Morphological Productivity of Non-verbal Persian Derivational Suffixes in "Dependency Corpus"

Anahid Dashti[1], Vida Shaghaghi[2]

1. INTRODUCTION

Productivity covers a wide range of linguistic phenomena. Productivity is as much about syntactic phenomena; it can be related to morphological phenomena. The various definitions of productivity are reflected in the presentation of the approach and the method of productivity measurement. Studies on what and how to measure morphological productivity have been conducted since 1899.

It can be said, in 1976, Aronoff proposed the first method for quantitative measuring of morphological productivity in English. Baayen (2009), in productivity measurement, considered the possibility of finding new words made by derivational suffixes. Thus, he has developed quantitative morphological productivity measurement methods. Anshen and Aronoff (1998), describe productivity of how new words are made with a certain suffix. With this view, the productivity rate of each suffix is obtained. It means productivity becomes a probable continuum and it can be used to predict the rate of possible word construction. At both ends of this continuum, there are patterns that are "full-productive" and "non-productive".

Of course, Palg (2006), also discusses the use of suffixes as overall "possibility" of language for constructing new words. The logical consequence of all morphological productivity analyzes is the presentation of various methods for quantitative measuring of the productivity of suffixes. Most studies on this field focus on the measurement and reliability and validity of each measurement. Despite more than thirty years of quantitative studies, Bauer (2004) argues that one of the empirical problems with these measurements is that "we do not yet have a reliable measure for productivity". In Iran, Abbasi (2005), Rafiei (2008), Badakhshan (2010), Hemmasian (2010), Kheirabadi et al. (2010), Erfanian Qonsuli (2011), Mavvaji (2012), Amirarjmandi et al. (2013) and Farzaneh (2016) have studied on morphological productivity in Persian.

The main purpose of this study is to quantify the current productivity of non-verbal Persian derivational suffixes using different methods so that the probable productivity continuum of these suffixes is given. This article also introduced and

---

[1] Ph.D. candidate in General Linguistics, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran. (second Author); anahid_dashti@atu.ac.ir

[2] Ph.D. in General Linguistics, Professor at Department of Linguistics, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran. (corresponding author); shaghaghi@atu.ac.ir

analyzed different methods for quantitative measurement of productivity in researchers' studies over the last thirty years. In this study, morphological productivity is assessed using a corpus-based approach. Additionally, according to Bauer (2004) and Baayen (2009) quantitative measurement framework, four frequency methods and two probability estimation methods have been used to measure morphological productivity.

2. MATERIALS AND METHODS
The used corpus was the "dependency corpus". This corpus has about half a million words and was published in 2013. This study is based on 50 suffixes introduced by Sadeghi (1991-1993). First, the data entered the Excel spreadsheet. Then all the participating suffixes in the construction of the words were extracted. In the next step, the segmentation and finding the root of all words based on Persian dictionaries was done manually. Given the large number of suffixes, their mean frequency were categorized by "Clustering Analysis". the K-mean algorithm was used to categorize the morphological productivity value by the frequency of types, tokens, hapax legomena, and neologisms. Data were categorized into 5 clusters by using R software version 4.0.3.

3. RESULTS AND DISCUSSION
Findings show; calculating the frequency of types in this corpus does not provide a suitable criterion for determining the current productivity of suffixes. Because morphological productivity means the ability to create or emerge new words that did not exist before the present time. Baayen (2009), using mathematical formulas, has designed this method according to the size of the morphological category. Thus, measurement of a category that has many members and many complex words, is more productive.

According to the findings, in the measurement of morphological productivity with two methods of type and token frequency, we encountered several problems: first, in the high-productive cluster, once the suffixes "-æn" and "-e", and once the suffix "-e" are placed. The second problem, although the productivity rate of the suffixes is somewhat clear, neither a fixed value nor the productivity of the suffixes can be comparable. The third problem is that this measurement is only a result of past productivity and the extent of the use of realized words in the language. In other words, measurements of the frequency of types and tokens do not provide an indication of present productivity.

Morphological productivity is measured in two main methods according to hapax legomenon. One of these methods is called "productivity in the narrow sense". In this method, the number of hapax legomena of each given category is divided by the total frequency of its tokens in the corpus. We encountered other problems using this measurement; first, in this method, tokens of each suffix that are placed in the denominator of the fraction are variable. Therefore, the corresponding value in the measurement for each suffix in the corpora with different volumes changes frequently. Second, the suffix "-zar" in this measurement had full productivity while this was due to the similarity of the frequency of its type, token and hapax legomenon (ie with the frequency of one).

Thus, this type of measurement in small corpora gives misleading results because usually the frequency of type, token, and hapax legomenon are low. Third, the sum of the productivity values of the five clusters is greater than one. Hence, we conclude that a concept other than the probability of productivity has been measured whereas the probability of productivity must be values between the numbers zero and one. The last method of measuring morphological productivity in terms of

hapax legomenon is the "global productivity" method on which we tested the data. In this method, the frequency of hapax legomenon of each given category is divided by the set of hapax legomena of the Corpus.

4. CONCLUSION

The results of this method showed that in the "dependency corpus", the suffix "-i" has the highest "global productivity" with a probability of productivity (0.47), which is placed alone and with a very large difference compared to other suffixes on the right side of the continuum with the highest "global productivity". Then the two suffixes "-æn" and "-e", with a probability of productivity (0.10), are placed in the high-productive cluster. 10 suffixes "-an, -ane, -dar, -eʃ, -æk, -kar, -gan, -ænde, -jæt -ijje", are placed in the average-productive cluster with a probability of productivity (0.019). Twelve suffixes "-a, -ban, -bændi, -dan/dani, -estan, -kari, -gær, -gah, -gar, -mænd, -var, -in", with a probability of productivity (0.007) placed in the low-productive cluster. Twenty-five other suffixes "abad, ar, asa, -u, -bazi, -ʤe, - ʤi, -fam, -san, -æki -ænd, -kæde, -gane, -gæri, -gun, -gin, -man, -nak, -vare, -vær, -væʃ, -vænd, -jar, -ine, -zar", with very low probability of productivity (0.0017), were placed on the left side of the continuum in the non-productive cluster.

First, in this method, to measure the productivity of each suffix, all hapax legomena are considered as the denominator. Therefore, the corresponding value in the measurement of productivity for all suffixes of a corpus of any volume is quite constant. Second, the sum of the productivity values of the five clusters is equal to one. In other words, the productivity values of each suffix in the morphological productivity continuum have a probability between zero and one. Thus, what has been tested by this method has measured the probability of current morphological productivity. Therefore, this method has proven its reliability and corresponds to the linguistic intuition.