




Productivity of Derivational Affixes in Surani Kurdish Language: Corpus-based Approach

Seyedeh Sareh Sadeghi^{1*} 
Najmoldin Birakhasi² 
Ebrahim Badakhshan³ 

Abstract

The present research has investigated the production of derivational affixes in the Surani Kurdish language, based on the corpus of “Kurdpress”. In morphology, the productivity of a suffix was expressed as the number of new words created by that suffix. In this way, the present research has investigated the frequency of the derivational affixes of the Kurdish language in the final position of the word, based on a corpus approach. In the first step of this research, a list of Kurdish suffixes was identified. Then, by using web crawling programming in Python language, the required data was prepared from the Kurdpress news network. The measuring criterion used for the productivity of each of these suffixes is based on the criterion of the p value. Finally, it was determined that the words “چی، دان، ین، زن، یلانه” were among the most productive suffixes in the Kurdish language with a productivity rate of more than 50%. Among the other things discussed in this study is the difference between the concept of productivity and the importance that an affix assigns to itself. The importance of an affix is known in the number of words that are used in the vocabulary list of the language.

Keywords: derivational affixes, productivity, corpus-approach, Kurdish language, morphology.

Extended abstract

1. Introduction

The use of affixes in a language varies in terms of their productivity in word formation. Some affixes tend to generate more new words than others and can

*1. Phd in linguistics, Department of linguistics, Faculty of Language and Humanities, Bu-Ali Sina University, Hamedan, Iran. (**Corresponding Author: S.S.Sadeghi94@basu.ac.ir,**)

2. Master's student in linguistics, Department of English Language and Linguistics, Faculty of Language and Literature, University of Kurdistan, Sanandaj, Iran. (necmebiraxasi@gmail.com)

3. Associate professor in general linguistics, Department of English Language and Linguistics, Faculty of Language and Literature, University of Kurdistan, Sanandaj, Iran.(ebadakhshan@uok.ac.ir)

be considered more active in the process. However, their productivity rates may change over time, resulting in shifts in their usage. This study is distinctive in that it delves into the investigation of Kurdish language derivational affixes in a rigorous manner, exploring their productivity rates based on a considerable amount of data. Additionally, the study examines the significance of existing variables in relation to the productivity rates of these affixes. The primary objective of the study is to identify which of the Kurdish language's derivational affixes is the most productive. The study aims to fill a gap in the literature by conducting a detailed analysis of this aspect of the Kurdish language.

2. Theoretical framework

As is typically the case, studies conducted in the realm of morphology are concerned with topics such as affixation, location and types of affixes, affix productivity, affix importance, and changes over time in affixes. The two common types of affixes observed in most languages are prefix and suffix affixes. A prefix is a dependent morpheme that is attached to the beginning of a base. A suffix is a dependent morpheme that is attached to the end of a base. Among the affixes used in a language, some are generally more productive in word formation and create a greater number of new words, while others have less productivity. Even the productivity of affixes may change over time, with some affixes being more productive at certain times and transformed into less productive affixes at other times. Affix productivity is defined as the number of new words created by that affix in word formation, and the importance of an affix is measured by the number of words that are created on a diachronic basis based on that affix. The frequency of a unit, along with all its repetitive forms, occurring multiple times in the text is referred to as its token frequency. The number of occurrences of a unit under study in the corpus, disregarding repetition, is referred to as its type frequency. Affix productivity is considered as a value obtained by dividing the number of single occurrences of affixes known as hapax legomena by the total number of tokens in the corpus. The formula for affix importance and productivity is presented below.

$$\text{Hapax Token Rate} = \text{HTR} = \frac{n(\text{hapax legomena})}{n(\text{token})}$$

$$\text{Type Token Rate} = \text{TTR} = \frac{n(\text{type})}{n(\text{token})}$$

3. Methodology

For corpus-based studies in Kurdish linguistics, the most important issue is the availability of suitable data for analysis. Therefore, the first step was to collect sufficient data for this research, using the Kurdish language corpus and web crawling techniques. The initial version of this corpus consisted of 69,000 news documents, containing various news items from different categories, which

were collected using a web crawler program written in Python version 3.4 that focused on news sources.

In the second step, a specific list of derivational suffixes in Kurdish was determined using Kurdish grammar books, such as books on grammar and a book that explains the structure of the Sorani dialect of the Kurdish language. Since the frequency of single-word counts is commonly used for examining the productivity of suffixes, after collecting the data and the list of derivational suffixes in Kurdish, the frequency of all words was calculated. This calculation was performed using Python scripts. Then, words with a frequency of one were identified, and it was determined how many words are formed by each of these suffixes. Suffixes that produced fewer than five single-word forms were removed to obtain more accurate results.

4. Result and discussion

After identifying the derivational suffixes, the frequency of different affix types, markers, and single-frequency words for each morpheme at the end of the examined word were considered. The results, along with the values of productivity and importance of the morphemes, are displayed in a table within the article. Productivity is generally considered as a spectrum in which less productive units are on the left side and more productive units are on the right side. Another noteworthy point in this study is that in studies of word structure, paying attention to the frequency of examined units alone cannot show accurate results. This is because various factors may influence the frequency of words, and therefore focusing solely on word frequency is not enough.

5. Conclusion and Suggestions

This study discusses the central concept of productivity. It was found that the suffixes “chi”, “dan”, “yan”, “zan”, and “ylaneh” make up over 50% of the most productive suffixes. The average productivity rate for suffixes in Kurdish language is 37.25%. Attention to the prefixes in Kurdish language, the order of affixes, and the hierarchy that plays a role in the productivity rate and importance of affixes are among the topics that can be explored in future research to complement this study.

Select Bibliography

- Kohanzad, P., Fallahi, M., Pahlevanzadeh B. A Corpus-based Study of the Productivity of Derivational Affixes in Persian. *Journal of Researches in Linguistics*. 2021; 2 (23): 219-240. [in Persian]
- Badakhshan E. Kurdish corpus project. International Institute for the Study of Kurdish Societies First Biennial Conference Germany, Frankfurt, 2017; 16-19.
- Gaeta L. Ricca D. Productivity in Italian word formation: A variable-corpus approach, Berlin: De Gruyter Mouton. 2006; 44(1): 57-89.

- Montero-Fleta, B. Suffixes in word-formation processes in scientific English. LSP Journal-Language for special purposes, professional communication, knowledge management and cognition. 2011; 2(2): 4-14.
- Motsch W. On inactivity, productivity and analogy in derivational processes. In the Contribution of Word-Structure-Theories to the Study of Word Formation. 2018; 1-30.
- Stefanowitsch A. Corpus linguistics: A guide to the methodology. Berlin: Language Science Press; 2020. DOI: 10.5281/zenodo.3735822
- Ten H., P. Productivity and Anticipation in Language Processing. SKASE Journal of Theoretical Linguistics. 2020; 17(4): 23-36.

