Baayen (2009) introduced three measures of productivity, each highlighting a distinct dimension of morphological productivity:

The first measure, the 'realized productivity', also referred to as 'type frequency' is defined as the total number of different words that have been coined by means of the affix under study (Bauer, 2001; Plag, 2003; Baayen, 2009). As Baayen (2009) clarifies, this measure shows the past productivity rather than the future productivity of affixes and can only be used for comparing the productivity of different affixes with the same function. In other words, it measures the actual frequency of a given morphological process in a corpus and reflects how often a specific prefix or suffix is used to form new words within a dataset. This measure is valuable for understanding the current usage patterns and trends in a language.

The second measure, namely 'the category-conditioned degree of productivity', displays the 'potential productivity' of a category (or an affix) and is realized by dividing the number of once

occurred words containing the affix under study, 'hapaxes', by the total number of tokens created by that affix in the corpus (Baayen, 1992, pp. 115-116; Baayen, 2009, p. 7). As stated by Baayen (2009), the number of hapaxes can indicate the productivity of an affix because, as he claims, the number of hapaxes for a particular word formation pattern is related to the number of neologisms resulting from the application of that process. This ratio expresses the growth rate of the vocabulary made by a certain affix. It also remarks "the probability of encountering a word type that has not yet been observed in the corpus or text under consideration" (Baayen, 1994).

The third measure presented by Baayen (1993, 1994, 2009) is 'expanding productivity', that he terms 'the hapax-conditioned degree of productivity' (HCDP), which equals to the number of hapaxes of an affix in a given corpus divided by the total number of hapax words in that corpus. This measure shows "the rate at which a morphological category¹ is expanding and attracting new members" and ranks productive processes according to their degree of productivity (Baayen, 2009, p. 7). In other words, expanding productivity assesses the growth rate of new formations over time. In fact, it examines how frequently new word types are added to the language through a specific morphological process. This measure is crucial for understanding the long-term evolution and innovation within a language.

Generally speaking, each measure portrays a special facet of productivity. Thus comparing and associating the results of the three methods mentioned above can be rewarding for estimating productivity.

Correspondingly, in the present study the results of evaluating the productivity of Persian derivational prefixes by employing realized, potential and expanding types of productivity measures were scrutinized using fuzzy statistical techniques.

3.2. Fuzzy Set Theory

As stated by Zadeh (1965, 1972), a fuzzy set is a class of objects with vague limits and a continuum of degrees of membership. He (1965, p. 339) believes that fuzzy set theory can offer "a natural way of dealing with problems in which the source of imprecision is the absence of sharply defined criteria of class membership rather than the presence of random variables".

In this regard, a basic concept in fuzzy set theory is the concept of a *fuzzy set* (class) A in U (universe of discourse), "characterized by a membership function $\mu_A(x)$ which associates with each point x in U its 'grade of membership' in A " (Zadeh, 1972, p. 5). Therefore, the more an element x belongs to A , the higher its grade of membership. As stated by Dubois and Prade (1980), the grades

¹ Baayen (2009, p. 2) refers to "the sets of words sharing aspects of form structure and aspects of meaning as morphological categories". He states that unproductive categories are those with fixed or decreasing membership and productive categories are those categories which have increasing membership.

of membership indicate ranking of the objects in the universe, induced by the predicate associated with A ; when there is such a ranking, it is more important than the membership values themselves.

In this sense, the authors assume that the degree of productivity of derivational prefixes in Persian can be studied in terms of fuzzy sets, too; since such prefixes do not have precisely defined criteria of membership to different productivity categories. It is worth mentioning that the aim of the present work is to introduce a new method, i.e. fuzzy set theory, capable of dealing with vague concepts to the area of morphology and productivity in particular, for analyzing and representing the linguistic data in the form of derivational prefixes.

4. Methodology

4.1. The Corpus

For the purpose of investigating the productivity of Persian derivational prefixes, we first selected Bijankhan's (2004) corpus of contemporary Persian which is a tagged corpus that covers daily news and common texts; it is appropriate for natural language processing research on Persian. In this collection, all documents are classified into 4300 different subject categories such as political, cultural, economic, and social, among others. Additionally, this corpus contains morpho-syntactic and semantic annotation of approximately 2.6 million words with a tag set holding 550 Persian part-of-speech tags. These tags are served for the retrieval of varied items and for easier inspection of the tendencies of the prefixes under study in terms of word-classes involved. In this research, the subject categories of the documents are not reflected upon and the focus is mainly on the derivational prefix productivity in the whole corpus.

Furthermore, each file consists of two main columns. The first column includes the words and in-text signs in the corpus. In the second column, the labels corresponding to each word is cited. Each label indicates a specific syntactic and morphological property. In the first step, each file in the corpus was separately read and their contents were saved in a structured database. To this end, a string array of memory cells called dictionary was defined which was empty at the beginning of the study. Then each file was read line by line. In each line, the word was segregated from its labels. This word was searched by the functions *strcmp* and *find* in MATLAB to be found in the dictionary and if there was not such a word recorded, it was written in the first empty entry in the dictionary. In this way, after investigating the whole documents in all the files, a dictionary of unique words and their specific properties was composed. As for the aim of this study, a list of all derivational prefixes was defined as a string array in MATLAB. To find words with a given prefix, *strcmp* function was used. Each prefix was compared with all the cells in the dictionary of all the words in the corpus. If the beginning of the cell matched with the prefix at hand, that word was selected as a derived word in the output.

4.2. Data Analysis Procedure

With the intention of applying Baayen's (2009) productivity measures for weighing the productivity of Persian derivational prefixes, the next step was to sum up the number of types or distinct words having the relevant prefix in the corpus (type frequency) and the number of all the words beginning with that prefix in the same corpus (token frequency).

Besides, the number of hapaxes of each derivational prefix and the number of all hapaxes in the corpus needed for computing the realized, expanding and potential productivity were calculated. After calculating each productivity measure for different prefixes, the results were presented on a separate diagram for each measure to see the distribution of the items for further categorizations.

