



Parsability and Ordering of Persian Derivational Prefixes Maryam Hemasian Etefagh¹ & Adel Rafiei² & Batool Alinezhad³ (71-96)

Abstract

Affix ordering is a theoretical discussion in morphology. There has been some debate over the existence of general principles or mechanisms governing affix ordering. In this regard, we can point to three major approaches. The first approach, the stratum-oriented model (Siegel, 1974; Allen, 1978; Kiparsky, 1982; Mohanan, 1986), asserts that the lexicon has a layered structure and can determine the affixes' combinations. The second approach relies on affix-particular Selectional Restriction (Fabb, 1988; Plag, 1999), and it takes into consideration the morphological, phonological, and semantic restrictions for affix ordering. The third approach conducted the theoretical framework of the present study providing a psycholinguistic model based on the idea that the affix ordering is a function of the degree of parsability of affixes from their bases (Hay 2000; 2002). According to this approach, affixes parsed easily in language processing cannot be close to the base relative to the affixes with a lower degree of parsability. Plag (2003) called the approach as Complexity-Based Ordering (CBO). The current study investigates how the Persian derivational prefix ordering can be described based on the CBO framework? to answer this question, by selecting 17 prefixes, 2200 derived words were extracted from the database. By considering prefixes co-occurrences, each prefix's rank was obtained using directed-graph and upper triangle square matrixes. According to this theoretical framework, we could draw the scatter plot and determine the rank of type and token frequency ratio by using the frequency of derived words and bases. Finally, we assessed the continuum of Persian prefix ordering according to data and using the rank of parsability, type, and token frequency ratio. In line with the findings of Hay and Plag (2004) on English affix ordering, this approach can explain the ordering of Persian prefixes.

Keywords: "affix ordering", "Complexity-Based Ordering Approach", "Persian derivational Prefixes", "rank of parsability", "ratio of parsability".

Received: 29, October, 2018; Accepted: 5, September, 2020

doi
10.22059/jolr.2020.268569.666420
Print ISSN: 2288-1026-Online ISSN: 3362-2676
<https://jolr.ut.ac.ir>

1. Ph.D. Candidate of Linguistics, University of Isfahan, Isfahan, Iran.
2. Email of the corresponding author: a.rafiee@fgn.ui.ac.ir
Assistant Professor of Linguistics, University of Isfahan, Isfahan, Iran.
3. Associate Professor of Linguistics, University of Isfahan, Isfahan, Iran.

1. Introduction

In recent decades, there has been some debate over the existence of general principles governing affix ordering. The three main approaches to affix ordering are the Stratum-oriented Approach (Siegel 1974; Allen 1978; Kiparsky 1982; Selkirk 1982; Mohanan 1986; Giegerich 1999), the affix-particular Selectional Restriction (Fabb 1988; Plag 1999), and the Complexity-based Ordering Approach (CBO) (Hay 2000, 2002). The present study is conducted in the framework of the last one. This article extends the empirical scope of the Complexity-based Ordering Approach (Baayen & Plag, 2009) by investigating the sequence of 17 Persian derivational prefixes. This approach believes that the morphological processing constraints of pre-lexical and lexical can govern the affix combinations by presenting a psychological-linguistic model. For investigating the Persian derivational prefix ordering, first of all, based on the prefix co-occurrences in the studied corpus, the square matrix and the directed graph are obtained then the rank of each affix is gained. The parsability ratios of each affix based on the frequency of occurrence of the type and token of derived words and their bases are obtained by using the scatter plot and the trend line. The Persian continuum of prefixes is drawn through calculating the co-occurrence rank (CO-ranks) and the type and token parsing ratio of derived words. According to CBO, we scrutinize how the Persian derivational prefixes can be justified in the parsability continuum.

2. Literature of Review

The affixation processing was first ascertained in the stratum-oriented model frameworks (level-ordering) by some researchers, such as Siegel (1974), Allen (1978), Kiparsky (1982), and Mohanan (1986). They assumed that the affixation restrictions could be explained by considering affixes in two different lexical strata. This model has some shortcomings in putting affixes in two strata. According to Fabb (1988), the second approach deals with the affixation selectional restrictions (morphological, phonological, and semantic). He also points to the restrictions on four classes of affixes. Fabb's (1988) Quadratic Classification also has serious shortcomings (Plag 1996, 1999). These include many counterexamples, the optionality of suffix classification, the vagueness of why a particular suffix should belong to a particular class, and the inability to predict all restrictions on possible or impossible combinations. Hay's psycholinguistic model proposes the Complexity-based Ordering approach (Hay 2000, 2002; Hay & Plag 2004; Hay & Baayen 2009). This model maintains that processing complexity can order affixes along a hierarchy through the degree of affixes' parsability. To determine the degree of parsability, we need to extract the token, type frequency of derived words and their bases, the token from the database, and type parsing ratio of derived words and their bases through some calculations (Hay & Plag 2004 & Baayen 2002). Affix ordering studies in the stratum-oriented approach and the selectional restrictions approach that only consider the type frequency of derived words have been investigated for Persian affixes, besides some shortcomings. For the first time, the present study explains affix ordering and affixation of the Persian prefixes based on the complexity-based ordering approach.

3. Methodology

Following Hay and Plag (2004), we have investigated the derived words of 17 Persian derivational prefixes to determine their parsability ranks and ratios. The prefixes under study are: «abar-, bāz, bar-, bol-, belā-, pād-, dar-, doʒ-, farā-, foru-, lā-, nā-, na-, vā-, var-, vāʒ-, ham-». For assigning the co-occurrence of prefixes regarding the attested combination, we first plotted the prefixes listed in alphabetical order in a square matrix. Using the R software (Gansner & et al.

1993), we drew our directed prefix graph according to the co-occurrence of the derivational prefixes. We use the upper triangular matrix to rearrange affixes for ranking them according to Plag and Baayen (2009); this ranking is called the degree of parsability. Then the degree of parsability ratio of each prefix was calculated using the ratio obtained by the type and token frequency of derived words and their bases. For this purpose, the derived words of 17 prefixes and their bases frequencies were extracted from the database and then entered into the Excel software. Following Hay and Baayen's (2001) tenor, the logarithmic scale measured derived words' frequency and bases. We used the scatter plot to investigate the frequency of derived words and their relation to the frequencies of their bases. Next, we illustrated the regression of base frequency on derived frequency by the least trimmed squares regression line called the parsing line by Hay and Baayen (2001). This line can divide the derived words into two categories: those that fall above the parsing line (the parsable derived words) and those below (the non-parsable derived words). By identifying the proportion of derived words falling above the trend line, we can distinguish two parsing ratios: type and token frequencies. The type parsing ratio can be estimated by how many proportions of the type frequency of the derived words fall above the parsing line, and the token parsing ratio is calculated by how many proportions of the token frequency of the derived words fall above the parsing line. Then, for each prefix, we compared the specified CO-rank or what is named as parsability ratio with the proportion of type and token parsing ratios. To understand the regression of token frequency on token parsing ratio and type parsing ratio, we used the scatter plots to investigate the psychological characteristics and mental processing in the Persian derived words. To obtain the type parsing ratio rank and the token parsing ratio rank, we first sorted them by the lowest to the highest degree in Excel and calculated their ranks. We used the type parsing ratio rank and the token parsing ratio rank to achieve the parsability continuum rank. The purpose of the continuum drawing is to show the order in which the Persian derivational prefixes are positioned, preceded, and delayed in attested and non-attested words.

4. Results

For obtaining the parsability rank of Persian derivational prefixes based on Selectional Restrictions and the co-occurrences of affixes, the present study used the square matrix, directed graph, and upper triangle square matrix. In this way, the ranking of all prefixes was not possible because some do not participate in co-occurrences with others. According to the Complexity-based-ordering approach, the regression of derived words is obtained by scatter plot for each affix. By dividing the proportion of type and tokens frequencies above the trend line by the total type and token frequencies of the derived word of a given prefix, we determined the type and token parsing ratio. Align with Hay and Baayen (2002, 2003) and Plag and Baayen (2009), we showed that the token frequency increases by decreasing the parsability ratio in the Persian derivational affixes and the less lexical processing. Eventually, to reach the parsability continuum, we used the CO-rank (parsability rank), the type parsing ratio rank, and the token parsing ratio rank of the given derived words of each prefix. Then, through the given ranks, the parsability continuum of Persian derivational prefixes was drawn. It showed that affixes on the left side are closer to bases, and affixes on the right side are farther from bases, and it can describe the Persian affixation. In line with the findings of Hay and Plag (2004) on English affix ordering, the Complexity-based-ordering approach can explain the ordering of Persian prefixes and affixation.