How to cite:

Sadeghi, Seyedeh Sareh., Birakhasi, Najmoldin. Badakhshan, Ebrahim. Productivity of Derivational Affixes in Surani Kurdish Language: Corpus-based Approach. *Zaban Farsi va Guyeshhay Irani*, 2023; 2(14): 35-51. DOI:10.22124/plid.2023.23607.1622

Copyright:

Copyright for this article is retained by the author(s), with first publication rights granted to *Zaban Farsi va Guyeshhay Irani (Persian Language and Iranian Dialects)*. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original work is properly cited.



میزان زایایی وندهای اشتقاقی در زبان کردی سورانی، رویکرد پیکره بنیان

ابراهیم بدخشان^۱

نجم‌الدین براخاصی^۲

سیده ساره صادقی^۳

چکیده

پژوهش حاضر با استفاده از پیکره کردپرس^(۱) به بررسی زایایی وندها در زبان کردی سورانی پرداخته است. در علم صرف، میزان زایایی یک وند را به میزان تعداد کلمات جدیدی بیان کردند که یک وند ساخته است. بدین ترتیب به بررسی میزان زایایی پسوند به صورت پیکره بنیاد پرداخته ایم. در گام نخست فهرستی از پسوندهای زبان کردی مشخص گردید. سپس با استفاده از برنامه نویسی خزش وب به زبان پایتون از وبسایت شبکه خبری کردپرس داده تهیه گردید. معیار سنجش برای زایایی هر یک از این وندها معیار باین مقدار p است. در نهایت مشخص گردید که وندهای «چی، -دان، -ین، -زن، -یلان» از زیاترین وندهای زبان کردی با مقدار زایای بالاتر از ۵۰ درصد بوده است. از دیگر موارد بحث شده در حوزه تفاوت مفهوم زایایی و میزان اهمیتی است که یک وند دارد. میزان اهمیت یک وند را در تعداد کلماتی می دانند که در فهرست واژگانی برای زبان به کار برده می شود و در علم صرف این دو مفهوم از یکدیگر متمایز شناخته می شوند.

واژگان کلیدی: زایایی، رویکرد پیکره بنیان، زبان کردی، وند.

✉ s.s.sadeghi94@basu.ac.ir

۱. دانش آموخته دکتری زبان شناسی همگانی، دانشکده زبان و ادبیات، دانشگاه بوعلی سینا، همدان، ایران. (نویسنده مسؤل)

necmebiraxasi@gmail.com

۲. دانشجوی کارشناسی ارشد زبان شناسی همگانی، دانشکده زبان و ادبیات، دانشگاه کردستان، سنندج، ایران.

ebadakshah@uok.ac.ir

۳. دانشیار گروه زبان انگلیسی و زبان شناسی، دانشکده زبان و ادبیات، دانشگاه کردستان، سنندج، ایران.

۱- مقدمه

آناتول استفانویچ در کتاب خود شرح کاملی از تعریف زبان‌شناسی پیکره‌ای ارائه می‌دهد که در آن تمامی عناصر سازنده تعریف زبان‌شناسی آزمایش و بررسی دقیق می‌شوند (Stefanovich, 2020:61). وی زبان‌شناسی پیکره‌ای را بررسی پدیده زبانی می‌داند که در توزیع تعاملی با پدیده‌ای دیگر در پیکره زبانی سنجیده می‌شود. به‌طور نمونه می‌توان به بررسی کلماتی که برحسب شیوه ساختارشان برحسب‌گذاری شده‌اند و مطالعه می‌شوند اشاره کرد. ابزاری که برای مطالعه در این حوزه زبان‌شناسی پیکره‌ای استفاده می‌شود رایانه است تا با کمک آن بتوان حجم زیادی از داده را مطالعه و ساماندهی کرد و زمان جمع‌آوری و مطالعه را کاهش داد. حجم بسامد از داده که امروزه در اینترنت و صفحات و شبکه‌های اجتماعی فراهم شده، سبب استقبال زبان‌شناسان به این رویکرد زبانی شده است. بدین جهت این پژوهش با رویکردی پیکره‌بنیاد به مطالعه میزان زایایی وندهای کردی گویش سورانی پرداخته است.

وندها در بیشتر زبان‌ها، به‌ویژه زبان‌های هندواروپایی، نقش عمده و مهمی در زایایی، خلاقیت و واژه‌سازی یا تصریف واژه‌ها ایفا می‌کنند. برخی وندها در واژه‌سازی زایاترند. در حقیقت زایایی به‌صورت طیفی از کم به زیاد نمایش داده می‌شود. همچنین زایایی وندها در طول زمان ثابت نیست و برحسب عوامل گوناگونی قابل تغییر است (Bauer, 2004). بدین ترتیب، در مجموع آنچه این پژوهش را از دیگر پژوهش‌ها متمایز می‌سازد آن است که برای نخستین بار به‌صورت جدی به مطالعه وندهای اشتقاقی زبان کردی گویش سورانی در چارچوب یک پیکره زبانی پرداخته است. تعیین میزان زایایی پسوند‌های اشتقاقی در حجم وسیعی از داده و همچنین میزان اهمیت وندهای موجود از موارد مطالعاتی این پژوهش بوده است. پاسخ به پرسش‌هایی از قبیل آنکه کدام‌یک از وندهای اشتقاقی در زبان کردی زایاتر است از جمله موارد مطرح شده در این پژوهش است. همچنین تهیه و گردآوری فهرست وندهای اشتقاقی زبان کردی، استفاده از پیکره عظیم دادگانی ده‌ساله از متون خبری کردی سورانی از جمله نکات مثبت پژوهش انجام شده است.

۲- پیشینه تحقیق

مطالعات حوزه صرف درباره موضوعاتی چون واژه‌سازی، مکان‌وندها، میزان زایایی وندها، اهمیت وندها و تغییرات در زمانی وندها است (mendaza, 2015; Bauer, 2004). توجه به اینکه وندهای

در طول زمان به چه صورتی در واژه‌سازی عمل کرده‌است، اینکه افزودن وندهی سبب چه تغییرات معنایی و دستوری در یک کلمه شده‌است، اینکه چه محدودیت‌هایی سبب می‌شود تا برخی وندها برای برخی از کلمات به کار نرود، توجه به انواع وندهای به‌کاررفته در زبان و میزان زایایی و اهمیتی که آنها در واژه‌سازی ایفا می‌کنند از نمونه مواردی بوده‌است که همواره در علم صرف بررسی شده‌است (Motsch, 2018).