In addition, as far as the fuzzy statistical analysis was concerned, the authors recommended a quintuple bunching of prefixes on a continuum of [0,1] into (completely) unproductive (0), almost unproductive, semi-productive, almost productive and (completely) productive (1), with some prefixes being close to one neighbor and farther away from the other. This was accomplished by studying the distribution and density of the dots corresponding to each prefix in the chart for each measure of productivity. Subsequently, the density of some bunching was different from the others in each part of the chart, thus the intervals were not necessarily the same. In (1) a trapezoidal-shaped membership function $f(x; a, b, c, d)$ for each productivity categorization, in which a, b, c , and d , are parameters of the membership function, is described:

$$(1) \quad f(x; a, b, c, d) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & b \leq x \leq c \\ \frac{d-x}{d-c} & c \leq x \leq d \\ 0 & d < x \end{cases}$$

Where x stands for a real number within the universe of discourse. This function is defined by a lower limit a , an upper limit d , a lower support limit b , and an upper support limit c , where $a < b < c < d$.

In order to validate the notion of hapax legomena, we also counted the number of words occurred twice/ dilegomena and three times/ trilegomena in the corpus for each given prefix and added their corresponding frequencies to the number of hapax legomena to be applied in category-conditioned degree of productivity (CCDP) and hapax-conditioned degree of productivity (HCDP) measure. Comparing the outcomes could help us to see whether the results of CCDP and HCDP taken from hapax legomena were in line with the results taken from dilegomena and trilegomena.

Next section presents the results of applying each measure of productivity on Persian derivational prefixes under study and their fuzzy analysis.

5. Results

This section provides details about the results obtained from Bijankhan's (2004) corpus for realized, potential and expanding productivity measures of each Persian derivational prefix.

5.1. Realized Productivity (RP)

This measure examines the total number of unique words (types) produced by a given prefix, reflecting its past productivity. Higher RP values indicate prefixes that have historically created a significant number of words (Baayen, 2009). The realized productivity degree of each Persian derivational prefix estimated in the corpus is presented in Table 1 along with some examples:

Table 1

Realized Productivity of Persian Derivational Prefixes in Bijankhan's (2004) Corpus

	prefix	RP	example
1	bi-	326	bi-pærva, bi-ædæb, bi-væfa, bi-tædbir, bi-særopa
2	por-	306	por-tævan, por-karbord, por-hadese, por-kar
3	na-	200	na-fenas, na-fenava, na-tævan- na-seza
4	be-	153	be-seza - be-çga, be-ruz
5	hæm-	145	hæm-rah, hæm-zæman, hæm-saje
6	sær-	119	sær-xof, sær-zæmin, sær-bar, sær-kar
7	kæm-	110	kæm-tævan, kæm-xabi, kæm-bud
8	ba-	90	ba-ædæb, ba-xoda, ba- væfa, ba- tædbir
9	piʃ-	64	piʃ-ro, piʃ-xan, piʃ-zæmine, piʃ-qæza
10	bær-	62	bær-xord, bær-daft, bær-gaft
11	færa-	46	færa-xan, færa-su, færa-bænæf
12	baz-	45	baz-xord, baz-nevisi, baz-pæxf
13	tæk-	39	tæk-sævar, tæk-ro, tæk-nævaz
14	dær-	39	dær-xast, dær-amæd, dær-xor, dær-sæd
15	foru-	28	foru-kef, foru-maye
16	va-	23	va-konef, va-gir, va-nemud
17	la-	18	la-obali
18	pej-	14	pej-gir, pej-kar
19	bela-	14	bela-estefade, bela-tæklif
20	pæs-	10	pæs-ro, pæs-ændaz, pæs-zæmine
21	æbær-	10	æbær-ræsanā
22	para-	9	para-olæmpik
23	fær-	6	fær-sajef
24	pad-	5	pad-zæhr
25	pæsa-	4	pæsa-næhvi, pæsa-modern
26	bæs-	4	bæs-amæd
27	em-	3	em-sal, em-fæb
28	doʒ-	3	doʒ-xim
29	dof-	3	dof-var, dof-nam
30	ændær-	3	ændær-un
31	vær-	2	vær-dæst, vær-fekæst
32	pira-	2	pira-pezeʃki
33	tæra-	1	tæra-bæri
34	bol-	1	bol-hævæs
35	febh-	0	-
36	bijf-	0	-

Looking at the data in Table 1 reveals that prefixes *bi-* (without), and *por-* (abundance) exhibit the highest number of type frequency among all prefixes respectively. This indicates that these prefixes were productive in the past and have produced a lot of new words which have been used so far, as well. Prefix *bi-* designates negation and prefix *por-* specifies abundance. Examining the data shows that prefix *bi-* is added to nouns or noun phrases and it generally forms an adjective or an adverb, such as *bi-ædæb*, *bi-fæk*, *bi-pærvæ*, *bi-sæbrane*. Prefix *por-* is also added to many nouns and noun phrases and creates adjectives, such as *por-tævan*, *por-bar*, and *por-næmæk*. These findings indicate that *bi-* and *por-* have been highly productive in forming new words that remain in use.

The realized productivity (RP) values of *na-*, *be-*, and *ham-* placed them almost near the highly productive prefixes while the prefixes *vær-*, *pira-*, *tæra-*, *bol-*, *febh-*, and *bif-* have the lowest type frequency in the corpus which proves that these prefixes were almost unproductive in the past. In the next step, regarding the distribution of the prefixes on the chart, they are divided into 5 fuzzy sets with different productivity intervals according to their realized degree of productivity. Figure 1 displays the realized productivity of prefixes and their membership in 5 productivity categories i.e. unproductive (0), almost unproductive, semi-productive, almost productive and productive (1),

