
CHAPTER 4

Measuring productivity in the Assamese Prefixes

In the previous chapter, we sketched out the prefixes and discussed their semantic descriptions; in this chapter, we evaluate the productivity of the same using data obtained from the corpus consisting of digital texts (sample A) and an Assamese dictionary (sample B).

4.1 Sample A: Corpus

As mentioned earlier, Sample A is the data collected from a sample of one lac words. The sample comprises of contemporary texts collected from digital platforms. Below, Table 4.1 presents the statistical data gathered from sample A and Table 4.2 calculates the productivity of the prefixes based on Table 4.1 by applying the measuring methods mentioned in Chapter 2. Here, Type means every distinct word formed by a prefix, Token means individual occurrences regardless of how often they are repeated by a prefix and Hapax indicates the words formed by these prefixes that occur only once in the entire sample.

Table 4.1 Number of Types, Token and Hapaxes in Sample A

SL	NEG Prefixes	Type V	Token N	Hapax n1
1	অ- <i>a-</i>	208	843	117
2	অপ- <i>apa-</i>	13	15	6
4	দূৰ-/দুৰ- <i>dur-</i>	33	76	14
5	কু- <i>ku-</i>	3	12	1
6	নি- <i>ni-</i>	123	415	59
7	বি- <i>bi-</i>	29	203	12

Table 4.2 Results of Type Frequency, Token Frequency, Hapax Legomena, Token/Type ratio, Type/Token ratio and Token/Type ratio

Prefixes	Type (V)	Token (N)	Hapax (n1)	Type/Token (V/N)	Token/Type (N/V)	P* (n1/N)	n1/V
অ- <i>ɔ-</i>	208	843	117	0.247	4.053	0.139	0.563
অপ- <i>ɔpɔ-</i>	13	15	6	0.867	1.154	0.4	0.462
দূৰ-/দুৰ- <i>dur-</i>	33	76	14	0.434	2.303	0.184	0.424
কু- <i>ku-</i>	3	12	1	0.25	4	0.083	0.333
নি- <i>ni-</i>	123	415	59	0.296	3.217	0.142	0.480
বি- <i>bi-</i>	29	203	12	0.143	7	0.059	0.414