وندها را براساس نوع فرایند ساخت‌واژی به دو دسته؛ وند اشتقاقی و وند تصریفی تقسیم می‌کنند. وند اشتقاقی وندهی است که با افزودن به یک واژه، معنا یا نقش آن واژه را تغییر می‌دهد. وند تصریفی وندهی است که با افزودن به یک واژه، صورت‌های نحوی مختلف واژه را فراهم می‌کند. وندهای تصریفی برخلاف وندهای اشتقاقی واژه جدید نمی‌سازند بلکه موجب به وجود آمدن صورت‌های گوناگونی از یک واژه می‌شوند که در موقعیت‌های مختلف برای سازگاری با نحو مصرف می‌شوند. با به‌کارگیری وندهای تصریفی می‌توان ابعاد دستوری شمار، شخص، زمان دستوری، نمود، حالت و ... را باز نمود. پژوهشگران ملاک‌های متفاوتی برای متمایز ساختن وندهای تصریفی از اشتقاقی مطرح کرده‌اند. به‌طور نمونه مارتین هاسپلمت (1996) ملاک‌های دوازده‌گانه‌ای برای تمایز وند صرفی و اشتقاقی ارائه می‌دهد. این ملاک‌ها برای وندهای اشتقاقی بدین‌صورت بیان شده‌است: به نحو ربط ندارد؛ اختیاری است؛ می‌توان آن را با کلمه بسیط جایگزین کرد؛ مفهوم جدید می‌سازد؛ معنای نسبتاً عینی دارد؛ خیلی به معنای پایه نزدیک است؛ کاربرد آن محدودیت دارد؛ به پایه نزدیک‌تر است؛ تکواژگونگی پایه بیشتر است^(۲)؛ معنای چندگانه صرفی نداریم؛ تکرارپذیر است؛ احتمالاً به لحاظ معنایی بی‌قاعده است. از منظری دیگر، وندها را می‌توان براساس محل اتصال به انواع مختلفی همچون پیشوند، پسوند، میانوند، درون‌وند، دگروند، زبروند و پیراوند و تراوند تقسیم کرد (شقاقی، ۱۳۸۷: ۴۲). هر زبانی برحسب ویژگی‌های صرفی و ساخت‌واژی که دارد برخی از این وندها را شامل می‌شود و ممکن است برخی از وندها را در ساختار کلمات خود نداشته باشد.

باین‌حال دو نوع متداول وند که در اکثر زبان‌ها مشاهده می‌گردد وندهای پیشوندی و پسوندی است. پیشوند تکواژ وابسته‌ای است که به آغاز پایه متصل می‌شود. پسوند تکواژ وابسته‌ای است که به انتهای پایه متصل می‌شود. در میان‌وندهای یک زبان به‌طور معمول برخی از وندها در واژه‌سازی فعال‌ترند و تعداد واژگان جدید بیشتری با آنها ساخته می‌شود و تعداد دیگری از وندها زایایی کمتری دارند. حتی این میزان زایایی وندها ممکن است در طول

زمان تغییر کند و در زمان‌های خاص زیاتر باشند و گاه به وندهای سترون تبدیل شوند (Cowie, 2002).

در بافت‌های واژه‌سازی برای زایایی تعاریف متعددی بیان کرده‌اند. گایتا و ریکا برحسب آنچه رینر در مقاله خود بیان کرده‌است فهرستی از تعاریف زایایی را ارائه داده‌اند (Gaeta, 2006). این تعاریف عبارت‌اند از:

- لغاتی که با قانون واژه‌سازی مخصوص ساخته می‌شود.
 - لغات جدیدی که با قانون واژه‌سازی به‌خصوص در زمانی خاص ساخته می‌شود.
 - امکان ساخت لغات جدید با قانون واژه‌سازی به‌خصوص
 - احتمال لغات جدید ساخته‌شده با قانون به‌خصوص واژه‌سازی
 - شماری از لغاتی که با یک قانون واژه‌سازی به‌خصوص ساخته‌شده‌است.
 - رابطه بین لغات ممکن و بالقوه براساس قانون واژه‌سازی به‌خصوص
- علم صرف با رویکردی پیکره‌بنیان به مطالعه توزیع وندها و بازیایی مواردی که وند در کلمات رخ داده‌است می‌پردازد. زایایی را تعداد واحدهای جدیدی تعریف می‌کنند که در طول زمان برحسب ضرورت و نیاز افراد ساخته می‌شود و به‌صورت کلی به‌صورت طیفی که سمت چپ آن واحدهای نازا و سمت راست آن زیاترین واحدها قرار دارند در نظر می‌گیرند (Stefanovich, 2020: 315). به‌طور نمونه در زبان انگلیسی پسوند -ness ممکن است به صفت‌های زیادی برای ساخت اسامی انتزاعی متصل شود. در مقابل پسوند -th پسوندی است سترون و هرگز به صفت‌ها برای ساخت واژه‌های جدید متصل نمی‌شود. درجه‌ای از الگوی صرفی با استفاده از واژه‌های جدیدی که ساخته می‌شود اندازه‌گیری می‌شود. توجه به اینکه در گذشته چه مقدار واژه ساخته‌شده‌است در بررسی زایایی وند جایگاهی قابل بررسی ندارد. چندین روش در خصوص بررسی میزان زایایی فرایندهای واژه‌سازی بیان شده‌است که هر کدام برحسب نوع هدفی که دنبال می‌شوند با یکدیگر تفاوت دارند (ارجمندی و همکاران، ۱۳۹۲):

- شمارش برون داده‌های فرایندهای ساخت‌واژی
- استفاده از واژگان ممکن
- استفاده از واژگان تک‌بسامدی
- شمارش نوواژه‌ها

توجه به داده و هدف‌های این پژوهش سبب گردید تا برای بررسی میزان زبانی برحسب معیار باین^(۳) از روش واژگان تک‌بسامدی استفاده گردد. هنگامی که در مطالعات به بسامد کلمه، وند، گروه و موارد مشابه توجه می‌شود، موضوع یک واحد در نظر گرفته می‌شود. در ادامه توجه می‌گردد که چه مقدار از این واحدها تکرار می‌گردند و چه مقدار تنها یک‌بار به کار می‌روند. به اینکه هر واحد با تمام صورت‌های تکراری آن چند بار در متن رخ داده‌است توکن^۱ می‌گویند. تعداد وقوع واحد مورد مطالعه در پیکره را بدون در نظر گرفتن موارد تکراری گونه^۲ می‌گویند. به‌طور نمونه در متنی که در ادامه آمده‌است، اگر به کلمه کتاب توجه شود، مشاهده می‌گردد که میزان توکن این واحد (کتاب‌ها، کتاب، کتاب، کتاب، کتابش، کتابی) برابر با ۶ و میزان گونه این واحد (کتاب‌ها، کتاب، کتابی، کتابش) برابر با ۴ است.

«مادر کتاب‌ها را بر روز میز گذاشت. سه کتاب بود. کتاب جغرافیا، دینی و یک کتاب قصه برای سارا. سارا سریع آمد و کتابش را از روی میز برداشت. علی هم بی‌حوصله کتابی از روی میز برداشت و رفت.»

بدین صورت اگر مطالعات صرفی را در حوزه و ندهای اشتقاقی یا تصریفی انجام دهیم می‌توان میزان توکن و گونه یک وند را مشخص گردانید. به‌طور نمونه اگر به متنی که در ادامه می‌آید توجه کنید، کلمات «همسایه‌ها، پول‌ها، پول‌ها و کارها» دارای نشانه جمع «ها» هستند. در این میان کلمه «پول‌ها» دو بار آمده‌است و بقیه یک بار. بدین ترتیب تعداد توکن وند «ها» برابر با ۴ و تعداد گونه آن در این متن کوتاه برابر با ۳ است. تعداد کلمات تک‌بسامدی با وند «ها» نیز ۲ است؛ «همسایه‌ها» و «کارها».