Figure 1

Fuzzy Representation of Realized Productivity of Persian Derivational Prefixes

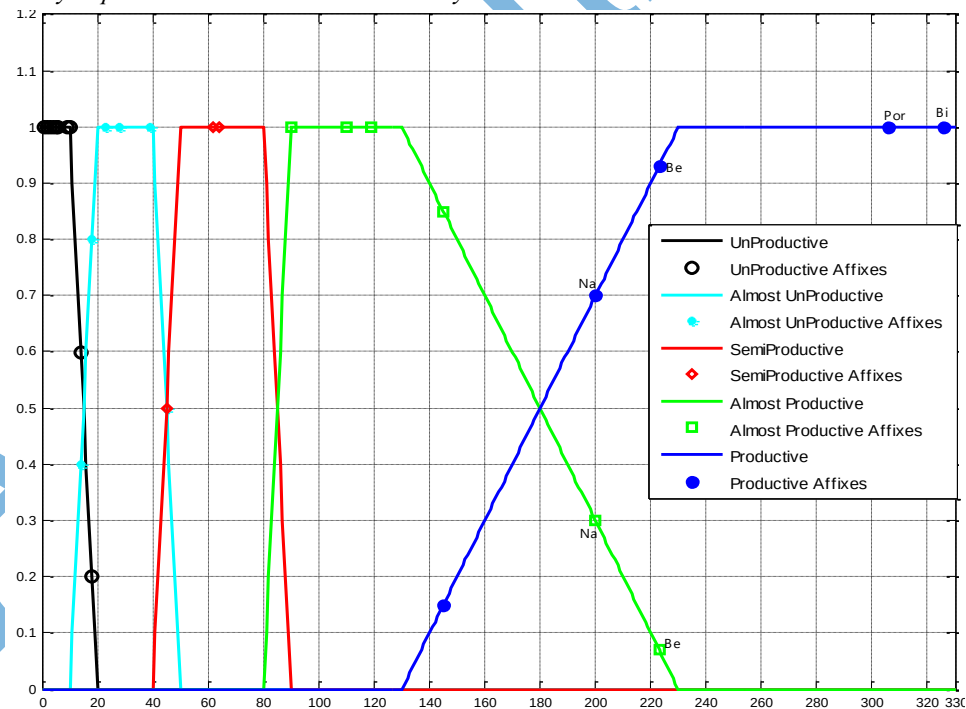


Figure 1 pictures that prefixes with realized productivity between 0 and 10, e.g. *para-*, *fær-*, *pad-*, *pæsa-*, *bæs-*, *em-*, *doz-*, *dof-*, *ændær-*, *vær-*, *pira-*, *tæra-*, *bol-*, *febh-*, *bif-* are considered

completely unproductive in the sense that they didn't produce many new words in the past. Prefixes with the realized productivity between 10 to 40 are regarded almost unproductive in the past. Moreover, prefixes *foru-*, *tæk-*, *dær-*, *va-* seem to be the prototypes of almost unproductive category by the degree of membership 1. The type frequency of the range 40 to 90 are devoted to semi-productive category in the past including prefixes such as *ba-*, *færa-*, and *baz-*. Besides, prefixes *pij-* and *bær-* are representatives of the most semi-productive cases. Prefixes with the realized productivity of more than 80 can be considered as semi-productive and almost productive since the semi-productive interval overlaps with the almost productive interval [80, 230]. Prefixes *ba-*, *kæm-*, *sær-*, and *hæm-* are typical prefixes in the almost productive category. The past productivity of prefixes *bi-*, *por-*, *na-*, *be-*, and *hæm-* with the realized productivity between 130 to 330 are the highest. The more the productivity measure moves from 130 and gets closer to 330, the less it belongs to the almost productive class and the more it belongs to the completely productive class. According to this measure of productivity, prefixes *por-*, and *bi-* with the degree of membership 1 to productive category, are the most productive prefixes among other prefixes so far. The parameters for fuzzy categorization of realized productivity of Persian prefixes are defined in Table 2:

Table 2

The Parameters (A, B, C, and D) for Each Fuzzy Category Function According to RP Measure for Persian Derivational Prefixes

fuzzy category	a	b	c	d
Unproductive (U)	0	0	10	20
Almost unproductive (AU)	10	20	40	50
Semi-productive (SP)	40	50	80	90
Almost productive (AP)	80	90	130	230
Productive (P)	130	230	350	400

According to these parameters, the fuzzy function for the semi-productive category is defined as (2):

$$(2) \quad \begin{cases} 0 & x \leq 40 \\ \frac{x-40}{50-40} & 40 \leq x \leq 50 \\ 1 & 50 \leq x \leq 80 \\ \frac{90-x}{90-80} & 80 \leq x \leq 90 \\ 0 & 90 < x \end{cases}$$

Thus, the degrees of membership of prefixes *baz-*, *færa-*, *bær-*, *pij-*, and *ba-* to the semi-productive category are 0.5, 0.6, 1, 1, and 0, respectively

5.2. Potential Productivity

Potential productivity or the category-conditioned degree of productivity (CCDP) measure which is the ratio of the number of hapax legomena to the total number of tokens formed by a given

affix in the corpus (Baayen, 2009; Plag, 1999) is evaluated for Persian derivational prefixes in Bijankhan's (2004) corpus. The results are demonstrated in Table 3:

Table 3

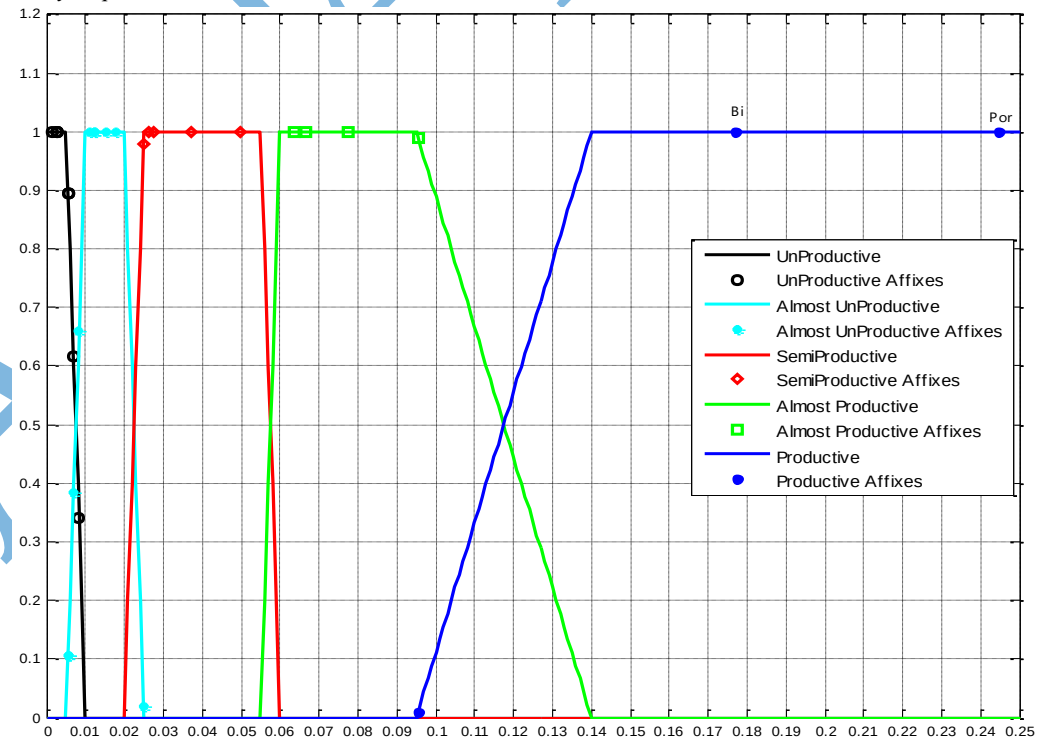
CCDP Measure of Persian Derivational Prefixes