4.1.1 Type frequency

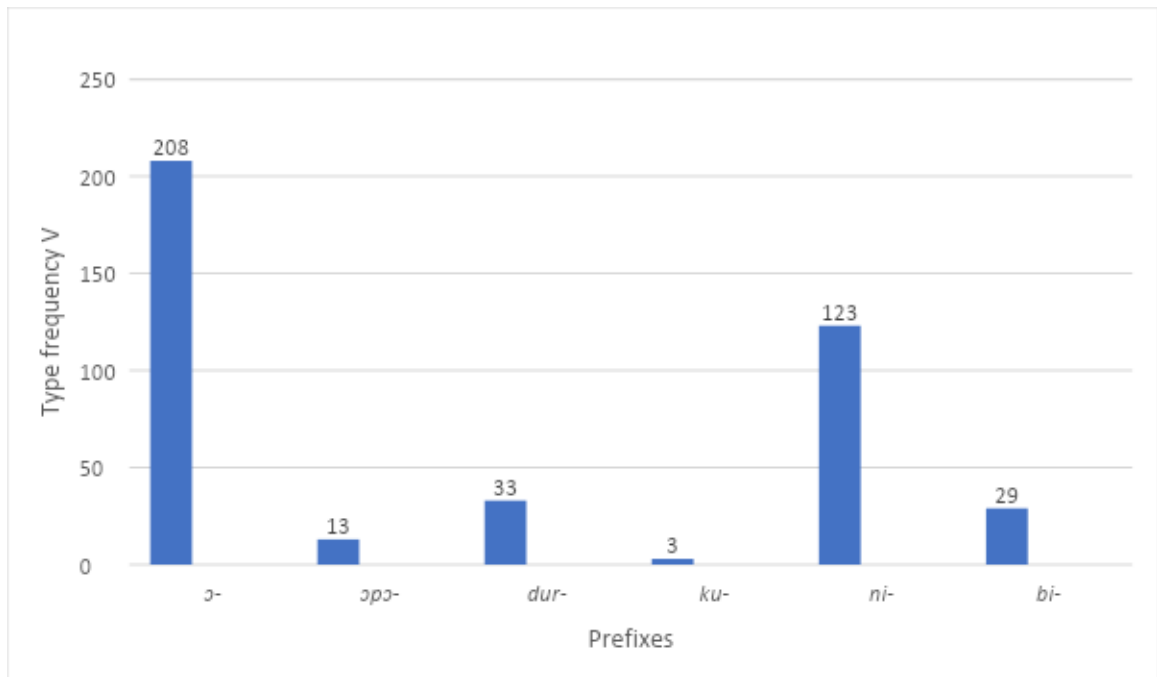


Fig 4.1 Type frequency of prefixes

In terms of Type frequency V, *ɔ-* has the highest frequency while *ku-* displays the lowest frequency. The prefixes *ni-*, *dur-*, *bi-* and *ɔpɔ-* display subsequent productivity after *ɔ-*. However, amongst the prefixes, *ɔ* and *ni-* have distinctly higher frequencies than the rest with 208 and 123 types respectively, which means *ɔ-* and *ni-* are more productive than the others.

4.1.2 Token frequency

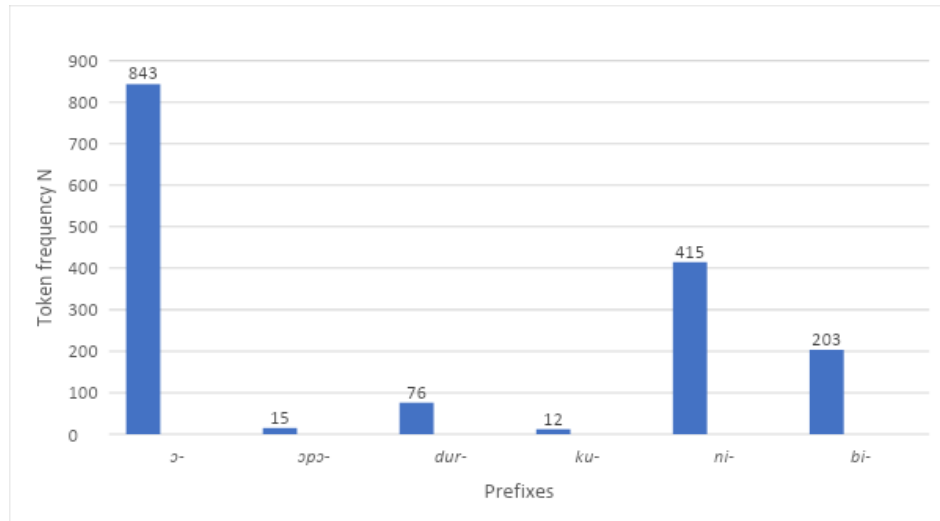


Fig 4.2 Token frequency of prefixes

Like Type frequency, ∂- has the highest frequency in terms of Token frequency N as well followed by ni- and ku- having the lowest frequency. However, bi- is more frequent than dur- unlike the previous one. The position of ∂∂- also remains the same. From here, we see that the result of Type frequency and Token frequency are nearly the same.