«بدون اینکه بگذاریم کسی از همسایه‌ها بفهمد، شب اول با ترس‌ولرز زیاد وارد خانه‌اش شدیم و دزدی کردیم. بعد هم با خود کنار آمدیم که: این‌که دزدی نیست تازه او به این پول‌ها دیگر هیچ احتیاجی هم ندارد. تازه می‌توانیم کمی هم از این پول‌ها را از طرفش صرف کارها خیر کنیم تا هم خودش سود برده باشد و هم ما.»

زبانی به معنی توانایی تولید صورت‌های جدید براساس قواعد موجود در زبان از ویژگی‌های مهم زبان است که در تمامی بخش‌های آن، از جمله حوزه ساخت‌واژه، وجود دارد. اکثر صاحب‌نظران زبانی را یک پیوستار می‌دانند که از کمترین مقدار زبانی تا بیشترین

1. Token
2. Type

مقدار زایایی در آن مشاهده می‌گردد. محدود کردن مفهوم زایایی به بسامد کلمات از اشتباهاتی است که سبب تقلیل مفهوم زایایی می‌گردد. مطالعات نشان می‌دهد که گاه بسامد بالای کلمات به خاطر آن است که وند در تعداد محدودی از کلمات رخ می‌دهد ولی خود کلمه‌ای که وند در آن رخ داده‌است از کلمات پرکاربرد با بسامد بالاست. گاه هم وند خود در تعداد بی‌شماری از کلمات رخ می‌دهد و همین سبب افزایش شمار بسامد وند است (Stefanowitsch, 2020). بدین ترتیب با استفاده از فرمول زایایی HTR سعی می‌شود تا در مطالعات زبانی میزان تأثیر عوامل بیرونی و بسامد تنها را بر نتیجه نهایی کاهش دهند و به جواب منطقی برسند. در فرمول زایایی را ارزشی می‌دانند که از تقسیم تعداد کلمات تک‌بسامدی که تنها یک بار برحسب ضرورت و نیاز افراد ساخته شده‌است بر تعداد کل گونه‌های واحد بررسی شده به دست آمده‌است. کلمات تک‌بسامد کلمات واژه‌های جدید هستند و هرچه تعدادشان بیشتر باشد یعنی واژه‌های بیشتری با وند مربوطه ساخته شده‌است. توکن‌ها نیز میزان واژه‌های ساخته شده از وند مربوطه بدون توجه به صورت‌های تکراری‌شان آن در متن هستند. فرمول زایایی باین برحسب کلمات تک‌بسامدی در ادامه به صورت نمادین نمایش داده شده‌است:

$$HTR = \frac{n(\text{hapax legomena})}{n(\text{token})}$$

مطابق فرمول بالا در مثال بیان شده در توضیحات توکن و گونه می‌توان میزان زایایی وند «ها» را در متن عدد $\frac{2}{3}$ دانست. تعداد کلمات تک‌بسامدی با وند «ها» نیز ۲ است؛ کلمات «همسایه‌ها» و «کارها». تعداد توکن آن ۳ است؛ کلمات «همسایه‌ها»، «پول‌ها» و «کارها». البته بررسی میزان زایایی می‌بایست در حجم عظیمی از داده بررسی شود و مثال بخش پیشینه تنها برای فهم بیشتر خواننده است.

از دیگر موارد حوزه صرف میزان اهمیتی^۱ است که یک وند دارد. میزان اهمیت یک وند در تعداد کلماتی است که در فهرست واژگانی برای زبان به کار برده می‌شود. این بدان معناست که به ساخت کلمات جدید توجهی نمی‌شود بلکه به تعداد کلماتی که با وند خاص وجود دارد توجه می‌شود. بدین ترتیب وندی که در کلمات بیشتری به کار برده شده‌است مهم‌تر از وندی است که در تعداد معدودی از کلمات به کار رفته‌است. هنگام بررسی میزان اهمیت یک وند نمی‌توان به این بسنده کرد که چون بسامد یک وند بالاست، اهمیت زیادی دارد. چون ممکن

1. Importance

است یک وند در چندین کلمه به‌خصوص رخ دهد و این کلمات بسامد بالایی داشته باشند نتوانند میزان اهمیت وند را نشان دهند. یا در جهت عکس این قضیه می‌توان گفت که گاه وندی در تعداد زیادی از کلمات با بسامد بسیار کمتر از حد طبیعی به‌کار برده می‌شود و همین امر سبب می‌شود تا نتوان میزان اهمیت یک وند را بررسی کرد. استفانوویچ در این زمینه مثالی دارد؛ در زبان انگلیسی اگر دو وند *icle*- و *mini*- را بررسی کنیم متوجه می‌شویم که هرکدام بسامد به ترتیب مقادیر ۲۰۷۷۲ و ۱۷۰۲ دارند. اما اگر به گونه‌ای دسته‌ای از وندها توجه کنیم و به تکرار کلمات تکراری توجه نکنیم متوجه می‌شویم که هر یک میزان بسامد گونه‌ای برابر با ۹۲ و ۳۶۷۳ دارند و همین به‌خوبی نشان می‌دهد که اگرچه وند *icle*- دارای بسامد بالایی است، ولی این مقدار به‌صورت ۹۲ کلمه بوده‌است که در بافت‌های مختلف پیکره تکرار شده‌است و از این ۹۲ کلمه سه کلمه وجود داشته‌است که ۹۲ درصد مواقع رخ داده. همین ۹۲ درصد را اگر برای وند *mini*- بررسی کنیم مشخص می‌گردد که تعداد ۲۵۳ گونه از کلمات برحسب این وند به‌کار می‌رود. درنهایت نیز مشخص می‌گردد که چون برای میزان ۹۲ درصد تعداد کلمات وند *mini*- بیشتر است و ۲۵۳ با ۳ تفاوت بسیار زیادی دارد، میزان اهمیت وند *mini*- از وند *icle*- بیشتر است. استفانوویچ درنهایت در کتاب پیکره‌شناسی خویش فرمولی ارائه می‌دهد که برحسب آن می‌توان میزان اهمیت یک وند را نیز بررسی کرد که برابر با میزان تعداد گونه‌ی یک وند بر کل کلمات با تکراری است که از یک وند ساخته می‌گردد. فرمول زیر نشان‌دهنده‌ی میزان اهمیت وند هست.

$$TTR = \frac{n(\text{type})}{n(\text{token})}$$

این نکته نیز مهم است که اهمیت یک وند و میزان زایایی آن دو مفهوم مجزا هستند و طریقه‌ی محاسبه‌ی مجزایی دارند. یک وند ممکن است میزان اهمیت بالایی داشته باشد و در کلمات زیادی به‌کار برده شود، ولی کلمات شناسایی‌شده مربوط به زمان قبل بوده باشد و از کلماتی نباشد که به‌تازگی ساخته شده‌است. درحالی‌که مفهوم زایایی به کلمات جدیدی اشاره دارد که با بسامد یک به‌تازگی بر جمع کلمات واژگان یک زبان اضافه شده‌است.