prefix	CCDP	prefix	CCDP	prefix	CCDP	prefix	CCDP
doz-	0.5000	kæm-	0.0354	sær-	0.0062	febh-	0.0000
por-	0.2151	færa-	0.0342	baz-	0.0053	bif-	0.000
æbær-	0.1250	foru-	0.0259	hæm-	0.0045	ændær-	0.0000
tæk-	0.1226	na-	0.0241	va-	0.0040	vær-	0.0000
pæsa-	0.1111	belä-	0.0240	fær-	0.0027	em-	0.0000
ba-	0.0891	pæs-	0.0192	bær-	0.0012	dof-	0.0000
para-	0.0769	be-	0.0109	dær-	0.0011	tæra-	0.0000
la-	0.0672	pij-	0.0072	bæs-	0.0003	pira-	0.0000
bi-	0.0495	pad-	0.0063	bol-	0.0000	pej-	0.0000

Table 3 highlights that the probability of facing new forms made by prefixes *doz-*, *por-*, *æbær-*, *tæk-* and *pæsa-* is the most whereas the chance of finding new well-formed neologisms coined by *febh-*, *bif-*, *ændær-*, *vær-*, *em-*, *dof-*, *tæra-*, *pej-*, and *pira-* is the least. Figure 2 presents the fuzzy categorization of the data in Table 3:

Figure 2

Fuzzy Representation of CCDP of Persian Derivational Prefixes



In keeping with Figure 2, the intervals of [0.0.01], [0.003, 0.025], [0.02, 0.06], [0.055, 0.1] and [0.095, 1] are established for completely unproductive, almost unproductive, semi productive, almost productive and completely productive categories respectively (based on CCDP). This measure illustrates the contribution of each prefix to the growth rate of the total vocabulary. The parameters a, b, c, and d for each category are put forward in Table 4:

Table 4

The Parameters (a, b, c, and d) for Each Fuzzy Category Function According to CCDP Measure for Derivational Prefixes

Fuzzy category	a	b	c	d
Unproductive (U)	0	0	0.003	0.01
Almost unproductive (AU)	0.003	0.01	0.02	0.025
Semi-productive (SP)	0.02	0.025	0.055	0.06
Almost productive (AP)	0.055	0.06	0.095	0.14
Productive (P)	0.095	0.14	1	1

In line with Table 9, the fuzzy function for the completely productive category should be identified as equation (3):

$$(3) \quad f(x) = \begin{cases} 0 & x \leq 0.095 \\ \frac{x - 0.095}{0.1 - 0.095} & 0.095 \leq x \leq 0.14 \\ 1 & 0.14 < x < 0.3 \end{cases}$$

Consequently, the degree of membership to the completely productive category of prefixes *doz-*, *por-*, and *æbær-* is 1. Whereas the prefix *doz-* shows the highest degree of CCDP, looking at the words coined by this prefix in the corpus signifies that there are only 3 infrequent words, *doz-kam* 'sorrowful', *doz-xuyi* 'enmity', and *doz-ximan* 'enemy', in the corpus, none of them are neologisms in contemporary Persian. Thus the result for *doz-* as being productive in this sense does not seem to be acceptable in authors view.

After investigating the results of CCDP measure for Persian prefixes, in order to examine the reliability of the concept of hapax legomena being used in this measure, we also count the number of twice and three-times occurred words for each prefix in the corpus and add them to their related hapax number to be used in the CCDP measure. It is necessary to mention that the denominator in CCDP formula is considered as the total number of all hapaxes plus the number of all twice and three times occurred words. Table 5 delivers the results of CCDP concerning hapax-legomena, dilegomena and trilegomena:

Table 5*CCDP of Persian Derivational Prefixes Considering the Hapax-Legomena, Dilegomena and Trilegomena*

prefix	CCDP	prefix	CCDP	prefix	CCDP	prefix	CCDP
doz-	0.7500	bi-	0.0727	sær-	0.0103	bæs-	0.0003
ændær-	0.5000	kæm-	0.0525	hæm-	0.0083	vær-	0.000
pæsa-	0.3333	foru-	0.0479	baz-	0.0080	em-	0.0000
por-	0.3062	bela-	0.0479	pad-	0.0063	dof-	0.0000
tæk-	0.1935	færa-	0.0410	va-	0.0056	tæra-	0.0000
æbær-	0.1667	pæs-	0.0385	pej-	0.0055	pira-	0.0000
para-	0.1346	na-	0.0359	fær-	0.0027	bol-	0.0000
ba-	0.1206	be-	0.0186	dær-	0.0021	febh-	0.0000
la-	0.0896	pij-	0.0128	bær-	0.0014	bif-	0.0000

Comparing the results in Table 5 with those in Table 3 signals that the number of derivational prefixes with zero degree of CCDP decreases while there isn't a significant change in the ranking of the most productive prefixes in this measure. The prefix *pej-* is considered as completely unproductive in Table 3 while it is regarded as almost unproductive in Table 5. Looking at the data in the corpus shows that this prefix has produced many words in contemporary Persian e.g. *pej-neveft*, *pej-gærd* 'prosecution', etc. and is not completely unproductive. This is more in line with our intuition as well.