4.1.3 Type/Token (V/N) and Token/Type (N/V) frequency

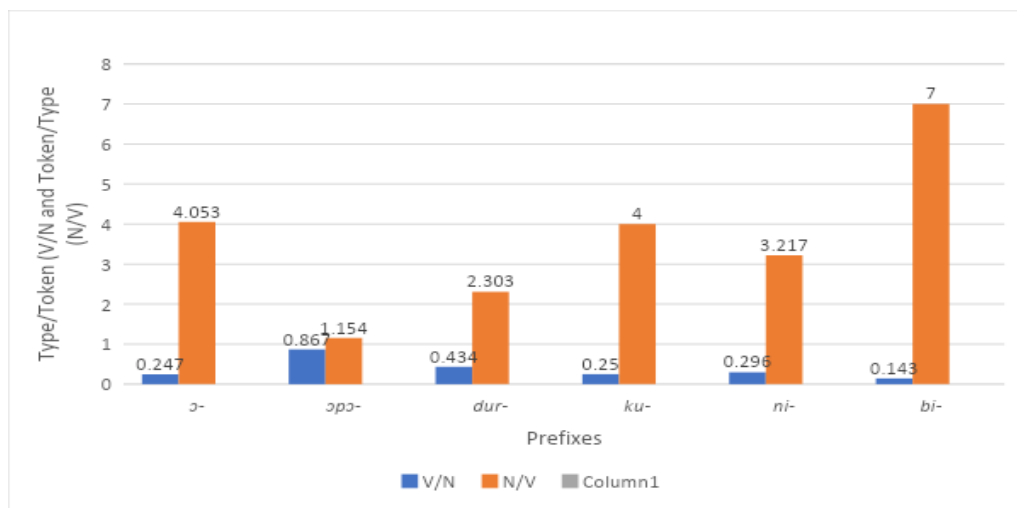


Fig 4.3 V/N and N/V ratio of prefixes

After Type frequency and Token frequency methods, the next methods that are going to be employed show the probability of occurring the prefixes in new word formation and they use the variables type (V), token (N) and hapax (n1). V/N and N/V method incorporates types and tokens to measure the ratio. We have already mentioned that the higher value of Type/ Token (V/N) indicates higher productivity, and the higher value of Token/Type (N/V) indicates the opposite, i.e., the lower productivity (Baayen and Lieber 1991; Plag 1999, Hulse 2010). However, unlike Type frequency and Token frequency, the outcome of this measure is different. Here, *ɔ-* is ranked in the second position from the bottom unlike the previous two. Again, *ɔpɔ-* which was in the second last position in V and N methods, displays the highest frequency in this method followed by *dur-*, *ni-*, *ku-*, *ɔ-* and *bi-*.

4.1.4 Productivity in the strict sense (P): $n1/N$

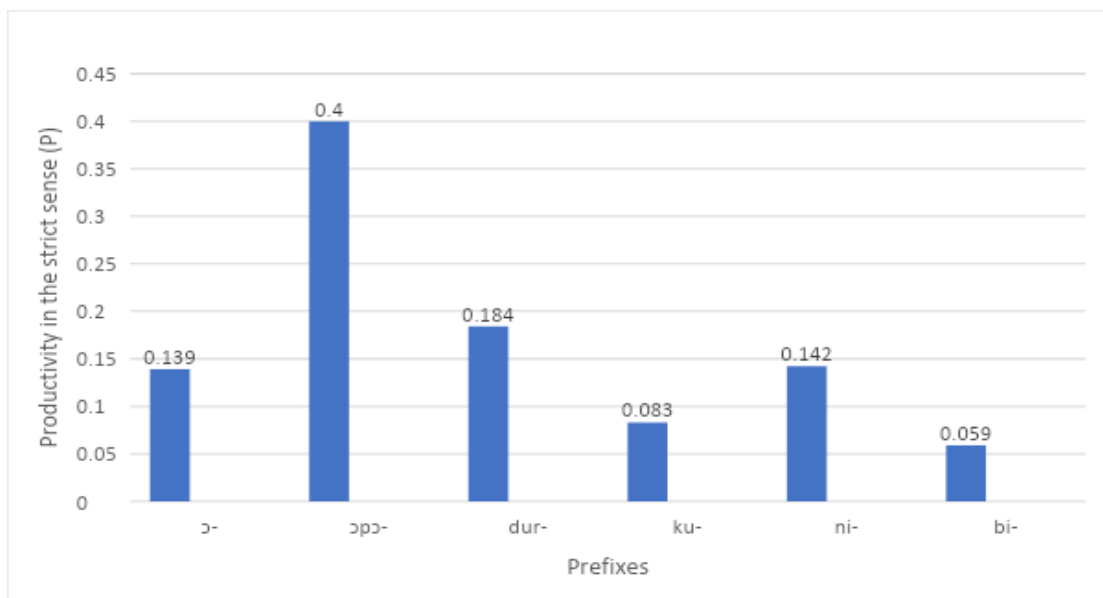


Fig 4.4 'Productivity in the strict sense': $n1/N$ ratio for the prefixes

Methods 4.1.4 and 4.1.5 utilise hapax legomena. While the ratio is calculated by using hapax and token frequency in 4.1.4, it is calculated by using hapax and type frequency in 4.1.5. The higher ratio in Productivity in the strict sense (P) indicates higher productivity. It calculates that *ɔpɔ-* has the highest productivity and *bi-* has the lowest productivity rate.