در طول زمان مطالعاتی در زمینه‌ی زایایی وندها در زبان‌های مختلف انجام شده‌است. در زبان ایتالیایی پس از محاسبه‌ی میزان زایایی هر وند مطابق قانون باین، به مقایسه‌ی میزان زایایی هر وند با توکن یکسان پرداخته شده‌است و مشخص شده که کدام وندها زایاتر است (Gaeta & Ricca, 2006). در زبان فارسی نیز پس‌ازآنکه میزان زایایی وندها برحسب قانون باین

(Baayen, 2009) مشخص گردیده است، به بررسی معناپذیری کلمات ساخته شده و زایایی و ندها پرداخته شده است (کهن‌زاد و همکاران، ۱۳۹۸). در پژوهشی دیگر نیز زایایی صرفی از منظر کمی و کیفی بررسی شده و نیز رابطه بسامد و کاربردشناسی مطالعه شده است (علوی‌مقدم، ۱۳۸۶). در زبان انگلیسی میزان زایایی و ندها بررسی شده است (Montero-Fleta, 2011). همچنین توجه به عوامل معنایی که بر میزان زایایی کلمات تأثیر دارد مطالعه شده است (Schweikhard, 2018). توجه به زایایی و پیش‌بینی واژگانی که در پردازش زبانی صورت می‌گیرد نیز موضوع بررسی شده دیگر است (Ten Hacken, 2020). مطالعات در زبان‌های گوناگون نشان می‌دهد که در زبان کردی ابتدا می‌بایست زایایی و ندهای اشتقاقی محاسبه و سپس دیگر موارد مطالعاتی زایایی بررسی گردد.

۳- روش‌شناسی

در مطالعات زبان کردی آنچه توجه پژوهشگر را به خود جلب می‌کند و اهمیت دارد نوع داده استفاده شده است. به دلیل تنوع بالای زبان کردی و همچنین نبود پیکره‌های بزرگ دیجیتالی و غیردیجیتالی امکان مطالعات پیکره‌بنیاد کمتر بوده است. به همین دلیل در نخستین گام می‌بایست دادگان مناسب و کافی برای پژوهش تهیه گردد. پس سعی شد تا با استفاده از پیکره دادگان زبان کردی سورانی و برنامه خزش وب دادگان کافی برای مطالعه فراهم شود. پروژه پیکره زبان کردی نخستین پیکره آنلاین موجود به زبان کردی گویش سورانی است که در دانشگاه کردستان شروع به فعالیت نموده است (Badakhshan, 2017). در جمع‌آوری متون این پیکره از وبسایت‌های خبری کردی مانند کردپرس استفاده شده است و پس از نرمال‌سازی^(۴) گزارش آماری زیر به دست آمده است. نسخه اولیه این پیکره شامل ۶۹۰۰۰ سند خبری است که محتوی آن اخبار گوناگون است. اسناد ۱۴۸۹۸۰۶۲ کلمه دارد که ۴۳۶۶۵۵ واژه منحصر به فرد است. پراکندگی زمانی از سال ۱۳۸۹ تا ۱۳۹۴ است. خبرگزاری کردپرس از آن جهت که رسانه‌ای رسمی برای پوشش اخبار کردستان ایران، عراق، ترکیه و سوریه بر حسب کردی سورانی است، اعتبار کافی و لازم برای سنجش داده‌های کردی دارد. به همین دلیل با برنامه خزش وب سعی شد تا منابع خبری از سال ۱۳۹۴ تا ۱۴۰۰ تهیه گردد. برنامه خزش وب با زبان پایتون نسخه ۳,۴ نوشته شده است. در نهایت نزدیک به ۳۰۰۰۰ سند جدید

به دادگان پیشین اضافه گردید. پس از جمع‌آوری داده کافی به مشخص کردن وندهای زبان کردی پرداخته شد^(۵).

در ادامه فهرست مشخصی از پسوندهای اشتقاقی در زبان کردی با استفاده از کتاب‌های دستور زبان کردی (چمن‌آرا، ۱۳۹۹) و توصیف ساخت‌واژه گویش سورانی زبان کردی (رحیمی، ۱۳۹۹) مشخص گردید. به دلیل پیچیدگی‌های مطالعات وندهای تک‌حرفی از مطالعه این‌گونه وندها صرف‌نظر گردید. برای نمونه اگر به پسوند ه در زبان فارسی توجه کنیم می‌بینیم که این وند در سه نوع ساختار وجود دارد:

- به بن مضارع می‌چسبد و اسم‌افزار می‌سازد، مانند: استره (از استردن)، ماله، رنده (از رندیدن)، تراشه، پرونده، لوله (از لولیدن)، سنبه (از سنبیدن).

- به بن مضارع می‌چسبد و اسم مصدر می‌سازد، مانند: خنده، مویه، گریه، آموزه.

- بیانگر اسم معرفه است، مانند: زنه، پسره، کارمنده.

از جهت دیگر همین پسوند در انتهای بسیاری از کلمات ریشه کلمه است و وند نیست. پس چون شناسایی اولیه کلمات ونددار با برنامه‌نویسی انجام می‌پذیرد و برنامه کلمه را بدون در نظر گرفتن ساختار و براساس نشانه جست‌وجو می‌کند مشکلاتی در شناسایی کلمات پسونددار ایجاد می‌گردد. به همین دلیل مطالعه پسوندهای تک‌حرفی به زمان دیگری موکول گردید.

جدول ۱- فهرست وندهای بررسی شده

پسوند	آوانگاری	پسوند	آوانگاری
-گار	/gar/	-گه	/gæ/
-گا	/gɑ/	-گهر	/gæɾ/
-مند	/ænd/	-ایی	/ayi/
-یژه	/zæ/	-نده	/ndæ/
-هوهری	/æwæri/	-ئنه	/enæ/
-یار	/yar/	-یده	/idæ/
-ان	/an/	-مان	/man/
-ۆلکه	\ olkæ \	-هر	/ær/
-ی‌هت	/yat/	-ۆک	/ok/
-ی‌هر	/yær/	-انی	/ani/
-یده	/idæ/	-اوی	/awi/