5.3. Expanding Productivity, P^*

The values of P^* , i.e. the hapax-conditioned degree of productivity (HCDP) calculated are organized in Table 6 for Persian derivational prefixes:

Table 6*HCDP of Persian Derivational Prefixes in Bijankhan's Corpus*

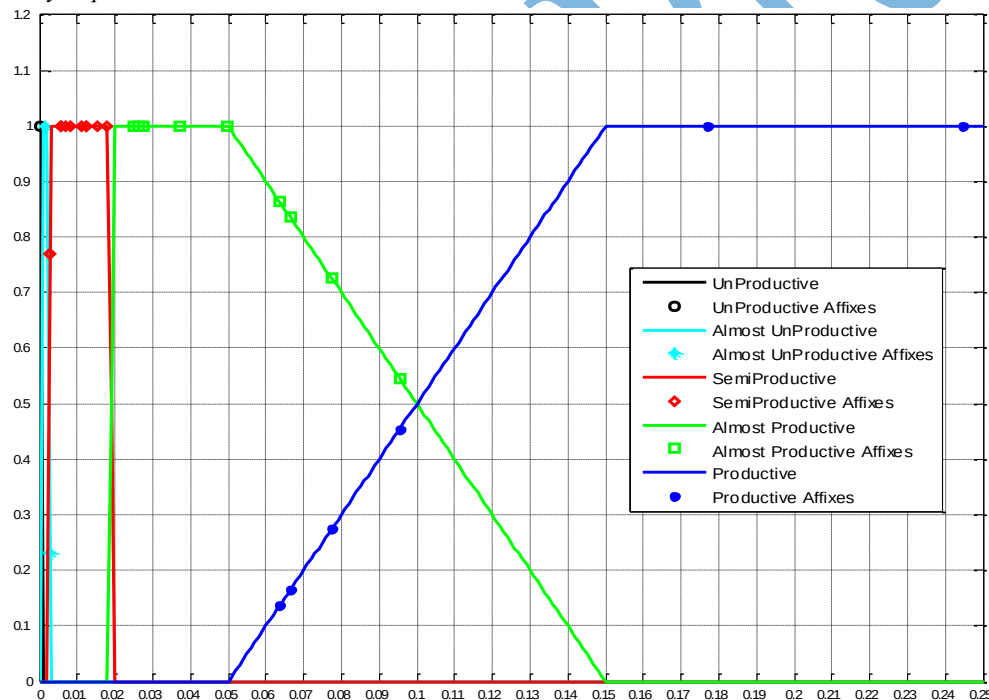
Prefix	HCDP	Prefix	HCDP	Prefix	HCDP	prefix	HCDP
por-	0.2448	tæ-	0.0263	bela-	0.0055	bif-	0.0000
bi-	0.1770	pij-	0.0249	para-	0.0028	febh-	0.0000
na-	0.0954	bær-	0.0180	pad-	0.0028	pej-	0.0000
kæm-	0.0775	baz -	0.0166	pæs-	0.0028	pira-	0.0000
ba-	0.0664	foru -	0.0124	dof-	0.0028	tæra-	0.0000
be-	0.0609	la-	0.0111	fær-	0.0014	doz-	0.0000
hæm -	0.0498	dær-	0.0083	bol-	0.0014	em-	0.0000
sær -	0.0373	æbær-	0.0069	bæs-	0.0014	vær-	0.0000
færa -	0.0277	va-	0.0055	pæsa-	0.0000	ændær-	0.0000

In accordance with Table 6, prefixes *por-*, *bi-*, *na-*, *kæm-*, *ba-*, *be-*, *hæm-*, and *sær-* have the highest numbers of HCDP, that is to say, the probability of encountering new word formations with these prefixes is higher than others. Prefixes like *bif-*, *febh-*, *pej-*, *pira-*, *tæra-*, *doʒ-*, *em-*, *vær-*, and *ændær-* have zero degree of HCDP implying that the growth rate of the vocabulary created by these prefixes is zero.

As to the fuzzy analysis of the data, the potential productivity results for each prefix are taken as inputs and the degree to which prefixes belong to each of the appropriate productivity fuzzy categories are determined via membership functions. The output denotes the degree of membership in the interval from 0 to 1. The fuzzy representation of the productivity of derivational prefixes based on HCDP is available in Figure 3:

Figure 3

Fuzzy Representation of CCDP of Persian Derivational Prefixes



According to Figure 3, the intervals $[0, 0.001]$, $[0.0005, 0.0030]$, $[0.0020, 0.020]$, $[0.0180, 0.15]$, and $[0.05, 1]$ are defined for completely unproductive, almost unproductive, semi productive, almost productive and completely productive categories respectively. It is worth mentioning that these intervals would overlap. The more the productivity measure moves from the beginning of the overlapped interval and gets closer to the end of it, the higher the degree of membership to the next category. The parameters for each fuzzy productivity category function based on HCDP measure are stated in Table 7:

Table 7

The Parameters (a, b, c, and d) for Each Fuzzy Category Function According to HCDP Measure of Persian Derivational Prefixes

Fuzzy category	a	b	c	d
unproductive	0	0	0.0005	0.001
Almost unproductive	0.0005	0.001	0.0020	0.003
Semi-productive	0.0020	0.003	0.0180	0.020
Almost productive	0.0180	0.20	0.05	0.15
productive	0.05	0.15	1	1

As per these parameters, the fuzzy function for the fuzzy set of semi-productive is as (4):

$$(4) \quad \begin{cases} 0 & x \leq 0.0020 \\ \frac{x - 0.0020}{0.003 - 0.0020} & 0.0020 \leq x \leq 0.003 \\ 1 & 0.003 \leq x \leq 0.0180 \\ \frac{0.020 - x}{0.020 - 0.0180} & 0.0180 \leq x \leq 0.020 \\ 0 & 0.020 \leq x \end{cases}$$

According to this function, the degree of membership of prefix *fær-* to the semi productive category is 0.7. Prefixes *bær-*, *baz-*, *foru-*, *la-*, *dær-*, *æbær-*, *va-*, *bela-*, and *para-* have the degree of membership 1 to the semi productive category, which means that they are the best instances of this category.