تقطیع پذیری و ترتیب پیشوندهای اشتقاقی زبان فارسی

مریم حمصیان اتفاق

دانشجوی دکتری زبان‌شناسی، دانشکده زبان‌های خارجی، دانشگاه اصفهان، اصفهان، ایران

عادل رفیعی^۱

استادیار زبان‌شناسی، دانشکده زبان‌های خارجی، دانشگاه اصفهان، اصفهان، ایران

بتول علی نژاد

دانشیار زبان‌شناسی، دانشکده زبان‌های خارجی، دانشگاه اصفهان، اصفهان، ایران

تاریخ دریافت مقاله: ۹۷/۰۸/۷؛ تاریخ پذیرش مقاله ۹۹/۰۶/۱۵

چکیده

ترتیب وندها از جمله مباحث نظری مطرح در حوزه صرف است. تاکنون بحث‌های زیادی در خصوص وجود اصول یا مکانیسم‌های کلی ناظر بر ترتیب وندها صورت گرفته است. در این راستا می‌توان به سه رویکرد اصلی اشاره کرد. اولین رویکرد با نام مدل لایه محور (سیگل، ۱۹۷۴؛ آلن، ۱۹۷۸؛ کیپارسکی، ۱۹۸۲؛ سلکرک، ۱۹۸۲؛ موهانان، ۱۹۸۶؛ گیگریچ، ۱۹۹۹) معتقد است واژگان دارای ساختار لایه‌های هستند و این ساختار تعیین‌کننده مشخصه‌های ترکیبی وندها است. دومین رویکرد متکی بر محدودیت‌های گزینشی وندمحور (فب، ۱۹۸۸؛ پلاگ، ۱۹۹۹) است. این محدودیت‌ها شامل محدودیت‌های گروهی وندها و نیز محدودیت‌های واجی، صرفی، معنایی و نحوی آنها است. ادعای این رویکرد این است که توصیف ترکیبات موجود و محتمل وندها در گرو توصیف عملکرد این محدودیت‌ها است. سومین رویکرد که چارچوب مقاله حاضر را تشکیل می‌دهد، با ارائه یک مدل روان‌شناختی- زبان‌شناختی مبتنی بر این ایده است که ترتیب وندها تابعی از میزان تقطیع‌پذیری وندها از پایه (های، ۲۰۰۰؛ ۲۰۰۲) است. بر طبق این رویکرد که پلاگ آن را رویکرد ترتیب مبتنی بر پیچیدگی نامید وندهایی که به راحتی در پردازش زبانی تقطیع می‌شوند نمی‌توانند در جایگاهی نزدیک‌تر به پایه نسبت به وندهای دارای میزان تقطیع‌پذیری کمتر قرار گیرند. سؤال پژوهش حاضر این است که چگونه می‌توان در این چارچوب، ترتیب پیشوندهای اشتقاقی زبان فارسی را توجیه کرد. بدین منظور هفده پیشوند انتخاب و تعداد ۲۲۰۰ واژه مشتق از پایگاه داده‌ها استخراج شد. با توجه به هم‌آیندی پیشوندها مرتبه هریک با استفاده از گراف جهت‌دار و ماتریس مربعی بالا مثلثی به دست داده شد. بر طبق چارچوب نظری اتخاذشده، با بسامد واژه‌های مشتق و پایه آنها نمودار پراکندگی ترسیم و مرتبه نسبت بسامد رخداد نوع و مرتبه نسبت بسامد نمونه تقطیع‌پذیر هر وند مشخص شد. در نهایت بر اساس داده‌ها و با استفاده از مرتبه تقطیع‌پذیری، مرتبه نسبت بسامد نمونه و مرتبه نسبت بسامد نوع تقطیع‌پذیر، پیوستار ترتیب پیشوندهای اشتقاقی زبان فارسی به دست داده شد. همسو با یافته‌های های و پلاگ (۲۰۰۴) مشخص شد که این رویکرد همراه با رویکرد محدودیت‌های گزینشی به‌خوبی می‌تواند ترتیب پیشوندهای اشتقاقی زبان فارسی را توجیه می‌کند.

واژه‌های کلیدی: ترتیب وندها، رویکرد ترتیب مبتنی بر پیچیدگی، پیشوندهای اشتقاقی زبان فارسی، مرتبه تقطیع‌پذیری، نسبت تقطیع‌پذیری

۱- مقدمه

مبحث ترتیب وندها از جمله مباحثی است که در دهه‌های اخیر مورد توجه زبان‌شناسان قرار گرفته است. تاکنون بحث‌های زیادی در خصوص وجود اصول یا مکانیسم‌هایی کلی ناظر بر ترتیب وندها صورت گرفته است. در این راستا می‌توان به سه رویکرد اصلی اشاره

کرد. اولین رویکرد با نام مدل لایه محور^۱ (سیگل، ۱۹۷۴؛ آلن، ۱۹۷۸؛ کیپارسکی، ۱۹۸۲؛ سلکرک، ۱۹۸۲؛ موهانان، ۱۹۸۶؛ گیگریچ، ۱۹۹۹) حاکی از آن است که واژگان دارای ساختار لایه‌ای هستند و این ساختار تعیین‌کننده مشخصه‌های ترکیبی وندها است. دومین رویکرد متکی بر محدودیت‌های گزینشی وند محور^۲ (فب، ۱۹۸۸؛ پلاگ، ۱۹۹۹) است. این محدودیت‌ها شامل محدودیت‌های گروهی وندها و نیز محدودیت‌های واجی، صرفی، معنایی و نحوی هر یک از وندها است. ادعای این رویکرد آن است که توصیف ترکیبات ممکن و غیرممکن وندها در گرو توصیف عملکرد این محدودیت‌ها است. سومین رویکرد با ارائه یک مدل روان‌شناختی- زبان‌شناختی مبتنی بر این ایده است که ترتیب وندها تابعی است از میزان تقطیع‌پذیری^۳ وندها از پایه (های، ۲۰۰۰؛ ۲۰۰۲). بر طبق این رویکرد وندهایی که به راحتی در پردازش زبانی تقطیع می‌شوند نمی‌توانند در جایگاهی نزدیک‌تر به پایه نسبت به وندهای دارای میزان تقطیع‌پذیری کمتر قرار گیرند. پلاگ (۲۰۰۳) این رویکرد را ترتیب وند مبتنی بر پیچیدگی^۴ نامیده است. جمعی از محققان همچون به این، پلاگ و همکاران (۲۰۱۵) با انجام طرح‌های تحقیقاتی در زبان‌های انگلیسی، آلمانی و مالت در خصوص رابطه عوامل دخیل در ترتیب وندها با استفاده از رویکرد ترتیب مبتنی بر پیچیدگی درصد رسیدن به نظریه جامع و کاملی در مبحث ترتیب وندافزایی به صورت بین زبانی هستند. پژوهش حاضر در پاسخگویی به اینکه تا چه میزان می‌توان ترتیب پیشوندهای اشتقاقی زبان فارسی را در چارچوب رویکرد ترتیب مبتنی بر پیچیدگی (باین و پلاگ، ۲۰۰۹) با توجه به دو معیار مرتبه تقطیع‌پذیری و نسبت تقطیع‌پذیری توجیه کرد به بررسی هفده پیشوند اشتقاقی زبان فارسی پرداخته است. برای این منظور ابتدا داده‌ها از پیکره داده‌های زبان فارسی پژوهشگاه علوم انسانی و مطالعات فرهنگی^(۱) استخراج و با توجه به هم‌آیندی وندها^۵ و ماتریس مربعی بالا مثلثی^۶ و گراف جهت‌دار^۷ مرتبه تقطیع‌پذیری^۸ پیشوندها به دست آمد. پس از تعیین بسامد رخداد واژه‌های مشتق و نیز پایه‌های آنها با قراردادن لگاریتم واژه‌های مشتق در محور افقی و لگاریتم پایه‌ها در محور عمودی برای هر

-
1. stratum-oriented model
 2. selectional restriction
 3. parsability
 4. Complexity-based ordering
 5. co-occurrence
 6. upper triangular square matrix
 7. directed graph
 8. rank

پیشوند نمودار پراکندگی^۱ در نرم‌افزار اکسل ترسیم و برای هر واژه مشتق با توجه به جایگاه آن و جایگاه پایه مربوطه نقطه‌ای بر روی نمودار ترسیم شده است. با توجه به میانگین نقاط قرار گرفته در هر یک از نمودارهای پراکندگی خط رگرسیون مربوطه ترسیم و با استفاده از معادله خط، نسبت بسامد نوع^۲ و نسبت بسامد نمونه^۳ تقطیع‌پذیری هر یک از پیشوندها محاسبه شده است. با توجه به مرتبه نسبت بسامد نوع و مرتبه نسبت بسامد نمونه تقطیع‌پذیر و مرتبه تقطیع‌پذیری، پیوستار ترتیب پیشوندهای اشتقاقی زبان فارسی به دست آمده است. نتایج نشان می‌دهند تقدم و تأخر پیشوندهای اشتقاقی زبان فارسی همراستا با جایگاه قرارگیری آنها در پیوستار^۴ تقطیع‌پذیری است و لذا ادعای رویکرد ترتیب مبتنی بر پیچیدگی در مورد ترتیب پیشوندهای اشتقاقی زبان فارسی تأیید می‌شود. مقاله حاضر در پنج بخش تدوین شده است. پس از مقدمه در بخش دوم به مروری بر پیشینه تحقیق و رویکردهای مطرح شده در خصوص ترتیب وندها پرداخته‌ایم. در بخش سوم به معرفی رویکرد ترتیب مبتنی بر پیچیدگی‌های و پلاگ (۲۰۰۲) و هم‌آیندی وندها می‌پردازیم. در بخش چهارم به بررسی پیشوندهای اشتقاقی زبان فارسی در چارچوب نظریه ترتیب مبتنی بر پیچیدگی‌های و پلاگ (۲۰۰۲) مرتبه تقطیع‌پذیری، نسبت تقطیع‌پذیری و همچنین پیوستار تقطیع‌پذیری پیشوندهای اشتقاقی زبان فارسی خواهیم پرداخت. بخش آخر نیز به نتیجه‌گیری و جمع‌بندی اختصاص داده شده است.

۲. پیشینه و مبانی تحقیق

شیوه ترکیب وندها ابتدا در چارچوب مدل لایه محور مورد توجه برخی از محققان قرار گرفت (پلاگ و باین، ۲۰۰۹). از جمله کسانی که در این چارچوب به بررسی ترتیب وندها پرداخته‌اند می‌توان به سیگل (۱۹۷۴)، آلن (۱۹۷۸)، کیپارسکی (۱۹۸۲) و موهانان (۱۹۸۶) اشاره کرد. بر اساس این رویکرد وندها اعم از پیشوندها و پسوندها در دو لایه قرار می‌گیرند. رویکرد لایه‌ای علی‌رغم توجه نسبی برخی ترکیب‌های وندی که قبلاً ممکن نبود، دارای نارسایی‌هایی نیز است. از جمله این نارسایی‌ها می‌توان به ناتوانی این رویکرد در توجیه ترکیبات ممکن و غیرممکن وندهای حاضر در یک لایه معین، نداشتن دلیل واضح مبنی بر قرار گرفتن وندها در دو لایه مختلف و صرفاً اتکا به معیار ریشه‌شناختی و مشخصه‌های واجی جهت قراردادن وندها در دو لایه اشاره کرد (پلاگ و باین، ۲۰۰۹). در زبان فارسی نیز با بررسی‌های انجام‌شده بر طبق این مدل وندها اکثراً در لایه اول قرار گرفته‌اند و تکیه را به سمت خود کشیده‌اند. در واقع اکثر وندهای زبان

-
1. scatter plot
 2. relative type frequency
 3. relative token frequency
 4. continuum

فارسی تکیه‌بر هستند (غلامعلی زاده، ۱۳۹۰). با توجه به این مطلب مشخص می‌گردد که رویکرد لایه‌ای به آن شکلی که در زبان انگلیسی با معیار تغییرات آوایی به لایه‌بندی وندها می‌پردازد، در زبان فارسی کارآیی چندانی ندارد. دومین رویکرد ناظر بر ترتیب قرارگیری وندها با مطرح ساختن محدودیت‌های گزینشی مانند محدودیت‌های صرفی، واجی و معنایی به بررسی ترکیبات اشتقاقی ممکن و غیرممکن می‌پردازد (فب، ۱۹۸۸). به اعتقاد فب (۱۹۸۸) ترتیب قرارگیری پسوندهای اشتقاقی زبان انگلیسی بر پایه محدودیت‌های گزینشی هر وند است. وی بر اساس معیار هم‌آیندی، پسوندهای زبان انگلیسی را در چهار گروه قرار می‌دهد.

گروه اول: پسوندی‌هایی که به کلمه دارای پسوند متصل نمی‌شوند.

گروه دوم: پسوندی‌هایی که بعد از پسوندهای خاص دیگری می‌آیند.

گروه سوم: پسوندی‌هایی که به راحتی به هر پسوندی متصل می‌شوند.