\ zən\	-زن	\ ɪlæ\	-یل/یل‌ه
\ gin\	-گین	\ ɪlənæ\	-یلان‌ه
\ nək\	-ناک	\ ɪle\	-یل‌ئ
\ or\	-ؤر	\ ɪnæ\	-ینه
\ æki\	-ه‌کی	\ ɪni\	-ینی
\ kælæ\	-ک‌ه‌له	\ ælat\	-ه‌لات
\ olæ\	-ؤله	\ æwənæ\	-ه‌وان‌ه
\ ol\	-ؤل	\ ger\	-گ‌نپر
\ olæ\	-ؤله	\ gər\	-گر
\ æk\	-ه‌ک	\ nok\	-ن‌ؤک
\ æl\	-ه‌ل	\ gir\	-گیر
\ æmæni\	-ه‌م‌ه‌نی	\ æn\	-ه‌ن
\ æwɑ\	-ه‌وا	\ əndæ\	-نده
\ æwar\	-ه‌وار	\ wænd\	-ه‌وند
\ ilkæ\	-یل‌ک‌ه	\ ulæ\	-وول‌ه
\ ilæ\	-یل‌ه	\ ulkæ\	-وول‌ک‌ه
\ in\	-ین	\ olæ\	-ؤل‌ه
\ inæ\	-ین‌ه	\ æti\	-ه‌تی
\ yæti\	-ی‌ه‌تی	/ kilæ/	-کیل‌ه
\ en\	-ئ‌ن	\ æz\	-ه‌ز
\ enæ\	-ئ‌ن‌ه	\ ælæ\	-ه‌له
\ eni\	-ئ‌نی	\ ælæk\	-ه‌له‌ک
\ tʃæ\	-چ‌ه	\ zar\	-زار
\ tʃi\	-چ‌ی	\ Səz\	-ساز
\ xənæ\	-خ‌انه	\ stən\	-ستان
\ dar\	-دار	\ kæ\	-ک‌ه
\ dan\	-دان	\ kær\	-ک‌ه‌ر
\ æwær\	-ه‌و‌ه‌ر	\ læ\	-له
\ U\	-وو	\ mænd\	-م‌ه‌ند
\ ur\	-و‌ور	\ əndæ\	-ه‌نده
\ ulæ\	-و‌وله	\ æwən\	-ه‌وان
\ ulkæ\	-و‌ول‌ک‌ه	\ wənæ\	-وان‌ه
/læ/	-ل‌ه	/ bənd/	-ب‌ه‌ند

/ ælan/	-ه‌آن	/ Uwæ/	-هوه
/ waz/	-واز	/ næ/	-انه
/ bayi/	-بایی	/ ol/	-ؤل
/ bæɾ/	-بهر	/ ari/	-اری
/ xan/	-خان	/ aɫæɫ/	-ا٬ت
/ ak/	-اک	/ ælæ/	-له
/ dæ/	-ده	/ ane/	-انه‌ئ
/ daɾ/	-دەر	/ awa/	-اوا
/ ban/	-بان	/ awæɾ/	-اوفر
/ zilæ/	-ژیل‌ه	/ bazi/	-بازی
/ zæ/	-ژه	/ ayæti/	-ای‌ه‌تی
/ olkæ/	-ؤل‌که	/ baz/	-باز
/ kaɾ/	-کار	/ ʒaɾ/	-چار
/ kæ/	-که	/ ʃkæ/	-چکه
/ okæ/	-ؤکه		

پس از آنکه پیکره و وندهای اشتقاقی در زبان کردی جمع‌آوری گردید، به دلیل آنکه از روش شمارش کلمات تک‌بسامدی به بررسی میزان زایایی وندها پرداخته می‌شود، با استفاده از برنامه‌نویسی بسامد تمامی کلمات پیکره محاسبه گردید و سپس به کلماتی پرداخته شد که بسامدی برابر با عدد یک دارند. در نهایت کلماتی که بسامد یک داشتند و وندهای آنها به موارد اشاره شده در بالا ختم می‌شد شناسایی گردید. پس از آنکه مشخص شد هر ونده در ساخت چند کلمه تک‌بسامدی حضور داشته‌است، برای کسب نتایج صحیح‌تر از محاسبه زایایی وندهایی که میزان تأثیرشان در ساخت‌واژه تک‌بسامدی کمتر از بسامد ۵ بود صرف‌نظر گردید. جداولی براساس بسامد فرمول زایایی ترسیم گردید که در ادامه می‌آید:

جدول ۲- نحوه قرار دادن اطلاعات برحسب فرمول زیایی

نسبت کلمات تک‌بسامدی به توکن وندها				
		گونه		مجموع
		تک‌بسامدی	غیر تک‌بسامدی	
وندها	۱ وند
	۲ وند
	مجموع

برای بررسی میزان اهمیت وندها نیز جداولی مشابه جداول بالا برحسب فرمول TTR ترسیم گردید و درنهایت نتایج در جداولی مشخص نمایش داده شد.

۴- یافته‌ها

پس از آنکه وندهای اشتقاقی مشخص گردید، پایتون میزان بسامد گونه، توکن، تک‌بسامدی برای هر کلمه با استفاده از برنامه‌نویسی به زبان پرداخته شد و مشخص گردید که هر یک از وندها در بخش انتهایی کلمه به چه صورت رخ داده‌است. برای مطالعه دقیق‌تر، با سامانه به بررسی کلمات تک‌بسامدی شناسایی شده پرداخته شد تا موارد به‌اشتباه شناسایی شده حذف گردند. همچنین وندهایی که تعداد کلمات تک‌بسامدی آنها از مقدار ۵ کمتر بود نادیده گرفته شد. نتایج با ارزش میزان زیایی و اهمیت وندها در جدول ۳ در ادامه نمایش داده شده‌است.

جدول ۳- نتایج به‌دست‌آمده از میزان زیایی و اهمیت هر پسوند در پیکره

میزان اهمیت	میزان زیایی	تک‌بسامدی	گونه	توکن	آوانویسی	وند
۰,۱۷۴۸	۰,۳۸۲	۲۶۴	۶۹۱	۳۹۵۱	/xanæ/	انه
۰,۰۴۰۴	۰,۳۱۲۸	۵۱	۱۶۳	۴۰۲۵	/ayæti/	ای‌تی
۰,۰۳۶۳	۰,۳۶۳۶	۸	۲۲	۶۰۶	/baz/	باز
۰,۰۸۱۴	۰,۲۷۲۷	۶	۲۲	۲۷۰	/ʃkæ/	چکه
۰,۰۳۲۹	۰,۵	۶۳	۱۲۸	۳۸۸۴	/ʃi/	چی
۰,۰۴۹۹	۰,۲۵۲۸	۲۲	۸۷	۱۷۴۲	/xanæ/	خانه
۰,۰۲۳۹۲	۰,۳۱۸۳	۷۸	۲۴۵	۱۰۲۴۲	/dar/	دار
۰,۰۶۵۴	۰,۵	۱۱	۲۲	۳۳۶	/dan/	دان
۰,۰۳۵۳	۰,۳۷۵	۶	۱۶	۴۵۲	/saz/	ساز
۰,۰۴۰۸	۰,۳۵۷۹	۱۵۰	۴۱۹	۱۰۲۴۷	/stæn/	ستان