Baayen (2009) holds that neologisms are found largely among the hapaxes and to a lesser extent among words occurring twice, three times, etc. (also Baayen & Renouf, 1996; Plag, 2003). Therefore, in order to verify the notion of hapax legomena, we counted the number of words that occur twice and three times in the corpus and added their corresponding frequencies with the number of hapax legomena to see whether the results remain the same or change in the HCDP and HCDP measures. These results for HCDP are provided in Table 8:

Table 8

HCDP of Persian Derivational Prefixes Considering the Number of Dilegomena and Trilegomena

prefix	HCDP (1+2+3)	prefix	HCDP (1+2+3)	prefix	HCDP (1+2+3)	prefix	HCDP (1+2+3)
por-	0.2303	tæk-	0.0274	para-	0.0064	bæs-	0.0009
bi-	0.1718	færa -	0.0219	va-	0.0064	bol-	0.1009
na-	0.0941	baz -	0.0165	pæs-	0.0037	febh-	0.0000
kæm-	0.0759	foru -	0.0155	pej-	0.0037	pira-	0.0000
be-	0.0686	bær-	0.0137	ændær-	0.0027	tæra-	0.0000
hæm-	0.0612	dær-	0.0137	dof-	0.0027	døz-	0.0000
ba-	0.0594	la-	0.0110	pæsa-	0.0027	em-	0.0000
sær -	0.0411	bela-	0.0073	pad-	0.0018	vær-	0.0000
pij-	0.0293	æbær-	0.0073	fær-	0.0018	bij-	0.0000

Concerning Table for HCDP measure of twice and three times occurring words for Persian prefixes, the results of productive prefixes are almost the same as the results taken from once occurred words. Taking the results of unproductive prefixes into account, they were similar for both measures except for the prefix *pej-*. The HCDP (1+2+3) degree of prefix *pej-* changed noticeably compared to its HCDP measure. This prefix has devised contemporary words like *pej-ro* ‘pioneer’, *pej-vænd* ‘connection’, *pej-neveft* ‘appendix’, etc. thus, prefix *pej-* is not unproductive. It seems that the HCDP (1+2+3) measure offers closer-to-reality results. The discussion of these results will be offered in Section 6.

6. Discussion

The present study addressed the research questions concerning the realized, potential, and expanding productivity of Persian derivational prefixes and the applicability of fuzzy set theory to their classification. To this end, Baayen’s (1992, 2009) productivity measures were calculated for all derivational prefixes extracted from the Bijankhan’s (2004) corpus, and the resulting productivity values were subsequently modeled through fuzzy membership functions. This approach enabled a multidimensional analysis of productivity and provided a noble representation of the gradual differences among Persian prefixes.

With respect to realized productivity (RP), the prefixes *bi-* and *por-* emerged as the most productive derivational prefixes in the corpus, with realized productivity values of 326 and 306, respectively. Both prefixes exhibit a high degree of semantic transparency and combine with a broad range of nominal bases. The prefix *bi-* expresses negation, absence, or deprivation and productively forms adjectives and adverbs such as *bi kar* ‘unemployed’, *bi fomar* ‘countless’, *bi xab* ‘sleepless’, *bi næzi:r* ‘unparalleled’, and *bi pærva* ‘fearless’. Similarly, *por-* denotes abundance or intensity and readily forms adjectives such as *por karbord* ‘widely used’, *por tælaf* ‘hard-working’, *por dæramæd* ‘high-income’, *por tævan* ‘powerful’, and *por xanænde* ‘best-selling’. The high realized productivity of these prefixes is consistent with the observation that semantically transparent and morphologically regular affixes tend to remain productive because speakers can easily interpret and extend them to novel lexical formations (Baayen, 2009; Bauer, 2001). Similar observations have been reported in studies of Persian word formation, where *bi-* and *por-* are regularly identified as highly productive resources for lexical expansion due to their broad combinatorial potential and low semantic restrictions.

The analysis of potential productivity yielded a more complex picture. In principle, derivational affixes with high potential productivity are expected to have a greater probability of generating new lexical items, whereas derivational affixes with low values are approaching saturation and contribute less to vocabulary growth. However, the results demonstrate that statistical measures of potential productivity should be interpreted with caution. Although *doz-* obtained the highest CCDP value, a closer examination of the corpus reveals that only three

derivatives are attested: *dozkam* 'sorrowful', *dozxuji* 'enmity', and *dozximan* 'enemy'. All three forms are infrequent, highly lexicalized, and absent from contemporary neologistic processes. Consequently, despite its high numerical score, *doz-* cannot reasonably be considered a productive prefix in present-day Persian.

This discrepancy highlights a well-known methodological issue in productivity research. As Baayen (1992, 2009) emphasizes, productivity cannot be inferred solely from quantitative indices; it must also be evaluated in relation to an affix's capacity to generate new lexical items. Likewise, Plag (2003) argues that productive affixes typically occur across a diverse range of bases and remain available for the creation of novel formations. From this perspective, the apparent productivity of *doz-* is likely a statistical artifact resulting from the small number of derivatives associated with the prefix. The present findings, therefore, suggest that quantitative productivity measures should always be interpreted alongside qualitative linguistic evidence.

The results of expanding productivity (HCDP) provide a more reliable picture of the prefixes that actively contribute to lexical growth in contemporary Persian. The prefixes *por-*, *bi-*, *na-*, *kæm-*, *ba-*, *be-*, *hæm-*, and *sær-* achieved the highest HCDP values, indicating a greater likelihood of participating in ongoing word formation. These prefixes are characterized by broad semantic applicability and relatively weak selectional restrictions, properties that have repeatedly been associated with high morphological productivity. In contrast, prefixes such as *bif-*, *sebh-*, *pej-*, *pira-*, *tæra-*, *doz-*, *em-*, *vær-*, and *ændær-* exhibited negligible or zero HCDP values. These prefixes tend to occur in a limited set of lexicalized forms and are subject to stronger semantic and lexical constraints, factors known to reduce productivity.

A comparison between HCDP values based exclusively on hapax legomena and those calculated from words occurring one, two, and three times [HCDP (1+2+3)] revealed broadly similar productivity rankings. An important exception, however, was the prefix *pej-*. While its hapax-based score suggested a lack of productivity, the HCDP (1+2+3) measure assigned it a noticeably higher value. This outcome appears more consistent with actual usage patterns, as *pej-* participates in several relatively transparent word formations, including *pejro* 'pioneer', *pejvænd* 'connection', and *pejneve/t* 'appendix'. The discrepancy advocates that hapax-based measures may underestimate the productivity of low-frequency but still active affixes. Consequently, the HCDP (1+2+3) measure appears to provide a more realistic representation of productivity for Persian derivational prefixes.