گروه چهارم: پسوندهای مشکل‌زا که در ترکیبات یا تنها به پایه متصل می‌شوند و یا به پسوندهای خاص دیگری متصل می‌شوند.

در زبان فارسی با اتخاذ رویکرد محدودیت‌های گزینشی حمصیان (۱۳۸۹) با بررسی هفده پیشوند و ۳۷ پسوند به بررسی وندهای اشتقاقی زبان فارسی پرداخته است. وی با بررسی ساختارهای مشتق حاصل از این وندها الگوهای ساختاری واژه‌های مشتق زبان فارسی را با ساختار الگوی PRS و در یازده ساختار قرار داده است. در ادامه این پژوهش از چهار دسته‌ای که فب پسوندها را در آن قرار داده است وندهای زبان فارسی اعم از پیشوندها و پسوندهای اشتقاقی تنها در دو دسته قرار گرفته‌اند. این دو دسته عبارت‌اند از پسوندها و پیشوندی‌هایی که فقط به پایه متصل می‌شوند و پسوندها و پیشوندهای مشکل‌زا که هم به پایه و هم به پیشوند و پسوند خاص متصل می‌شوند. مواردی که در دسته‌بندی‌های دوم و سوم فب اعم از آن دسته از وندهایی که فقط به وند خاص متصل می‌شوند و آن دسته از وندهایی که به طور آزادانه به هر وندی متصل می‌شوند در زبان فارسی مشاهده نشده است. از نظر پلاگ (۱۹۹۶؛ ۱۹۹۹) طبقه‌بندی چهارگانه فب (۱۹۸۸) دارای ضعف‌هایی چون وجود مثال‌های ضدونقیض زیادی مبنی برقرار دادن وندها در لایه‌های این طبقه‌بندی است. دوم اینکه طبقه‌بندی وندها اختیاری است و مشخص نیست چرا یک پسوند معین باید متعلق به یک طبقه خاص باشد. سوم اینکه این طبقه‌بندی برای همه محدودیت‌های حاکم بر ترکیبات ممکن و یا ترکیبات غیرممکن به کار نمی‌رود. پلاگ به جای قائل شدن به گروه‌بندی وندها بر این باور است که هر وند معین در مدخل واژگانی خود دارای مشخصه‌های واجی، صرفی، معنایی، نحوی و نیز ویژگی‌های الحاقی خاص خود است که ترکیبات ممکن و غیرممکن یک وند با پایه و دیگر وندها را رقم می‌زند. وی علاوه بر محدودیت‌های وندمحور پسوندافزایی

پایه‌محور را برای ترکیبات پسوندی مطرح می‌سازد. اما پسوندافزایی پایه‌محور نیز در بعضی موارد به خصوص در مورد پسوندهایی که در دومین گروه فب قرار می‌گیرند قادر به توضیح چرایی این ترکیبات نخواهد بود. چون در این موارد این پسوندها هستند که ترکیبات پسوندافزایی را مشخص می‌سازد نه پایه‌ها. رویکرد لایه محور و رویکرد محدودیت‌های گزینشی مطرح‌شده تا بدین جا تنها به بررسی ساختار ترتیب وندها پرداخته و صرفاً بسامد نوع واژه‌های مشتق را مدنظر قرار داده‌اند. از این رو های (۲۰۰۰) سومین رویکرد وندافزایی را با عنوان ترتیب مبتنی بر پیچیدگی و با مطرح ساختن مسئله روان‌شناختی زبانی بر طبق نموده‌های ذهنی و واژگانی ارائه کرده است (های، ۲۰۰۰؛ ۲۰۰۲، های و پلاگ، ۲۰۰۴؛ پلاگ و باین، ۲۰۰۹). در واقع زبان‌شناسان بر این باورند که به طور کلی زبان و بالاخص صرف اشتقاقی انباشتی از مشخصه‌های منحصر به فرد زبانی نیست بلکه می‌توان قوانین کلی برای تعمیم دادن آنها به تمامی زبان‌ها را مطرح ساخت. بر این اساس در سومین رویکرد، ترتیب وندها بر اساس میزان تقطیع‌پذیری وند از پایه در راستای پیوستار تقطیع‌پذیری صورت می‌گیرد. به طوری که در یک سوی پیوستار تقطیع‌پذیرترین وندها و در سوی دیگر تقطیع‌ناپذیرترین وندها قرار می‌گیرند. برای به دست آورد میزان تقطیع‌پذیری وندها می‌توان علاوه بر بسامد نوع واژه‌های مشتق از بسامد نمونه آنها نیز استفاده کرد. در واقع در این رویکرد پردازش واژگانی وابسته به میزان بسامد نمونه و بسامد نوع واژه‌های مشتق از یک وند و نسبت تقطیع‌پذیری بسامد نوع و بسامد نمونه است (های و پلاگ، ۲۰۰۴ و باین، ۲۰۰۲). های و پلاگ (۲۰۰۴) با انتخاب ۱۵ پسوند زبان انگلیسی و ترکیب کردن آنها بدون در نظر گرفتن هیچ یک از محدودیت‌های گزینشی قائل به وجود ۲۱۰ ترکیب دو پسوندی شده‌اند. آنها با استفاده از منابعی چون BNC، CELEX^(۷) و OED و اینترنت به این مطلب پرداخته‌اند که آیا وجود ترکیبات موجود و ممکن در هر یک از این منابع را می‌توان بر اساس رویکرد ترتیب مبتنی بر پیچیدگی توصیف کرد یا آنکه رویکرد محدودیت‌های گزینشی تحلیل مناسب‌تری از وجود یا عدم وجود هریک از ترکیبات دو پسوندی را ارائه خواهد کرد؟ فرضیه‌های و پلاگ (۲۰۰۴) بر این امر استوار است که محدودیت‌های گزینشی و محدودیت‌های پردازشی هر دو حاکم بر ترکیبات موجود و ممکن پسوندهای اشتقاقی بوده و مکمل یکدیگر می‌باشند. زیرکل (۲۰۱۰) در مقاله‌ای با عنوان ترکیبات پیشوندی در زبان انگلیسی: عوامل ساختاری و پردازشی، با بررسی پیشوندهای اشتقاقی زبان انگلیسی درصدد یافتن قواعدی حاکم بر ترتیب قرارگیری آنها است. وی نظر به اینکه ترتیب قرارگیری پسوندها از نظرهای و پلاگ (۲۰۰۴) و پلاگ و باین (۲۰۰۹) تحت حاکمیت محدودیت‌های گزینشی و عوامل ساختاری و پردازشی است به بررسی پیشوندها پرداخته است. برای این منظور با انتخاب ۱۵ پیشوند زبان انگلیسی که در الگوی ساختار PPR قرار گرفته از منابع CELEX، BNC، OED

و COCA و اینترنت واژه‌های مشتق را انتخاب کرده است. از نظر وی با وجود ۱۵ پیشوند می‌توان انتظار داشت که ۲۲۵ ترکیب دو پیشوندی وجود داشته باشد. وی ابتدا در ماتریس مربعی پیشوندها را به ترتیب حروف الفبا در ستون عمودی و در ردیف افقی بالای جدول قرار داده و تمامی ترکیبات ممکن بالفعل را با عدد یک و بقیه را با عدد صفر مشخص کرده است. از ۲۲۵ ترکیب ممکن ۷۹ ترکیب از منابع مورد بررسی یافت شد که وی نقش محدودیت‌های گزینشی وندها و پایه‌ها را دلیل کاهش تعداد ترکیبات پیشوندی می‌داند. از نظر وی در هم‌آیندی پیشوندها به مانند پسوندها محدودیت‌های آوایی، صرفی، نحوی و معنایی وجود دارد اما شدت این محدودیت‌ها کمتر و از انعطاف-پذیری بیشتری برخوردار است. به عنوان مثال برخلاف پسوند -en یا -th پیشوندها چون حاوی تکیه می‌باشند بیشتر شبیه واژه بوده و در اتصال به پایه بدون در نظر گرفتن هم‌آیندی با دیگر پیشوندها ناقض محدودیت‌های آوایی نمی‌شوند. لذا محدودیت‌های آوایی برخلاف ترکیب پسوندی تأثیری بر ترکیبات پیشوندی ندارند. همچنین محدودیت‌های معنایی تأثیر چندانی بر پیشوندافزایی ندارند. پیشوندها معمولاً متعلق به طبقه خاص نحوی نیستند و به همه گروه‌های نحوی متصل می‌شوند، مانند: پیشوند dis- که در ترکیبات نحوی (disconnect (verb), disadvantage (noun), dissimilar (adj) کار می‌رود. در نهایت وی این‌گونه نتیجه‌گیری می‌کند که چون پیشوندافزایی نسبت به پسوندافزایی کمتر تابع محدودیت‌های گزینشی است، پس از نظر دستوری نیز محدودیت‌های کمتری بر پیشوندافزایی حاکم است. از این رو تعداد ترکیبات دو پیشوندی در این منابع مورد مطالعه بیشتر از تعداد ترکیبات دو پسوندی است. نسبتی که وی در این مقاله از پیشوندافزایی به دست آورده است حدود ۲۴٪ است، که این میزان دو برابر نسبتی است که های و پلاگ (۲۰۰۴) از ترکیبات ثبتي دو پسوندی به دست آورده‌اند. با توجه به اینکه مقاله حاضر به بررسی پیشوندهای اشتقاقی زبان فارسی در چارچوب رویکرد ترتیب مبتنی بر پیچیدگی پرداخته، لازم است این رویکرد را با توجه به معیارهای مطرح در آن به تفصیل معرفی و سپس به توصیف پیشوندهای اشتقاقی زبان فارسی در چارچوب آن بپردازد.