که	/kæ/	۷۳۵۴	۱۲۳	۴۱	۰,۳۳۳۳	۰,۱۶۷۲
کەر	/kær/	۲۴۶	۵۰	۱۷	۰,۳۴	۰,۲۰۳۲
گا	/gɑ/	۴۹۵۶	۱۰۱	۲۵	۰,۲۴۷۵	۰,۰۲۰۳
گه	/gæ/	۲۷۷۱	۴۴	۱۴	۰,۳۱۸۱	۰,۰۱۵۸
له	/læ/	۷۹۰۵	۱۹	۹	۰,۴۷۳۶	۰,۰۰۲۴
نه	/læ/	۴۲۸	۳۱	۹	۰,۲۹۰۳	۰,۰۷۲۴
م‌دند	/mænd\	۲۹۹	۴۴	۱۳	۰,۲۹۵۴	۰,۱۴۷۱
وو	/u/	۱۶۸۰	۳۱	۱۲	۰,۳۸۷	۰,۰۱۸۴
وان‌ه	/wanæ/	۵۸۷	۶۰	۲۰	۰,۳۳۳۳	۰,۱۰۲۲
یل‌ه	/ilæ/	۵۵	۲۱	۹	۰,۴۲۸۵	۰,۳۸۱۸
ین	/in/	۱۳۲	۸	۴	۰,۵	۰,۰۶۰۶
ین‌ه	/inæ/	۱۴۱	۲۴	۷	۰,۲۹۱۶	۰,۱۷۰۲
ی‌دتی	/yæti/	۲۰۹۹	۱۸۸	۸۲	۰,۴۳۶۱	۰,۰۸۹۵
ئن	/en/	۸۹۴	۷۲	۳۱	۰,۴۳۰۵	۰,۰۸۰۵
ایی	/ayi/	۳۱۷۸۸	۶۲۳	۲۵۷	۰,۴۰۶	۰,۱۹۹۱
ەر	/ær/	۹۰۳۴	۱۹۳	۷۰	۰,۳۶۲۶	۰,۰۲۱۳
یار	/yar/	۲۰۹۲۳	۳۱۱	۹۶	۰,۳۰۸۶	۰,۰۱۴۸
ؤک	/ok/	۱۶۴۹۲	۱۳۴	۵۰	۰,۳۷۳۱	۰,۰۰۸۱
انه	/anæ/	۷۲۴۲	۶۲۳	۲۱۴	۰,۳۴۳۴	۰,۰۸۶
ائی	/ani/	۹۰۹	۲۷	۱۲	۰,۴۴۴۴	۰,۰۲۹۷
اوی	/awi/	۴۱۲۲۲	۱۰۵۶	۴۳۳	۰,۴۱	۰,۰۲۵۶
ه‌کی	/æki/	۷۳	۱۳	۶	۰,۴۶۱۵	۰,۱۷۸
زن	/zæn/	۱۳۱	۱۰	۵	۰,۵	۰,۰۷۶۳
گین	/gin/	۷۹	۱۴	۶	۰,۴۲۸۵	۰,۱۷۷۲
ناک	/nak/	۲۸۰	۱۲	۵	۰,۴۱۶۶	۰,۰۴۲۸
اندن	/andæn/	۵۷۱۶	۳۳۴	۸۶	۰,۲۵۷۴	۰,۰۵۸۴
اری	/ari/	۱۶۲۱۴	۲۰۵	۸۱	۰,۳۹۵۱	۰,۰۱۲۶
اوا	/awɑ/	۲۷۰۴	۳۴	۱۲	۰,۳۵۲۹	۰,۰۱۲۵
اوه	/awæ/	۳۰۴۱	۱۵۱	۶۲	۰,۴۱۰۵	۰,۰۴۹۶
واز	/waz/	۵۰۸	۱۵	۵	۰,۳۳۳۳	۰,۰۲۹۵
بایی	/bæyi/	۱۶۸۰	۴۱	۱۹	۰,۴۶۳۴	۰,۰۲۴۴
چه	/tʃæ/	۶۷۷۱	۸۱	۲۵	۰,۳۰۸۶	۰,۰۱۱۹
خان	/xan/	۴۱۶	۱۸	۷	۰,۳۸۸۸	۰,۰۴۳۲
خان‌ه	/xanæ/	۳۳۶	۳۷	۵	۰,۱۳۵۱	۰,۰۴۳۲

ده	/dæ/	۳۷۲	۲۲	۷	۰,۳۱۸۱	۰,۰۵۹۱
دەر	/dær/	۵۷۳	۴۷	۱۳	۰,۲۷۶۵	۰,۰۸۲
بان	/yan/	۴۷۵	۳۵	۱۴	۰,۴	۰,۰۷۳۶
ژه	/zæ/	۵۴۵	۱۵	۵	۰,۳۳۳۳	۰,۰۲۷۵
شت	/ʃt/	۱۱۶۴	۳۷	۱۱	۰,۲۹۷۲	۰,۰۳۱۷
کار	/kær/	۹۵۱۶	۲۴۴	۶۵	۰,۲۶۶۳	۰,۰۲۵۶
گر	/gær/	۴۷۴	۲۹	۸	۰,۲۷۵۸	۰,۰۶۱۱
گیر	/gir/	۷۷۷۴	۹۴	۳۰	۰,۳۱۹۱	۰,۰۱۲
نده	/ændæ/	۲۱۶۳	۶۷	۲۱	۰,۳۱۳۴	۰,۰۳۰۹
هئه	/ælä/	۶۰۶	۲۹	۷	۰,۲۴۱۳	۰,۰۴۷۸
هئد	/ænd/	۲۷۰۱	۶۲	۱۸	۰,۲۹۰۳	۰,۰۲۲۹
هتی	/æti/	۶۸۳۴	۱۹۳	۸	۰,۰۴۱۴	۰,۰۲۸۲
یار	/yær/	۲۰۹۲۳	۳۱۱	۹۶	۰,۳۰۸۶	۰,۰۱۴۸
یان	/yan/	۷۲۹۵	۶۹۸	۳۵۷	۰,۵۱۱۴	۰,۰۹۵۶
یل	/Il/	۲۸۴	۱۲	۵	۰,۴۱۶۶	۰,۰۴۲۲
یلانه	/lönæ\	۳۱	۹	۵	۰,۵۵۵۵	۰,۲۹۰۳
هوانه	/æwanæ/	۴۱۸	۲۱	۷	۰,۳۳۳۳	۰,۰۵۰۲
ینه	/inæ/	۵۶۹	۵۵	۲۱	۰,۳۸۱۸	۰,۰۹۶۶

بدین ترتیب، فهرستی از زیاترین وندها با زایایی بالاتر از ۵۰ درصد از جدول ۳ مشخص گردید و معلوم شد میزان زایایی وندهای اشتقاقی در جایگاه پایانی کلمه ۳۷,۲۵ درصد است.

۵- بحث و نتیجه‌گیری

بررسی میزان زایایی وندهای اشتقاقی در جایگاه پایانی کلمه و میزان اهمیت وندهای کردی از اهداف اصلی این پژوهش بود. فراهم کردن داده کافی و قابل اعتماد از نخستین اقدامات بود. با استفاده از پیکره زبان کردی و برنامه خزش وب نزدیک به ۹۰۰۰۰ سند خبری در ۱۰ سال اخیر بررسی شد. براساس کتب دستوری زبان کردی که به توصیف ساختار گویش سورانی پرداخته بود، فهرستی از وندهای اشتقاقی تهیه گردید. در مرحله بعد به محاسبه میزان زایایی برای هر یک از وندها پرداخته شد. زایایی را تعداد واحدهای جدیدی تعریف می‌کنند که در طول زمان برحسب ضرورت و نیاز افراد ساخته می‌شود و میزان اهمیت یک ونده را در تعداد کلماتی می‌دانند که در فهرست واژگانی برای زبان به کار برده می‌شود. برای بررسی این دو

مفهوم در زبان کردی، ابتدا و سپس تعداد کلمات گونه، توکن و تک‌بسامدی برای هر وند در پیکره تهیه‌شده مشخص گردید. آنچه از مشاهدات و نمودارهای ترسیم‌شده به‌دست‌آمد نشان می‌دهد آن وندهای «چی، دان، ین، زن، وه، یلان» از زایاترین وندها هستند که کلمات جدیدی را به زبان کردی اضافه کرده‌اند. همچنین زایایی در فرایند واژه‌سازی به‌صورت طیفی از کم به زیاد است. در این طیف در سمت چپ وندهای نازا و در سمت راست زایاترین وندها قرار دارند. وندهای «هوان، یلان، گین، یله» از مهم‌ترین وندهای کردی هستند. در مطالعات پیکره‌بنیاد توجه به بسامد واحدها نمی‌تواند نتایج دقیقی را نشان دهد. چون عوامل زیادی ممکن است در میزان بسامد کلمات تأثیرگذار باشد.