Another notable finding is the strong correspondence between realized productivity and expanding productivity. Prefixes that ranked highly in RP generally occupied the highest positions in HCDP as well. This convergence suggests that prefixes with a large inventory of established derivatives also retain a significant capacity for future lexical expansion. Particularly notable are *bi-*, *por-*, *na-*, and *hæm-*, which consistently ranked among the most productive prefixes across

multiple measures. Their behavior supports the view that productivity is closely linked to semantic transparency, frequency, and combinatorial flexibility.

The fuzzy analysis further demonstrated that productivity is best conceptualized as a gradient rather than categorical phenomenon. Instead of imposing rigid boundaries between productive and unproductive affixes, the fuzzy model classified prefixes into five overlapping categories ranging from productive to unproductive. Membership in these categories was determined by the distribution of productivity values and the distances between adjacent prefixes along the productivity continuum. Consequently, some derivational prefixes exhibited partial membership in more than one category simultaneously. This finding is consistent with theoretical accounts that regard morphological productivity as a matter of degree rather than an all-or-nothing property (Bauer, 2001; Hay & Baayen, 2005).

Methodologically, the study also highlights the usefulness of fuzzy set theory for modeling uncertainty in morphological analysis. Unlike traditional approaches, which require discrete classifications, fuzzy modeling captures the inherently gradual nature of lexical innovation and affix productivity. The results therefore support the view that productivity constitutes a multidimensional continuum shaped by frequency, semantic transparency, lexical restrictions, and the availability of an affix for novel word formation.

Furthermore, when compared with the findings of previous studies, several distinctions and alignments emerge, highlighting the contributions and implications of the present research.

First, while prior studies such as Abbasi (2016) and Abedini et al. (2019) relied on single productivity measures, this study employed all three of Baayen's productivity measures—realized productivity (RP), potential productivity (HCDP), and expanding productivity (CCDP). This comprehensive integration enables a broader view of prefix productivity. Notably, the study confirms that prefixes *bi-*, *por-*, and *na-* are among the most productive across both RP and HCDP measures. However, discrepancies arise with the CCDP measure, which categorizes the rarely used prefix *doz-* as productive due to its low token count in the corpus. This inconsistency underscores the limitations of CCDP and aligns with Abbasi's (2016) observation regarding the need for critical assessment of individual measures.

Additionally, this research builds on earlier observations of scalar productivity (e.g., Faragordi, 2010; Abedini et al., 2019) by providing a systematic methodology to quantify this scalar nature. Through fuzzy set theory, prefixes are assigned degrees of membership to overlapping categories (e.g., productive, semi-productive, unproductive), represented visually in fuzzy diagrams. This approach refines prior categorizations, such as those by Faragordi (2010), who relied on frequency-based thresholds without a clear justification. The fuzzy framework enables a more dynamic and flexible classification, particularly for prefixes like *tæra-* and *doz-*, which exhibit borderline productivity levels.

The study also revisits the concept of hapax legomena, revealing its limitations as a sole indicator of neologisms. While prior research, such as Gaeta and Ricca (2003), acknowledged

inconsistencies in hapax-based measures, their solutions emphasized corpus size normalization rather than addressing conceptual ambiguities. This study demonstrates that archaic and rarely used terms (e.g., *Dezhkhaiman*) skew hapax-based calculations, necessitating manual evaluation in ambiguous cases. This critical reevaluation contributes to the ongoing debate about the reliability of hapax legomena as a proxy for innovation in word formation.

Besides, while Aftabi *et.al.* (2021) applied fuzzy set theory to analyze Persian derivational suffix productivity, the present study extends this framework to derivational prefixes, introducing parameters (a, b, c, and d) to define productivity categories and measure transitions between them. This methodological innovation provides a clearer representation of productivity gradience, revealing significant density shifts within specific categories.

Finally, it is worth noting that the prefixes identified as productive in this study (e.g., *bi-*, *por-*) align with Abedini et al.'s (2019) findings, despite differences in methodology. However, Aftabi *et.al.* challenges the classification of *bol-* as a prefix, arguing that it no longer functions as such in contemporary Persian.

Overall, the findings demonstrate that Persian derivational prefixes differ substantially in their productivity profiles and that these differences are more accurately represented through a fuzzy framework than through categorical classifications. Future research may extend this approach to other word-formation processes, comparing productivity across registers and discourse types, and investigate the interaction of productivity with factors such as lexical blocking, entrenchment, and semantic specialization. Such investigations would contribute to a more comprehensive understanding of morphological productivity and lexical change in Persian.

7. Conclusion

This paper studied the fuzzy nature of productivity of Persian derivational prefixes using the data in Bijankhan's (2004) corpus based on realized, potential, and expanding productivity measures proposed by Baayen (2009) and fuzzy set theory of Zadeh (1965, 1972). Our analysis disclosed that classifying the productivity measures calculated for each prefix into five overlapping categories of productivity enabled us to give each prefix a degree of membership to various productivity categories and proved the feasibility of using fuzzy set theory for investigating morphological productivity. This opens the door to develop fuzzy analysis for other linguistic and morphological vague concepts. The results found so far are to be taken as insights of a preliminary exploration of the notion of productivity. Nevertheless, several avenues remain open for future research.

First, the analysis was restricted to derivational prefixes extracted from the Bijankhan corpus. Future studies may extend the proposed framework to other morphological processes in Persian, e.g., compound formation, thereby providing a more comprehensive account of Persian word-formation patterns.

Second, the findings are corpus-dependent and reflect the linguistic characteristics of the Bijankhan's (2004) corpus. Applying the same methodology to larger and more diverse corpora, including contemporary web-based corpora, spoken corpora, and domain-specific corpora, would allow researchers to examine the stability and variability of productivity patterns across different registers and genres. Such investigations may reveal whether the productivity rankings observed in the present study are consistent across linguistic contexts.