۳. رویکرد ترتیب مبتنی بر پیچیدگی

مهم‌ترین اصل زیربنایی رویکرد ترتیب مبتنی بر پیچیدگی این است که ترتیب وندها - در کنار معیار محدودیت‌های گزینشی حاکم بر وندها - تابعی از میزان تقطیع‌پذیری وند از پایه است. در واقع این رویکرد بر اساس پردازش واژگانی در درک گفتار بنا نهاده شده است. در خصوص پردازش واژگانی دو راهکار صرفی وجود دارد: اینکه در هنگام مواجهه با یک واژه مشتق‌گوشور آن را به صورت یک واحد در نظر بگیرد یا اینکه آن را

همان‌طور که در شکل ۱ نشان داده شده بیشترین تعداد یال‌هایی که وندها را به یکدیگر متصل می‌کنند جهت رو به پایین دارند. همان‌طور که مشاهده می‌شود نرم‌افزار به‌گونه‌ای چیدمان پسوندها را قرار داده که کمترین تعداد یال جهت رو به بالا که نشان از موارد استثناء و هم‌آیندی‌های دو طرفه (وندهایی که هم قبل و هم بعد از یکدیگر می‌توانند قرار گیرند) هستند را نشان دهد. اینک برای مشخص شدن چگونگی قرار گرفتن پسوندها نسبت به یکدیگر با توجه به گراف به دست آمده از نرم‌افزار آر، ماتریس‌های مربعی بالا مثلثی را که نشان از چیدمان گراف است، ارائه داده می‌شود. با توجه به ۳۱ پسوند اشتقاقی زبان انگلیسی و با توجه به ترتیب‌های مختلف که این پسوندها می‌توانند نسبت به یکدیگر به‌گونه‌ای داشته باشند که کمترین تعداد یال جهت رو به بالا را نشان دهد، تعداد ۴۰۳۷ ماتریس مربعی بالا مثلثی به دست داده می‌شود. جدول ۱ یکی از ماتریس‌های مربعی بالا مثلثی ۳۱ پسوند اشتقاقی زبان انگلیسی را نشان می‌دهد.

	en	ment	ary	or	ster	ian	er	elle	ive	ary	ist	age	ess	erly	ly	ing	ship	hood	dom	less	fold	ful	ous	ism	ness	ly	ful	wise	
th	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
en	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ment	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ary	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
or	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ster	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ian	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
er	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
elle	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ive	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ary	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ist	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
age	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ess	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
erly	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
ly	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
ing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ship	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
hood	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
dom	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
less	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fold	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ful	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ism	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ness	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ly	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ful	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wise	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

جدول ۱. ماتریس مربعی بالا مثلثی هم‌آیندی ۳۱ پسوند اشتقاقی زبان انگلیسی

این دست از ماتریس‌ها نشان می‌دهند که میزان هم‌آیندی وندها و ایجاد ترکیبات ممکن، قطعاً بر پایه محدودیت‌های حاکم بر وندها به وجود می‌آید. های و پلاگ با استفاده از این ۴۰۳۷ ماتریس بالا مثلثی به دست آمده از گراف جهت‌دار و با میانگین گرفتن مرتبه قرارگیری هر وند در این ماتریس‌ها توانستند مرتبه تقطیع‌پذیری هر یک از وندهای مورد مطالعه در این بررسی را مشخص کنند. با استفاده از مرتبه تقطیع‌پذیری که از هم‌آیندی وندها در ماتریس مربعی به دست می‌آید می‌توان ترتیب قرارگیری وندها نسبت به پایه و نسبت به یکدیگر را مورد بررسی قرار داد. بر اساس نظر های و پلاگ (۲۰۰۴) هرچه وندهای مرتبه تقطیع‌پذیری بالاتری داشته باشد از پایه

دورتر است و از میزان تقطیع‌شدگی بالاتری برخوردار است و در مقابل هرچه مرتبه تقطیع‌پذیری یک وند کمتر باشد به پایه نزدیک‌تر بوده و در پردازش واژگانی کمتر از پایه خود جدا می‌شود. جدول ۲ مرتبه تقطیع‌پذیری ۳۱ پسوند اشتقاقی زبان انگلیسی را نشان می‌دهد.

جدول ۲. مرتبه تقطیع‌پذیری ۳۱ پسوند اشتقاقی زبان انگلیسی

	SUFFIX	RANK		SUFFIX	RANK		SUFFIX	RANK		SUFFIX	RANK
1	th	0.89	9	ette	8.62	17	lyAJ	14.58	25	fulA	21.87
2	en	1.78	10	aryA	9.27	18	ery	16.19	26	fold	21.87
3	ment	2.67	11	ive	9.27	19	ling	16.27	27	wise	25.80
4	or	4.40	12	ist	11.44	20	ship	18.63	28	lyAV	25.83
5	ster	4.42	13	ce	11.54	21	dom	18.71	29	fulN	25.86
6	aryN	4.55	14	ish	12.70	22	hood	18.81	30	ism	25.86
7	ian	6.24	15	ess	13.17	23	less	21.78	31	ness	25.88
8	er	7.13	16	age	14.21	24	ous	21.83			

نکته حائز اهمیت در مرتبه تقطیع‌پذیری که در جدول ۲ آمده این است که برخی از پسوندها در هم‌آیندی با دیگر پسوندها به کار نرفته و امکان مرتبه‌دهی به آنها وجود ندارد. از این رو‌های (۲۰۰۱) برای تعیین مرتبه این‌گونه وندها از نسبت تقطیع‌پذیری استفاده می‌کند. های و پلاگ (۲۰۰۴) علاوه بر مرتبه تقطیع‌پذیری که از هم‌آیندی وندها حاصل می‌شود و لذا در بعضی موارد قادر به مرتبه‌دهی نیست برای تعیین ترتیب وندها به بررسی معیار نسبت تقطیع‌پذیری هر وند نیز پرداختند تا در نهایت با توجه به هر دو معیار بتوانند پیوستار تقطیع‌پذیری را به دست دهند. در ادامه ابتدا نحوه محاسبه نسبت تقطیع‌پذیری بیان شده و سپس نحوه شکل‌گیری پیوستار تقطیع‌پذیری بر اساس دو معیار مرتبه تقطیع‌پذیر و نسبت تقطیع‌پذیر مشخص می‌شود. های (۲۰۰۱) برای به دست آوردن نسبت تقطیع‌پذیری وندهای مورد بررسی از بسامد نوع و بسامد نمونه واژه‌های مشتق از هریک از وندها و پایه‌های آنها استفاده کرده است. نسبت تقطیع‌پذیری به میزان کاربرد و واژه‌سازی هر وند اشاره می‌کند. همان‌طور که های در این رویکرد اشاره می‌کند برخلاف دیگر رویکردها که صرفاً با استفاده از بسامد نوع واژه‌های مشتق به بررسی ترتیب وندها پرداخته‌اند در این رویکرد علاوه بر بسامد نوع از بسامد نمونه واژه‌های مشتق نیز بهره گرفته شده است. بنا به نظر وی هرچه بسامد (اعم از بسامد نوع و بسامد نمونه) واژه مشتق نسبت به پایه آن بیشتر باشد پردازش صرفی سازه‌محور کمتر و آن واژه مشتق به صورت یک واژه کل در نظر گرفته می‌شود. در مقابل هرچه بسامد واژه مشتق نسبت به پایه کمتر باشد پردازش صرفی سازه‌محور بیشتر و بالطبع تقطیع‌پذیری بیشتر است (های، ۲۰۰۱). این رویکرد برای ایجاد رابطه میان بسامد رخداد واژه مشتق و پایه از نمودار پراکندگی و قرار دادن واژه‌های مشتق بر روی محور افقی و همچنین پایه‌های تشکیل‌دهنده آنها در محور

عمودی و محاسبه میانگین روند آنها (رگرسیون حداقل مربعات معمولی^۱) برای هر وند استفاده کرده است. های و پلاگ (۲۰۰۴) میانگین روند داده‌ها در نمودار پراکندگی را که در این نمودارها با عنوان رگرسیون حداقل مربعات معمولی شناخته می‌شود به عنوان خط تقطیع‌پذیری در نظر گرفته‌اند. با استفاده از خط تقطیع‌پذیری می‌توان واژه‌های مشتق حاصل از یک وند را به دو دسته تقسیم کرد: واژه‌هایی که در بالای خط قرار گرفته‌اند، واژه‌هایی با بسامد رخداد پایه بیشتر هستند به طوری که برای درک آنها ابتدا واژه تقطیع شده و سپس معنای آن درک می‌شود و واژه‌هایی که در پایین خط قرار گرفته‌اند، واژه‌هایی با بسامد رخداد پایه کمتری بوده و به صورت یک واژه کل درک می‌شوند. در واقع با دو دسته واژه که در این رویکرد آنها را واژه‌های تقطیع‌پذیر و تقطیع‌شدگی می‌نامیم مواجه خواهیم شد که می‌توان نسبتی برای این واژه‌ها در نظر گرفت و آن را نسبت تقطیع‌پذیری نامید (های و باین، ۲۰۰۲). نسبت تقطیع‌پذیری وندها با دو مقوله بسامد نوع و بسامد نمونه واژه‌های مشتق از هر وند در ارتباط است (های و باین، ۲۰۰۲). برای محاسبه نسبت تقطیع‌پذیری لازم است نسبت بسامد نوع و نسبت بسامد نمونه واژه‌های مشتق بالای خط تقطیع‌پذیر (خط روند نمودار پراکندگی هر یک از وندها) به بسامد نوع و بسامد نمونه کل واژه‌های در بردارنده وند مورد نظر در پیکره محاسبه گردد. این نسبت مشخص می‌سازد که شنونده برای درک، دریافت و پردازش واژه مشتق از یک وند تا چه میزان به تقطیع‌پذیری و تفکیک اجزای سازنده آن واژه می‌پردازد (های و باین، ۲۰۰۳). های و پلاگ (۲۰۰۴) با توجه به نسبت تقطیع‌پذیری بسامد نوع و نسبت تقطیع‌پذیری بسامد نمونه حاصل از واژه‌های مشتق از هر وند و مقایسه آن با دیگر وندها توانستند برای هر یک از وندها مرتبه نسبت بسامد نوع تقطیع‌پذیر و مرتبه نسبت بسامد نمونه تقطیع‌پذیر را تعیین کنند. در ادامه آنها با توجه به مرتبه تقطیع‌پذیری، مرتبه نسبت بسامد نوع و مرتبه نسبت بسامد نمونه و میانگین گرفتن از آنها و قراردادن آن در محور عمودی نمودار پراکندگی و مرتبه تقطیع‌پذیری در محور افقی نمودار توانستند به پیوستار تقطیع‌پذیری پسوندهای اشتقاقی زبان انگلیسی برسند. این پژوهش در راستای پاسخگویی به این مسائل است که تا چه میزان در زبان فارسی می‌توان با توجه به معیارهای مطرح‌شده در رویکرد ترتیب مبتنی بر پیچیدگی اعم از تقطیع‌پذیری، مرتبه تقطیع‌پذیری و نسبت تقطیع‌پذیری ترتیب پیشوندهای اشتقاقی زبان فارسی را توجیه کرد و کارایی این رویکرد را محک زد. در ادامه به بررسی داده‌های زبان فارسی در چارچوب رویکرد ترتیب مبتنی بر پیچیدگی می‌پردازیم.