این از نخستین پژوهش‌هایی است که در زبان کردی سورانی به‌صورت پیکره‌بنیان انجام گردیده‌است و می‌تواند مقدمه‌ای بر مطالعات آینده باشد. توجه به پیشوندهای زبان کردی، ترتیب قرارگیری وندها و توجه به سلسله‌مراتبی که در میزان زایایی وندها و یا اهمیت آنها نقش دارند، از مواردی هستند که می‌تواند در آینده موضوع تحقیق در زمینه تکمیل این پژوهش باشد. رابطه میزان زایایی و یا بررسی نوع ساختار هجا از مواردی است که می‌توان در زایایی بررسی کرد.

پی‌نوشت

۱. <https://www.kurdpress.com>

۲. تکواژگونگی را به سه دسته واجی، تکمیلی ضعیف و تکمیلی قوی تقسیم می‌کنند که می‌تواند به ترتیب دسته‌بندی به‌صورت صورت‌های تکواژ جمع، صورت افعال بی‌قاعده و صورت‌های ساده و تفضیلی صفات بی‌قاعده باشد. در این دسته‌بندی به‌تدریج از میزان شباهت صوری کاسته می‌شود.

۳. در این روش شاخص زایایی برابر است با واژه‌های تک‌بسامدی ساخته شده با یک فرایند ساخت‌واژی خاص، تقسیم بر کل نمونه‌های تک‌بسامدی.

۴. از چالش‌هایی که در زبان فارسی برای سامانه‌های پردازش متن وجود دارد، مشکلات نگارشی زبان فارسی است. به همین دلیل متخصصان با نرمال‌سازی تلاش می‌کنند تا این مشکلات کمتر شود. برای مثال استفاده نابه‌جا از فاصله به‌جای نیم‌فاصله است. توجه به استفاده از علائم نگارشی فارسی و استفاده از حروف غیرعربی می‌تواند ایرادهای نگارشی متن

را کم کند. از دیگر موارد وجود کلمات چنداملایی و شناسایی کلمات مرکب و در نظر گرفتن آنها به‌عنوان یک کلمه است.

۵. خوانندگانی که تمایل دارند پیکره را مطالعه کنند می‌توانند با ایمیل به نویسندگان مقاله اطلاع دهند تا برایشان ارسال گردد.

منابع

- ارجمندی، امیر. و همکاران. ۱۳۹۲. «زایایی فرایند ترکیب در زبان فارسی»، *زبان‌شناخت*، ۴(۷): ۱-۱۴.
- چمن‌آرا، بهروز. ۱۳۹۹. دستور زبان کردی، سنندج: پژوهشکده کردستان‌شناسی.
- رحیمی، محمد، ۱۳۹۹. توصیف ساخت‌واژه گویش سورانی زبان کردی. تهران: مؤلف.
- شقاقی، ویدا. ۱۳۸۷. مبانی صرف، تهران: سمت.
- علوی مقدم، سیدبهنام. ۱۳۸۶. «صرف و واژگان: واژه‌سازی و زایایی»، *زبان و زبان‌شناسی*، ۱(۵): ۱۴۹-۱۵۷.
- کهن‌زاد، پروانه، محمدهادی فلاحی، و بهاره پهلوان‌زاده. ۱۴۰۰. «بررسی پیکره‌بنیاد زایایی وندهای اشتقاقی زبان فارسی»، *زبان‌شناسی و گویش‌های ایرانی*، ۲(۲۳): ۲۱۹-۲۴۰.
- Baayen R. H. Corpus linguistics in morphology: morphological productivity. *Corpus linguistics. An international handbook*. edited by Anke Lüdeling, Merja Kyto, Berlin: De Gruyter Mouton; 2009: 900-919.
- Bauer L. Morphological productivity. Cambridge: Cambridge University Press; 2004.
- Badakhshan E. Kurdish corpus project. International Institute for the Study of Kurdish Societies First Biennial Conference Germany. Frankfurt ; 2017: 16 - 19.
- Cowie C. Dalton-Puffer Ch. Diachronic word-formation and studying changes in productivity over time: Theoretical and methodological considerations, In *A changing world of words*. edited by Javier E. Díaz Vera. New York: Rodopi; 2002. 410-437.
- Gaeta L. Ricca D. Productivity in Italian word formation: A variable-corpus approach. Berlin: De Gruyter Mouton; 2006. 57-89.
- Haspelmath M. Word-class-changing inflection and morphological theory. In *Yearbook of morphology*. edited by Geert Booij, Jaap Marle. New York: Springer Dordrecht; 1996. 43-66.
- Mendoza R. M. Matching productivity indexes and diachronic evolution: The Old English affixes ful-, -isc-, -cund, and -ful. *Canadian Journal of Linguistics/Revue canadienne de linguistique*. 2015; 60(1): 1-24.

- Montero-Fleta B. Suffixes in word-formation processes in scientific English. LSP Journal-Language for special purposes, professional communication, knowledge management and cognition. 2011; 2(2): 1-11.
- Motsch W. On inactivity, productivity and analogy in derivational processes. In The Contribution of Word-Structure-Theories to the Study of Word Formation. Akademie der Wissenschaften der DDR, Zentralinstitut für Sprachwissenschaft, 2018; 1-30.
- Schweikhard N. Semantic promiscuity as a factor of productivity in word formation. Computer-Assisted Language Comparison in Practice. 2018: 1:50-65.
- Stefanowitsch A. Corpus linguistics: A guide to the methodology, Berlin: Language Science Press; 2020: 308-352. DOI: 10.5281/zenodo.3735822
- Ten H. P. Productivity and Anticipation in Language Processing. SKASE Journal of Theoretical Linguistics. 2020; 17(4):1-17.

روش استناد به این مقاله:

صادقی، سیده سارا؛ براخاصی، نجم‌الدین. و بدخشان، ابراهیم. میزان زبانی وندهای اشتقاقی در زبان کردی سورانی، رویکرد پیکره‌بنیان، زبان فارسی و گویش‌های ایرانی، ۱۴۰۱؛ ۲(۱۴): ۳۵-۵۱. DOI:10.22124/plid.2023.23607.1622

Copyright:

Copyright for this article is retained by the author(s), with first publication rights granted to *Zaban Farsi va Guyeshhay Irani (Persian Language and Iranian Dialects)*. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original work is properly cited.