Third, future research may incorporate diachronic corpora to investigate changes in the productivity of derivational prefixes over time. A historical perspective would provide valuable insights into the evolution of Persian morphology and the emergence, expansion, or decline of specific derivational processes.

Fourth, cross-linguistic studies could employ the same fuzzy analytical framework to compare the productivity of derivational affixes across genetically related and unrelated languages. Such comparisons would help determine whether the productivity patterns identified in Persian reflect language-specific characteristics or broader typological tendencies in morphological systems.

Finally, although the present study utilized the three productivity measures proposed by Baayen (2009), future research may explore additional linguistic variables, such as semantic transparency, lexical frequency distributions, register variation, and psycholinguistic factors affecting affix usage. Furthermore, alternative fuzzy inference systems and hybrid computational approaches combining fuzzy logic with machine learning techniques may enhance the precision and explanatory power of productivity assessment models.

Generally, the integration of corpus linguistics, morphological productivity research, and fuzzy logic offers a promising interdisciplinary framework that can be further developed to achieve a better understanding of morphological productivity in Persian and other languages.

Conflict of interest: The authors state no conflict of interest.

References

- Abbasi, A. (2005). Mahdudiyat haye sarfi va nahvi dar zayayie farayande eshteqaq dar zabane farsi [Morphological and syntactic restrictions on the productivity of derivation in Persian]. *Zaban va Zaban Shenasi*, 2(4) [Language and linguistics]. 39-56 (in Persian).
- Abedini, H., Moloodi A. S., & Khormaei A. (2019). A Corpus-based Study of Derivational Prefixes in the Written Variety of Contemporary Persian. *Journal of Iranian Language Linguistics*, (2), 243-265. <https://doi.org/10.22099/jill.2020.35234.1170>
- Aftabi, S. Z., Ahangar, A. A., & Mishmast Nehi, H. (2021). Derivational Suffix Productivity in Persian: A Fuzzy Analysis. *Journal of Quantitative Linguistics*, 29(3), 387-411. <https://doi.org/10.1080/09296174.2021.1887575>
- Anshen, F., & Aronoff, M. (1988). Producing morphologically complex words. *Linguistics*, (26), 641-655. <https://doi.org/10.1515/ling.1988.26.4.641>
- Aronoff, M. (1976). *Word Formation in Generative Grammar*. MIT Press.
- Baayen, R. H. (1992). Quantitative aspects of morphological productivity. In G. E. Booij & J. van Marle (Eds.), *Yearbook of morphology* (pp.109-149). Kluwer Academic Publishers.
- Baayen, R. H. (1993). On frequency, transparency, and productivity. In G. E. Booij & J. van Marle (Eds.), *Yearbook of Morphology*. Kluwer Academic Publishers, 181-208. https://doi.org/10.1007/978-94-017-3710-4_7
- Baayen, R. H. (1994). Derivational productivity and text typology. *Journal of Quantitative Linguistics*, 1, 16-34. <https://doi.org/10.1080/09296179408589996>
- Baayen, R. H. (2009). Corpus linguistics in morphology: MCorphological productivity. In A. Lüdeling, & M. Kyto, (Eds.), *Corpus linguistics. An international handbook* (pp. 900-919). Mouton De Gruyter.
- Baayen, R. H., & Lieber, R. (1991). Productivity and English derivation: A corpus-based study, *Linguistics*, 29(5), 801-844. <https://doi.org/10.1515/ling.1991.29.5.801>
- Baayen, R. H., & Renouf, A. (1996). Chronicling The Times: Productive lexical innovations in an English newspaper. *Language*, 72, 69-96. <https://doi.org/10.2307/416794>
- Badakhshan, E. (2010). *A Corpus-based study of the productivity of compounding in contemporary Persian*. [Doctoral dissertation, Allameh Tabataba'i University].
- Bauer, L. (2001). *Morphological Productivity*. Cambridge University Press.
- Bijankhan, M. (2004). The role of the corpus in writing a grammar: An introduction to a software. *Iranian Journal of Linguistics*, 19(2), 48-67.
- Dubois, D., & Prade, H. (2000). *The Handbooks of Fuzzy Sets: Fundamentals of Fuzzy Sets*, 7. Springer, US.
- Faragordi, R. (2010). *Persian Prefixes*. [Master thesis, Allameh Tabataba'i University].
- Gaeta, L., & Ricca, D. (2003). Frequency and productivity in Italian derivation: A comparison between corpus-based and lexicographical data. *Italian Journal of Linguistics/Rivista di Linguistica*, 15(1), 63-98.

- Joyce, J. (1976). Fuzzy Sets and the Study of Linguistics. *Pacific Coast Philology*, 11, 39-42. Penn State University Press. <https://doi.org/10.2307/1316746>
- Kamp, H., & Partee, B. (1995). Prototype theory and compositionality, *Cognition*, 57, 129-191.
- Labov, W. (1973). The boundaries of words and their meanings. In Charles-James Bailey & Roger W. Shuy (Eds.), *New ways of analyzing variations in English 1* (pp. 340-371). Georgetown University Press.
- Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 2, 458-508. <https://doi.org/10.1007/BF00262952>
- Lasersohn, P. (1999). Pragmatic Halos. *Language*, 75(3), 522-551. <https://doi.org/10.2307/417059>
- Osherson, D., & Smith, E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9, 35-58. [https://doi.org/10.1016/0010-0277\(81\)90013-5](https://doi.org/10.1016/0010-0277(81)90013-5)
- Osherson, D., & Smith, E. (1982). Gradedness and conceptual conjunction. *Cognition*, 12, 299-318.
- Plag, I. (1999). *Morphological productivity: Structural constraints in English derivation*. Mouton de Gruyter.
- Plag, I. (2003). *Word-formation in English*. Cambridge University Press
- Rosch, E. (1977). Human categorization. In N. Warren (Ed.), *Advances in cross-cultural psychology* (Vol. 1). Academic Press.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8, 338-353. <https://doi.org/10.2307/2272014>
- Zadeh, L. A. (1972). A Fuzzy set-theoretic interpretation of linguistic hedges. *Journal of Cybernetics*, 2(3), 4-34. <http://dx.doi.org/10.1080/01969727208542910>

Corpora

- Bijankhan corpus. (2012). University of Tehran. <http://www.peykaregan.ir/dataset/> پیکره-بی جن خان