1. least trimmed squares regression

۴. تقطیع‌پذیری پیشوندهای اشتقاقی زبان فارسی

این مقاله به بررسی ترتیب پیشوندهای اشتقاقی زبان فارسی در چارچوب رویکرد ترتیب مبتنی بر پیچیدگی پرداخته تا مشخص شود این رویکرد تا چه میزان می‌تواند توجیه‌کننده ترتیب‌های ممکن و موجود پیشوندهای اشتقاقی زبان فارسی باشد. با توجه به موارد مطرح شده در بخش رویکرد ترتیب مبتنی بر پیچیدگی، برای تعیین مرتبه تقطیع‌پذیری و نسبت تقطیع‌پذیری به پیروی از های و پلاگ (۲۰۰۴) با استفاده از هم‌آیندی پیشوندها و میزان تقطیع‌پذیری وند از پایه به بررسی پیشوندهای اشتقاقی زبان فارسی پرداخته‌ایم. برای این منظور هفده پیشوند از کتاب *ساخت اشتقاقی واژه در فارسی/امروز (کلباسی، ۱۳۸۰)* انتخاب و واژه‌های مشتق از این هفده پیشوند اشتقاقی از پیکره پایگاه داده‌های زبان فارسی پژوهشگاه علوم انسانی و مطالعات فرهنگی^(۳) استخراج شده است^(۴). این پیشوندها عبارت‌اند از: «بر-، باز-، بر-، بل-، بلا-، پاد-، در-، دژ-، فرا-، فرو-، لا-، نا-، ن-، وا-، ور-، واژ-، هم-» پس از تعیین پیشوندهای مورد بررسی، در ابتدا هم‌آیندی پیشوندها را در قالب ماتریس مربعی تنظیم‌شده بر اساس حروف الفبا ترسیم کردیم. در این ماتریس اولین ردیف جدول، پیشوندهای جایگاه اول نسبت به پایه و اولین ستون پیشوندهای جایگاه دوم نسبت به پایه را نشان می‌دهد. همان‌طور که در ماتریس نشان داده شده است در صورت وجود هم‌آیندی پیشوندها در محل تلاقی آنها در ماتریس بسامد نوع هم‌آیندی آنها آورده شده است. لازم به ذکر است که برای مشخص کردن هم‌آیندی یا عدم هم‌آیندی پیشوندها، از ساختارهای PPR و PPRS حمصیان (۱۳۸۹) که ترتیب وندها را با رویکرد محدودیت‌های گزینشی توصیف کرده است، استفاده شد. جدول ۳ ماتریس مربعی هم‌آیندی پیشوندهای اشتقاقی زبان فارسی و بسامد نوع آنها را بر اساس حروف الفبا نشان می‌دهد.

	بر-	باز-	بر-	بل-	بلا-	پاد-	در-	دژ-	فرا-	فرو-	لا-	نا-	ن-	وا-	ور-	واژ-	هم-	جمع
بر-																		
باز-													۸					۸
بر-													۳۵				۲	۳۷
بل-																		
بلا-																		
پاد-																		
در-													۱۱				۳	۱۴
دژ-																		
فرا-																	۱	۱
فرو-																		
لا-																		
نا-																	۸	۸
ن-																		
وا-																		
ور-																	۷	۷
واژ-																		
هم-																		
جمع													۶۱				۱۴	۷۵

جدول ۳. ماتریس مربعی هم‌آیندی پیشوندهای اشتقاقی زبان فارسی و بسامد نوع ترکیبات آنها

آنها می‌توان چندین ماتریس مربعی بالا مثلثی به دست آورد به طوری که در این ماتریس‌ها در زیر خط قطر اصلی هیچ‌گونه هم‌آیندی پیشوندی وجود نداشته باشد.

جدول ۴. هم‌آیندی پیشوندهای اشتقاقی زبان فارسی بر اساس رویکرد ترتیب مبتنی بر پیچیدگی

	در-	یر-	ور-	نا-	قرا-	وا-	قرو-	یاز-	ایر-	یلا-	دژ-	یل-	یاد-	لا-	ن-	واژ-	هم-		
۱																			
۱	۱																		
۱		۱																	
۱			۱																
۱				۱															
۱					۱														
۱						۱													
۱							۱												
۱								۱											
۱									۱										
۱										۱									
۱											۱								
۱												۱							
۱													۱						
۱														۱					
۱															۱				
۱																۱			
۱																	۱		
۱																		۱	
۱																			۱

با توجه به گراف پیشوندهای اشتقاقی زبان فارسی و با استفاده از نرم‌افزار آر، تعداد ۲۷۰ ماتریس مربعی بالا مثلثی به دست آمده است. جدول ۴ یکی از این ۲۷۰ ماتریس مربعی بالا مثلثی را نشان می‌دهد. در این ماتریس عدد ۱ نشان‌دهنده هم‌آیندی پیشوندها است. بر اساس پلاگ و باین (۲۰۰۴) برای تعیین مرتبه تقطیع‌پذیری هر پیشوند با جمع کردن جایگاه قرارگیری هر یک از این پیشوندها در این ۲۷۰ ماتریس و تقسیم آن به این تعداد ماتریس می‌توان مرتبه تقطیع‌پذیری پیشوندهای اشتقاقی زبان فارسی را به دست داد. در ادامه به هر پیشوند یک مرتبه عددی می‌دهیم چون بر اساس رویکرد ترتیب مبتنی بر پیچیدگی جایگاه یک وند در سلسله‌مراتب قرارگیری وندها بیانگر میزانی است که آن وند می‌تواند مستقل از پایه خود پردازش شود (پلاگ و باین، ۲۰۰۹). در جدول ۵ مرتبه تقطیع‌پذیری پیشوندهای اشتقاقی زبان فارسی آورده شده است.

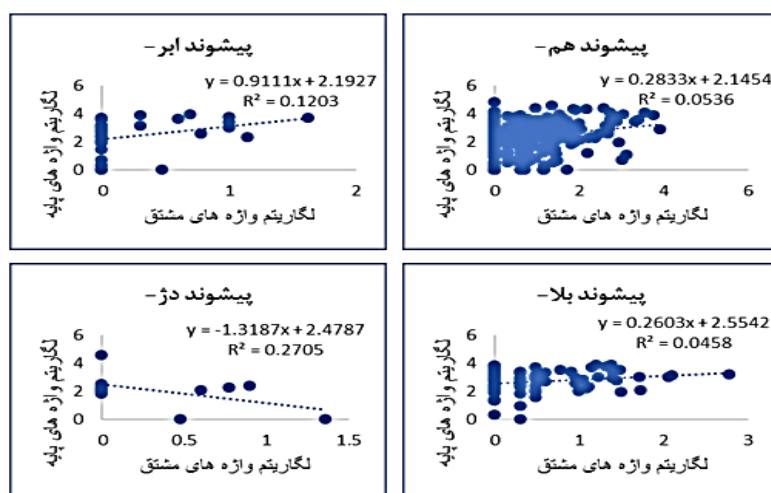
جدول ۵. مرتبه تقطیع‌پذیری هفده پیشوند اشتقاقی زبان فارسی

مرتبه	پیشوند اشتقاقی	مرتبه	پیشوند اشتقاقی	مرتبه	پیشوند اشتقاقی
۱	در-	۱	قرو-	۷	۱۲/۰۸
۲	یر-	۲	یاز-	۸	۱۲/۰۹
۳	ور-	۳	ایر-	۹	۱۲/۰۶
۴	نا-	۴	یلا-	۱۰	۱۲/۶۷
۵	قرا-	۵	دژ-	۱۱	۱۲/۸۶
۶	وا-	۶	یل-	۱۲	۱۱/۷۱

همان‌طور که در گراف جهت‌دار مشاهده شد تعدادی از پیشوندهای اشتقاقی اعم از پیشوندهای «واژ-»، «دژ-»، «بلا-»، «لا-»، «یل-» و «ایر-» در هم‌آیندی با دیگر وندها به کار نرفته‌اند در نتیجه بر اساس رویکرد ترتیب مبتنی بر پیچیدگی برای ترسیم

پیوستار تقطیع‌پذیری نمی‌توان جایگاه دقیقی نسبت به دیگر پیشوندها برای آنها بر روی پیوستار قائل شد. لذا در اینجا لازم است از نسبت تقطیع‌پذیری که از مشخصه‌های هر وند و بسامد نوع و بسامد نمونه واژه‌های ساخته‌شده با آن به دست می‌آید، مرتبه نسبت تقطیع‌پذیری را برای هریک از پیشوندهای اشتقاقی زبان فارسی به دست آورد. برای به محاسبه مرتبه نسبت تقطیع‌پذیری واژه‌های مشتق از هفده پیشوند مورد بررسی، پایه‌های سازنده همراه با بسامد وقوع واژه‌ها از پیکره استخراج و در نرم‌افزار اکسل آورده شد. به پیروی از های و باین (۲۰۰۱) برای بسامد رخداد واژه‌های مشتق و بسامد رخداد پایه‌های تشکیل‌دهنده آنها از مقیاس لگاریتم مبنای ده استفاده شد. دلیل استفاده از مقیاس لگاریتمی مبنای ده این است که چون میان بسامد وقوع پایه‌ها و بسامد وقوع واژه‌های مشتق حاصل از آنها در بسیاری از موارد فاصله عددی وجود داشته و قابل مقایسه نیست لذا با گرفتن لگاریتم می‌توان این اعداد را به یکدیگر نزدیک‌تر کرده و مقایسه دقیق‌تری بین آنها انجام داد. به عنوان نمونه بسامد رخداد واژه «برآشفته» ۳۲ و بسامد رخداد واژه «آشفته» ۳۰۲ است. در این مورد اختلاف عددی این دو نمونه ۲۷۰ است لذا با گرفتن لگاریتم می‌توان به اعداد ۱/۵ و ۲/۴۸ رسید. در این صورت اختلاف عددی این دو واژه برای مقایسه، ۰/۹۸ خواهد بود. به پیروی از های و باین (۲۰۰۳) برای بررسی میزان بسامد واژه‌های مشتق حاصل از یک وند و ارتباط آن با بسامد پایه‌های همان وند از نمودار پراکندگی (شکل ۳) استفاده کردیم. برای ترسیم این نمودار در نرم‌افزار اکسل لگاریتم بسامد وقوع واژه‌های مشتق از وند مورد بررسی را در محور افقی و لگاریتم بسامد وقوع پایه‌های سازنده آنها را در محور عمودی قرار دادیم. هر یک از نقاط ایجادشده در نمودار پراکندگی با توجه به مختصات آن نسبت به محور افقی (بسامد واژه مشتق) و محور عمودی (بسامد پایه) بیانگر یک واژه مشتق حاصل از پیشوند مورد بررسی است. در ادامه با استفاده از نرم‌افزار اکسل خط رگرسیون حداقل مربعات معمولی (خط تقطیع‌پذیری) که رابطه واژه‌های مشتق و پایه‌های آنها را با توجه به میانگین نقاط موجود در فضای بین دو محور نشان می‌دهد ترسیم کردیم. این خط که در نمودارهای شکل ۳ به صورت نقطه‌چین نشان داده‌شده واژه‌ها را به دو دسته؛ واژه‌های بالای خط و واژه‌های پایین خط تقسیم کرده است. به پیروی از های و باین (۲۰۰۳) داده‌های قرارگرفته در بالای این خط را واژه‌های مشتق تقطیع‌پذیر و داده‌های قرارگرفته در پایین این خط را واژه‌های مشتق تقطیع‌ناپذیر نامیده‌ایم. به عنوان مثال در خصوص پیشوندهای «ابر-»، «هم-»، «بلا-» و «دژ-» برطبق نمودار پراکندگی آنها واژه‌های «برقدرت»، «هم‌انگون»، «بلافاصله»، «دژخیم» که در زیر خط تقطیع‌پذیری قرارگرفته‌اند واژه‌های تقطیع‌ناپذیر و واژه‌های «برانسان»، «هم‌آوازی»، «بلا‌حاصل»،