4.1.5 Hapax/Type (n1/V) method

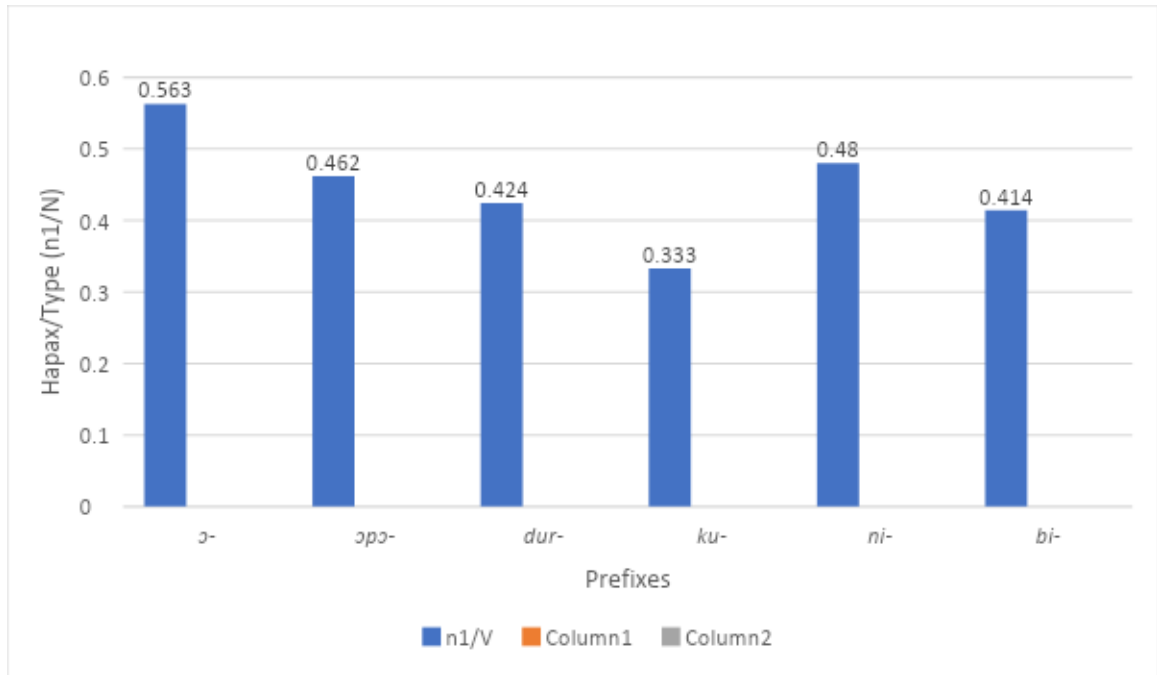


Fig 4.5 n1/V ratio of the prefixes

In this method, we see that *ɔ-* has the highest productivity rate followed by *ni-*, *ɔpɔ-*, *dur-* and *bi-*, while *ku-* has the lowest productivity rate.

Now, the prefixes are arranged in descending order in Table 4.3 based on the values calculated in Table 4.2:

Table 4.3 Ranking of prefixes in descending order of productivity by each method

Ranking	Type (V)	Token (N)	Hapax (n1)	Type/Token (V/N)	Token/Type (N/V)	P* (n1/N)	n1/V
1.	ɔ-	ɔ-	ɔ-	ɔpɔ-	ɔpɔ-	ɔpɔ-	ɔ-
2.	ni-	ni-	ni-	dur-	dur-	dur-	ni-
3.	dur-	bi-	dur-	ni-	ni-	ni-	ɔpɔ-
4.	bi-	dur-	bi-	ku-	ku-	ɔ-	dur-
5.	ɔpɔ-	ɔpɔ-	ɔpɔ-	ɔ-	ɔ-	ku-	bi-
6.	ku-	ku-	ku-	bi-	bi-	bi-	ku-

4.2 Interpretation

From Table 4.2 and Table 4.3, we see that the results of types, token and hapaxes have similar frequency rates. Only *dur-* and *bi-* have altered their ranking from third and fourth in Type frequency to fourth and third in Token frequency respectively. Except for these two, others remained the same.

In the probabilistic methods, which employ two variables to measure productivity rate, we see a different rate of productivity. Here, the results of these methods show a contrastive picture with that of Type and Token frequency methods. *ɔpɔ-* which was lowly productive in V and N methods, turns out to be the highest productive prefix in V/N or N/V and n1/N methods. Again, in the case of *ɔ-* also which was the highest productive prefix in V and N methods, has the lower productivity rate in the methods V/N or N/V and n1/N. However, as the overall result of n1/V is similar to N and V methods, *ɔ-* turns out as the highest frequent prefix in n1/V method as well.