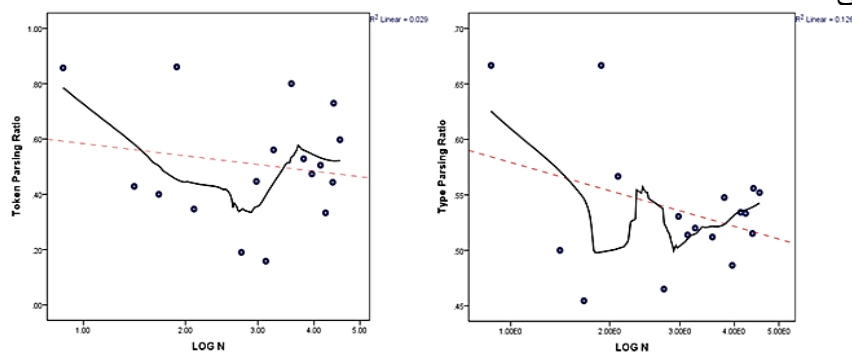
«دژسازگاری» که در بالای خط تقطیع‌پذیری قرار گرفته‌اند واژه‌های تقطیع‌پذیر در نظر گرفته می‌شوند. شکل ۳ به عنوان نمونه نمودار پراکندگی چهار پیشوند اشتقاقی زبان فارسی را نشان می‌دهد. طبق این نمودارها روند رابطه بسامد واژه‌های مشتق و بسامد پایه‌های آنها برای هر یک از پیشوندها متفاوت است. به عنوان مثال این رابطه برای پیشوندهای «هم-»، «ابر-» و «بلا-» رابطه‌ای مستقیم بوده و با افزایش بسامد واژه مشتق، بسامد پایه نیز به همان نسبت افزایش یافته است لذا، خط تقطیع‌پذیر رو به بالا قرار گرفته است. اما در مورد پیشوند «دژ-» همان‌طور که نمودار نشان می‌دهد این رابطه مستقیم نیست به طوری که با افزایش لگاریتم واژه‌های مشتق لگاریتم پایه‌های سازنده آنها افزایش پیدا نکرده و خط تقطیع‌پذیری دارای روندی نزولی و رو به پایین قرار گرفته است.



شکل ۳. نمودار پراکندگی پیشوندهای اشتقاقی «ابر-»، «هم-»، «بلا-»، «دژ-»

این وضعیت به‌خوبی نشان می‌دهد نه تنها میزان واژه‌سازی در زبان فارسی معیار با پیشوند «دژ-» رو به نزول است بلکه بیشتر واژه‌های ساخته‌شده با این پیشوند نیز به صورت یک واژه کل درآمده‌اند و شنونده به هنگام دریافت واژه‌های ساخته‌شده با این واژه برای پردازش نیاز به تقطیع‌پذیری و تفکیک آنها به اجزای سازنده ندارد. در ادامه با استفاده از فرمول خطی معادله درجه اول برای خط روند، واژه‌های قرارگرفته در بالای خط تقطیع‌پذیر را مشخص کردیم. آنگاه از تقسیم بسامد نوع و بسامد نمونه واژه‌های مشتق قرارگرفته در بالای خط تقطیع‌پذیر بر بسامد نوع و بسامد نمونه کل واژه‌های مشتق ساخته‌شده با پیشوند مورد بررسی نسبت بسامد رخداد نوع تقطیع‌پذیری و نسبت بسامد رخداد نمونه تقطیع‌پذیری را برای هر پیشوند به دست آوردیم. سپس برای هر یک از پیشوندها نسبت‌های تقطیع‌پذیری مشخص‌شده را با یکدیگر مقایسه و مرتبه نسبت بسامد نوع و نمونه هر پیشوند را نیز به دست آوردیم. از نظر های و باین

(۲۰۰۳) هرچه بسامد نمونه واژه مشتق افزایش یابد، نسبت بسامد نوع و نسبت بسامد نمونه تقطیع‌پذیر کاهش می‌یابد. استدلال آنها بر این است که با افزایش بسامد نمونه واژه مشتق، با توجه به اینکه واژه به گوش شنونده آشناتر می‌شود، برای درک و دریافت معنی آن نیازی به تقطیع واژه نیست بلکه آن واژه به صورت یک کل درک می‌شود. برای پی بردن به روند کلی تقطیع‌پذیری پیشوندهای اشتقاقی زبان فارسی و همچنین ارتباط آنها با بسامد نمونه واژه‌های تشکیل‌دهنده هر پیشوند توانستیم با ترسیم نمودار پراکندگی به روند کلی پیشوندهای اشتقاقی زبان فارسی نیز دست یابیم. هدف از این بررسی این است که دریابیم در واژه‌های مشتق پیشوندی زبان فارسی نیز -همان‌طوری که های و باین (۲۰۰۱) معتقدند- با افزایش بسامد رخداد نمونه واژه‌های مشتق و با توجه به مشخصه روان‌شناختی و پردازش ذهنی واژگان آیا نسبت بسامد نوع و نسبت بسامد نمونه تقطیع‌پذیر واژه‌های مشتق کاهش می‌یابد. همان‌گونه که در نمودار ۴ مشاهده می‌شود با جمع‌آوری اطلاعات هر یک از پیشوندها در نرم‌افزار اکسل و با ترسیم نمودار با نرم‌افزار اسپاس روند کلی رابطه بسامد نمونه و نسبت بسامد نوع و نسبت بسامد نمونه تقطیع‌پذیر همراه با خط روند LOESS را ترسیم کردیم. علاوه بر آن برای راحت‌تر نشان دادن این رابطه با استفاده از خط روند خطی و ترسیم آن به صورت نقطه‌چین جهت اصلی این روند نشان داده شده است. در تصویر سمت چپ لگاریتم بسامد نمونه هر یک از پیشوندها در محور افقی و اعداد حاصل از تقسیم بسامد نمونه تقطیع‌پذیر بر بسامد نمونه واژه‌های مشتق از هر وند (نسبت بسامد نمونه تقطیع‌پذیر) در محور عمودی جای گرفته به طوری که محور عمودی را می‌توان نسبتی از محور افقی دانست.



شکل ۴. نمودار پراکندگی تصویر سمت چپ نشان‌دهنده نسبت بسامد نمونه تقطیع‌پذیر به لگاریتم بسامد نمونه واژه‌های مشتق و تصویر سمت راست نشانگر نسبت بسامد نوع تقطیع‌پذیر به لگاریتم بسامد نمونه واژه‌های مشتق. همچنین در تصویر سمت راست لگاریتم بسامد نوع هر یک از پیشوندها در محور افقی و اعداد حاصل از تقسیم نسبت بسامد نوع تقطیع‌پذیر به بسامد نوع واژه‌های مشتق از هر

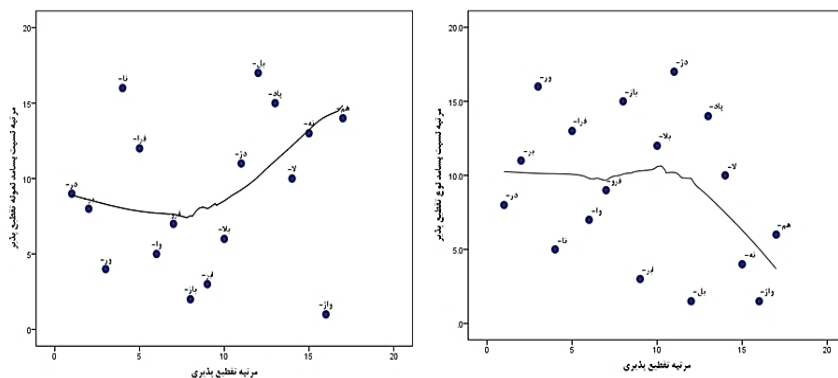
وند (نسبت بسامد نوع) در محور عمودی جای گرفته است. در این نمودار نیز محور عمودی نسبتی از محور افقی است. بر طبق این نمودارها روند کلی به سمت پایین بوده و حالت نزولی به روشنی مشاهده می‌شود. این تصاویر گویای این مطلب هستند که هرچه میزان بسامد نمونه واژه‌های مشتق از وندهای اشتقاقی زبان فارسی بیشتر می‌شود نسبت تقطیع‌پذیری نمونه واژه‌های مشتق و نسبت تقطیع‌پذیری نوع واژه‌های مشتق نیز کمتر می‌شود. در اصل هر چه میزان بسامد واژه‌های مشتق بالاتر می‌رود آن واژه‌ها به گوش شنونده آشناتر شده و در دریافت و پردازش معنی واژه‌های ساخته‌شده با پیشوند مورد نظر نیازی به تقطیع کردن واژه‌های مشتق حاصل از آنها نیست. در این صورت این واژه‌های مشتق پربسامد بیشتر به صورت یک واژه کل درک می‌شوند و پردازش آنها کمتر به صورت سازه‌محور صورت می‌گیرد. این مشاهدات در زبان فارسی هم‌راستا با مشاهدات های و باین (۲۰۰۱) است. برای رسیدن به پیوستار تقطیع‌پذیری از مرتبه نسبت بسامد نوع تقطیع‌پذیری پیشوندها و مرتبه نسبت بسامد نمونه تقطیع‌پذیری استفاده کردیم. برای به دست آوردن مرتبه نسبت بسامد نمونه تقطیع‌پذیر ابتدا آنها را بر اساس کمترین میزان تا بیشترین میزان در نرم‌افزار اکسل مرتب کرده و از شماره یک به آنها مرتبه داده شد. مرتبه نسبت بسامد نمونه و مرتبه نسبت بسامد نوع در سومین و چهارمین ستون جدول ۵ آورده شده است. طبق این جدول مرتبه نسبت بسامد نوع تقطیع‌پذیری برای دو پیشوند «بل-» و «واژ-» عدد ۱/۵ است. چون این دو پیشوند دارای نسبت بسامد نوع تقطیع‌پذیری یکسان بوده و در جایگاه اول و دوم قرار گرفته‌اند لذا، برای تعیین مرتبه آنها از عدد ۱/۵ استفاده کردیم تا مرتبه آنها نسبت به دیگر پیشوندها محاسبه گردد. با گرفتن میانگین از سه مرتبه به دست آمده برای هر پیشوند که عبارت‌اند از مرتبه تقطیع‌پذیر، مرتبه نسبت بسامد نمونه و مرتبه نسبت بسامد نوع، مرتبه میانگین به دست آمده است.

جدول ۶. مرتبه هر پیشوند در زبان فارسی بر اساس مرتبه تقطیع‌پذیری، مرتبه نسبت بسامد نمونه تقطیع‌پذیر،

مرتبه نسبت بسامد نوع تقطیع‌پذیر و میانگین آنها

پیشوند	مرتبه تقطیع‌پذیری	مرتبه نسبت بسامد نمونه تقطیع‌پذیر	مرتبه نسبت بسامد نوع تقطیع‌پذیر	میانگین مرتبه تقطیع‌پذیری، نسبت بسامد نمونه و نسبت بسامد نوع
در-	۱	۹	۸	۸/۵
بر-	۲	۸	۱۱	۹/۵
ور-	۳	۴	۱۶	۱۰
نا-	۴	۱۶	۵	۱۰/۵
فرا-	۵	۱۲	۱۳	۱۲/۵
وا-	۶	۵	۷	۶
فرو-	۷	۷	۹	۸
پاز-	۸	۲	۱۵	۸/۵
ایر-	۹	۳	۳	۳
پلا-	۱۰	۶	۱۲	۹
دژ-	۱۱	۱۱	۱۷	۱۴
بل-	۱۲	۱۷	۱/۵	۹/۲۵
پاد-	۱۳	۱۵	۱۴	۱۴/۵
لا-	۱۴	۱۰	۱۰	۱۰
ن-	۱۵	۱۳	۴	۸/۵
واژ-	۱۶	۱	۱/۵	۱/۲۵
هم-	۱۷	۱۴	۶	۱۰

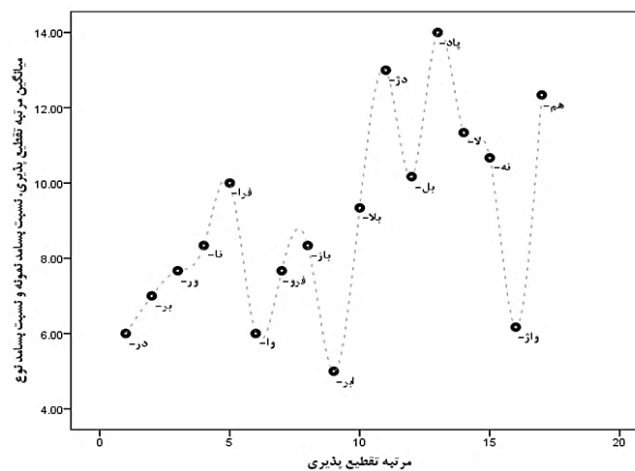
جدول ۶ مرتبه تقطیع‌پذیری بر اساس ماتریس مربعی، مرتبه نسبت بسامد نمونه، مرتبه نسبت بسامد نوع تقطیع‌پذیر و در ستون آخر میانگین مرتبه‌بندی مرتبه تقطیع‌پذیر، مرتبه بسامد نمونه و مرتبه بسامد نوع را نشان می‌دهد. با استفاده از مرتبه‌های به دست آمده از هر یک از موارد جدول ۵ برای هر کدام ابتدا نموداری رسم می‌کنیم تا جایگاه قرارگیری پیشوندهای اشتقاقی زبان فارسی را بسته به میزان نسبت بسامد نمونه، نسبت بسامد نوع و مرتبه تقطیع‌پذیری مشاهده کنیم. به عنوان نمونه جایگاه پیشوند «هم-» در مرتبه‌بندی نسبت بسامد نوع در پیوستار در جایگاه قبل از پیشوندهای «وا-»، «در-»، «بر-»، «فرا-»، «باز-»، «ور-» قرار گرفته است.



شکل ۵. نمودار پراکندگی رابطه مرتبه تقطیع‌پذیری و مرتبه نسبت بسامد نوع تقطیع‌پذیر (نمودار سمت راست) و رابطه مرتبه تقطیع‌پذیری و مرتبه نسبت بسامد نمونه تقطیع‌پذیر (نمودار سمت چپ)

در مرتبه‌بندی نسبت بسامد نمونه تقطیع‌پذیر نیز پیشوند «هم-» قبل از پیشوند «نا-» جای گرفته است. با توجه به واژه‌هایی چون «ناهمگون»، پیشوند «نا-» باید قبل از پیشوند «هم-» قرار گیرد. نکته‌ای که در پیشوندهای اشتقاقی زبان فارسی می‌توان با توجه به نمودارهای پراکندگی شکل ۵ بدان اشاره کرد این است که مرتبه نسبت بسامد نمونه تقطیع‌پذیر- همان‌طور که از پلاگ و باین (۲۰۰۹) انتظار می‌رود- با افزایش مرتبه تقطیع‌پذیری، مرتبه نسبت بسامد نمونه نیز افزایش می‌یابد که این روند را می‌توان با خط روند ترسیم‌شده در نمودار مشاهده کرد. اما رابطه مرتبه تقطیع‌پذیری پیشوندهای اشتقاقی زبان فارسی و مرتبه نسبت بسامد نوع -برخلاف انتظار- با افزایش مرتبه تقطیع‌پذیری مرتبه نسبت بسامد نوع کاهش یافته است. این روند نزولی با ترسیم خط روند در نمودار پراکندگی نشان داده شده است. شاید دلیل آن را بتوان با توجه به بحث زایایی پیشوندها و ارتباط آن با مرتبه تقطیع‌پذیری بیان کرد. همان‌طور که زایایی با افزایش بسامد نوع و بسامد نمونه در ارتباط است، هرچه وندی زایاتر باشد مرتبه نسبت تقطیع‌پذیری بسامد نوع و بسامد نمونه آن بیشتر است. بر اساس نظر های

و باین (۲۰۰۹) افزایش بسامد نمونه با افزایش مرتبه نسبت بسامد نوع و نمونه تقطیع‌پذیر خواهد بود و به تبع آن جهت خط تقطیع‌پذیری رو به بالا قرار خواهد گرفت. در مطالعات آتی می‌توان رابطه مرتبه تقطیع‌پذیری و زایایی را مورد مطالعه قرار داد. در ادامه برای رسیدن به پیوستار ترتیب وندها، نمودار پراکندگی که در محور عمودی، میانگین مرتبه نسبت بسامد نوع، مرتبه نسبت بسامد نمونه و مرتبه تقطیع‌پذیری و در محور افقی مرتبه تقطیع‌پذیری قرار گرفته ترسیم می‌کنیم. شکل ۶ نمودار پراکندگی میانگین مرتبه تقطیع‌پذیری، نسبت بسامد نوع و نسبت بسامد نمونه را نشان می‌دهد.



شکل ۶. پیوستار ترتیب پیشوندهای اشتقاقی زبان فارسی بر اساس رابطه مرتبه تقطیع‌پذیری با میانگین مرتبه تقطیع‌پذیری نسبت بسامد نمونه و نسبت بسامد نوع تقطیع‌پذیر

همان‌طور که در این نمودار پراکندگی پیوستار ترتیب وندها مشاهده می‌شود با توجه به میانگین مرتبه نسبت بسامد نوع و نسبت بسامد نمونه تقطیع‌پذیر جایگاه پیشوند «هم-» قبل از «ور-»، «نا-»، «فرا-» جای گرفته است. همان‌گونه که در بخش دوم نیز گفته شد، رویکرد ترتیب مبتنی بر پیچیدگی بر مبنای پیچیدگی‌های صرفی با دیدگاه تقطیع‌پذیری وندها (های، ۲۰۰۰) بر این باور است: «وندهایی که به راحتی در پردازش زبانی تقطیع می‌شوند نمی‌توانند در جایگاهی نزدیک‌تر به پایه نسبت به وندها با تقطیع‌پذیری کمتر قرار گیرند.» پس همان‌گونه که در نمودار گراف جهت‌دار هم مشاهده کردیم همان‌طوری که های و پلاگ (۲۰۰۹) می‌گویند شاید بهترین پیوستار برای ترتیب قرارگیری پیشوندهای اشتقاقی زبان فارسی پیوستار مرتبه تقطیع‌پذیری و میانگین مرتبه‌های به دست آمده از هم‌آیندی پیشوندهای اشتقاقی بر اساس محدودیت‌های گزینشی باشد. با توجه به بررسی‌های انجام‌شده در زبان فارسی می‌توان به این امر قائل شد که محدودیت‌های گزینشی به تنهایی نمی‌توانند پیوستاری مبنی بر ترتیب وندافزایی ارائه دهند چون همان‌طور که در داده‌های زبان فارسی مشاهده کردیم

در نمودار گراف جهت‌دار هفت پیشنهاد از هفده پیشنهاد مورد بررسی در این مقاله در هم‌آیندی با دیگر پیشنهادها شرکت نداشته پس نمی‌توان برای آنها به لحاظ محدودیت‌های گزینشی و هم‌آیندی آنها ترتیبی نسبت به دیگر وندها قائل شد. در این خصوص مرتبه‌های حاصله از تقطیع‌پذیری می‌تواند جایگاه وندها نسبت به یکدیگر را مشخص کند. همچنین همان‌گونه که از نمودارهای بسامد نوع و نمونه تقطیع‌پذیری برداشت شد پیشنهادهایی که در هم‌آیندی بر اساس محدودیت‌های گزینشی در جایگاهی نزدیک‌تر به پایه باید قرار گیرند نسبت به دیگر وندها در جایگاه دورتری قرار گرفته بودند. با توجه به این نکته مشخص می‌گردد این دو رویکرد در کنار هم می‌توانند برای همه وندهای اشتقاقی زبان فارسی مرتبه‌ای را در نظر گرفته و آنها را بر روی پیوستاری که توجیه‌کننده ترتیب پیشنهادهای اشتقاقی زبان فارسی هستند قرار دهند. در نمودار ۵ برای ایجاد رابطه میان مرتبه تقطیع‌پذیری و میانگین مرتبه تقطیع‌پذیری نسبت بسامد نوع و نمونه تقطیع‌پذیر نمودار پراکندگی ترسیم شد. طبق این نمودار با افزایش مرتبه تقطیع‌پذیری میانگین مرتبه نسبت بسامد نوع و نمونه تقطیع‌پذیر افزایش یافته است. پیشنهاد «در» بیشترین میزان تقطیع‌پذیری و پیشنهاد «هم» کم‌ترین میزان تقطیع‌پذیری را دارا هستند. های (۲۰۰۰) در مبحث تقطیع‌پذیری معتقد است تقطیع‌پذیری به صورت یک عدد مطلق نیست و با درجات مختلف می‌تواند بر روی یک پیوستار قرار گیرد. در زبان فارسی نیز ما می‌توانیم بر اساس تصویر ۶ سلسله مراتب قرارگیری آنها را نشان دهیم. بر اساس آنچه قبلاً گفته شد محدودیت‌های گزینشی تنها می‌توانند مرتبه پیشنهادهایی که در هم‌آیندی با یکدیگر شرکت داشته‌اند را معین کنند ولی اگر پیشنهادهایی در هم‌آیندی شرکت نکرده باشند نمی‌توان برای آنها به توجه به این رویکرد مرتبه‌ای نسبت به دیگر پیشنهادها قائل شد. اما با استفاده از نسبت بسامد نمونه و نسبت بسامد نوع وندهای تقطیع‌پذیر می‌توان پیشنهادهای اشتقاقی را در مرتبه‌های دقیق‌تری قرار داد. به عنوان نمونه بر اساس پیوستار ترتیب پیشنهادهای به دست آمده در تصویر ۶ در واژه‌های «درنیافته، برنخوردن، ورنکشیده، بازنگشتن» پیشنهادهای «در- بر-، و- باز-» در این پیوستار به ترتیب قبل از پیشنهاد «ن-» قرار گرفته‌اند. همچنین در واژه‌های «درهم پیچیده، برهم خوردگی، ناهمنوا، فراهمایی» پیشنهاد «در- بر-، نا- فرا-» به ترتیب قبل از پیشنهاد «هم-» در پیوستار دیده می‌شوند. از این رو است که در توصیف ترتیب پیشنهادهای اشتقاقی زبان فارسی می‌توان شاهد عملکرد تکمیلی هر دو رویکرد ترتیب مبتنی بر پیچیدگی و محدودیت‌های گزینشی شد.

۵. نتیجه

با بهره‌گیری از رویکرد ترتیب مبتنی بر پیچیدگی و با تکیه بر معیار مرتبه تقطیع‌پذیری، و مرتبه نسبت تقطیع‌پذیری می‌توان پیشنهادهای اشتقاقی زبان فارسی را بر روی یک پیوستار قرار داد. همان‌گونه که داده‌های زبان فارسی نشان می‌دهد هرچه بسامد نمونه واژه‌ها افزایش یابد نسبت بسامد تقطیع‌پذیری نوع و نمونه نیز کاهش پیدا می‌کند. به عبارتی هرچه بسامد واژه‌های مشتق ساخته‌شده با پیشوند مورد نظر بیشتر باشد برای درک و پردازش آنها نیازی به تقطیع‌پذیری نیست. همان‌گونه که ذکر شد رویکرد ترتیب مبتنی بر پیچیدگی صرفاً به مباحث پردازشی و تقطیع واژه‌های مشتق می‌پردازد در سوی دیگر با توجه به رویکرد محدودیت‌های گزینشی و مباحث مطرح شده در آن، اعم از محدودیت‌های صرفی، نحوی، واجی و معنایی، می‌توان به وجود یا نبود ترکیبات موجود و بالقوه در واژه‌های مشتق پرداخت. لذا رویکرد ترتیب مبتنی بر پیچیدگی و رویکرد محدودیت‌های گزینشی در کنار یکدیگر می‌توانند به طور مکمل توجیه مناسبی از ترتیب وندها به دست دهند. در واقع در مواردی که محدودیت‌های گزینشی قادر به مرتبه‌بندی پیشنهادهای نیست، رویکرد تقطیع‌پذیری می‌تواند با موارد مطرح شده در آن مرتبه هر وند را مشخص کرده و نیز مواردی که رویکرد تقطیع‌پذیری برای قراردادن پیشنهادهای و ترتیب آنها بر روی یک پیوستار دلیل منطقی ندارد محدودیت‌های حاکم بر هر پیشوند می‌تواند به توجیه آن بپردازد. این نتایج با نتایج های و باین (۲۰۰۹) نیز همسو می‌باشند.

پی‌نوشت‌ها

۱. در اینجا لازم است از جناب آقای دکتر مصطفی عاصی و همکاران محترم ایشان در پژوهشگاه علوم انسانی و مطالعات فرهنگی به خاطر فراهم ساختن امکان استفاده از پیکره پایگاه داده‌های زبان فارسی تشکر کنیم.
2. lexical database (Baayen, Piepenbrock, Gullikers 1995)
۳. پیکره پایگاه داده‌های زبان فارسی پژوهشگاه علوم انسانی و مطالعات فرهنگی در مجموع دارای شصت میلیون واژه ست. این پیکره حاوی هزار و پانصد اثر نظم و نثر ادبیات معاصر ایران و انواع متون زبانی اعم از اثرهای داستانی و غیرداستانی نثر، آثار شعری، نمایشنامه، فیلمنامه، ادبیات کودکان، مجله‌ها و نشریه‌های علمی، ادبی و تخصصی، روزنامه و نشریه‌های خبری، کتاب‌های درسی دانشگاهی و دبیرستانی، نامه‌های اداری، متون حقوقی و دیوارنوشته‌ها است.
۴. در انتخاب پیشنهادهای حاضر در این پژوهش، ذکر سه نکته ضروری است. نکته اول این‌که وندهای انتخاب‌شده تمامی پیشنهادهای زبان فارسی نیستند. با در نظر گرفتن حجم بسیار بالای داده‌ها و محاسبات آماری به این تعداد از پیشنهادهای که در کتاب ساخت اشتقاقی واژه در فارسی امروز (کلباسی، ۱۳۸۰) ذکر شده است بسنده شد. در آثار مهم دیگری که بر روی وندهای زبان انگلیسی انجام شده نیز از میان وندهای موجود وندهایی انتخاب شده‌اند. به عنوان مثال‌های و پلاگ (۲۰۰۴) به بررسی ۱۵ پسوند اشتقاقی زبان انگلیسی و زیرکل (۲۰۱۰) به بررسی ۱۵ پیشوند اشتقاقی پرداخته‌اند. این امر اشکالی در یافته‌های تحقیق ایجاد نمی‌کند چون در هر حال مبحث ترتیب وندها در میان است. بدیهی است که افزودن تمام پیشنهادهای می‌تواند به بهتر شدن تحقیق کمک کند. نکته دوم این‌که چون در این چهارچوب فرض بر آن است که ترتیب وندها متأثر از میزان تقطیع‌پذیری و آن هم متأثر از میزان زیایی وندها است، وندهای مثل «بل-»، «دژ-» و موارد دیگر نیز در فهرست پیشنهادهای اشتقاقی مورد بررسی قرار گرفته‌اند. نکته آخر اینکه در مورد پیشوند «ن-» -با توجه به تمایز بین صورت‌های تصریفی نظیر «و نمی‌رود» و نمونه‌های مشابه با صورت‌های اشتقاقی نظیر «نشکن» در مثلاً «ظرف نشکن» و نمونه‌های مشابه- صرفاً واژه‌های اشتقاقی استخراج شده و در مبنای تحلیل قرار گرفته‌اند.

منابع

- حمصیان، مریم (۱۳۸۹). بررسی آرایش وندهای اشتقاقی زبان فارسی، پایان نامه کارشناسی ارشد، دانشگاه اصفهان.
- غلامعلی زاده، خسرو و فیضی پیرانی، یداله (۱۳۹۰). بررسی لایه‌ها و سطوح واژگانی در وندهای زبان فارسی از نگاه صرف واژگانی، مجله پژوهش‌های زبان‌شناسی، سال سوم، شماره ۵، اصفهان: سامانه نشریات دانشگاه اصفهان.
- کشان، خسرو (۱۳۷۱). اشتقاق پسوندی در زبان فارسی/امروز، تهران، مرکز نشر دانشگاهی.
- کلباسی، ایران (۱۳۸۰). ساخت اشتقاقی واژه در فارسی/امروز، تهران، پژوهشگاه علوم انسانی و مطالعات فرهنگی.
- Allen, M. 1978. *Morphological investigations*. Ph.D. Thesis, University of Connecticut.
- Baayen, R. H. 1992. Quantitative Aspects of Morphological Productivity, In Booij G. & J. Van Marle, *Yearbook of Morphology 1991*, Dordrecht: Kluwer, 49-109.
- _____. 2002. *Word Frequency Distribution*, Dordrecht, Kluwer Academic Publishers.
- Fabb, N. 1988. English Suffixation is Constrained only by Selectional Restrictions, *Natural Language and linguistic Theory* 6, 527-539.
- Gholamali Zade, Kh. & Feizi Pirani, Y. 2011. Lexical Stratum and Layers in Persian Affixes Based on Lexical Morphology. *Journal of Researches in Linguistics*, 65-84. [In Persian].
- Giegerich, H. J. 1999. *Lexical Strata in English. Morphological causes, Phonological Effects*, Cambridge: Cambridge University Press.
- Hay, J. 2000. *Causes and Consequences of Word Structure*, Ph.D. dissertation. Northwestern University.
- _____. 2002. From Speech Perception to Morphology: Affix-ordering Revisited, *Language* 78.3, 527-555.
- Hay, J. & Baayen, H. 2002. Parsing and Productivity. *Yearbook of Morphology*, Dordrecht: Kluwer, 203- 255.
- Hay, J. & Plag, I. 2004. What Constraints Possible Suffix Combinations? On the Interaction of Grammatical and Processing Restrictions in Derivational Morphology, *Natural Language and Linguistic Theory* 22, 565-596.
- Hemasian, M. 2010. *The Survey of Persian Derivational Affixes*. MA. Thesis, University of Isfahan. [In Persian].
- Kalbasi, I. 2001. *The Derivational structure of Word in Modern Persian*. Tehran: Institute for Humanities and Cultural Studies. [In Persian].
- Keshani, Kh. 1992. *Suffix Derivation in Modern Persian*. Tehran: Iran University Press. [In Persian].
- Kiparsky, P. 1982a. From Cyclic Phonology to Lexical Phonology, in van der Hulst and Smith, 1982a.
- _____. 1982b. *Lexical Phonology & Morphology*, Seoul: Linguistics in the Morning Calm.
- Mohanan, K.P. 1986. *The Theory of Lexical Phonology*, Dordrecht: Reidel.
- Plag, I. 1996. Selectional Restriction in English Suffixation revisited: A Reply to Fabb 1988, *Linguistics*, 34, 769-798.

- . 1999. *Morphological Productivity: Structural Constraints in English Derivation*, Berlin: Mouton de Gruyter.
- . 2002. The Role of Selectional Restrictions, Phonotactics and Parsing in Constraining Suffix Ordering in English, in Geert E. Booij and Jaap van Marle (eds.), *Yearbook of Morphology 2001*, Foris, Dordrecht, 285–314.
- . 2003. *Word-formation in English*, Cambridge: Cambridge University Press.
- . 2004. Syntactic Category Information and the semantics of Derivational Morphological Rules, *folia Linguistica* 38, 193-225.
- Selkirk, E. 1982. *The Syntax of Words*, Cambridge: MIT Press.
- Siegel, D. 1974. *Topics in English Morphology*, Cambridge: MIT Pr