Nonetheless, a few prefixes seem to exhibit uniform productivity rates across all methods. *ni-* is in the 2nd, 2nd, 3rd, 2nd and 2nd rank in V, N, V/N or N/V, n1/N and n1/V methods respectively. Similarly, *dur-* is in the 3rd, 4th, 2nd, 2nd and 4th rank in the V, N, V/N or N/V, n1/N and n1/V methods respectively. Unlike *ɔ-* and *ɔpɔ-*, they do not produce widely divergent outcomes in the methods.

Again, in the case of *bi-* and *ku-*, we observe another tendency. *bi-* is the lowest frequent prefix in the methods N/V or V/N and n1/N methods. Other methods, likewise, place this prefix somewhat lower in their rankings. In V, N and n1/V methods, it occupies 4th, 3rd and 5th ranking respectively. Similarly, *ku-* has the lowest frequency in V, N and n1/V methods. It also demonstrates the propensity toward downside in the other two approaches, V/N or N/V and n1/N. It has 5th and 4th ranking in V/N or N/V and n1/N methods respectively.

Now, based on these trends, the prefixes can be loosely classified into three categories. *ni-* has a comparatively higher productivity rate throughout the methods than others; since the rank of *dur-* is somewhere in the middle, it may be said that it is medially productive, while *bi-* and *ku-* can be considered as lowly productive as both tend

to display their ranking from middle to towards the end in all the methods. However, the prefixes *ɔ-* and *ɔpɔ-* show extreme and contrastive productivity rates in different productivity methods. Either they are the highest productive (*ɔ-* in V, N and n1/V methods, *ɔpɔ-*, the highest in V/N or N/V method and 3rd highest in n1/V method) or the lowest productive prefixes (*ɔ-* second lowest in V/N and N/V method and 3rd lowest in n1/N method, *ɔpɔ-* second lowest in V and N methods).

It has already been discussed that different productive methods display different aspects of productivity, and that none of them are exhaustive. The prefixes that are productive in the V and N methods tell us that the past productivity of these prefixes is the highest and the same goes for the lowest productive prefixes. *ɔ-* is the highest productive prefix in V and N methods meaning it has the highest past and present productivity. Similarly, *ku-* has the lowest productivity rate in the V and N methods tells us that it has the lowest past productivity.

In the case of probabilistic methods, the higher the productivity, the higher they have the chances of forming new words in the future. *ɔpɔ-* being in the first rank in V/N or N/V and n1/N indicates that it has the highest possibility of forming new words in the future. However, the probabilistic method n1/V places *ɔ-* at the top, which hints that not only *ɔ-* has the highest past productivity, it may also have the capability of generating new words in the future. Similarly, *bi-* turns out to be the least frequent prefix in the V/N or N/V and n1/N method which means that it has the least probability of forming new words in the future.

Coming to the highest productive prefixes, across the methods, we can say that *ɔ-* is one of the highest productive negative prefixes of the language. Not only that it has the highest frequency in type, token and hapax wise, but it also has the highest frequency in one of the probabilistic methods n1/V. Van Marle (1992) argues that Type frequency is more useful than Token frequency to gauge productivity, as Type frequency allows or records the words that are produced by utilising a morphological process. If we go by this logic, then the result of n1/V is more convenient than the result of n1/N. Now, *ɔ-* occurring as the highest frequent prefix in the n1/V method, which is also highest in the V and N methods as well, establishes or resonates the fact that *ɔ-* is reasonably a high

productive prefix. *অপ-* also occupies the mid rank in n1/V, which says that it is medially productive in terms of building potential words.

4.3 Sample B: Dictionary

Our next sample, i.e., sample B is created with the data extracted from the Assamese dictionary *Hemkosh* ed. 2006 and 2016. The data are collected from both editions of the dictionary separately and arrived at the following values. Table 4.4 shows the total number of types in both editions along with the newly added words in the span of ten years, which give us the result of *Type frequency V*.

Table 4.4 Number of negatively-prefixed words in 2006 and 2016

SI	Negative Prefixes	2006	2016	New words
1	অ- <i>অ-</i>	1383	1720	337
2	অপ- <i>অপ-</i>	153	170	17
3	দুৰ-/দূৰ- <i>dur-</i>	103	121	18
4	কু-/কূ- <i>ku-</i>	49	77	28
5	নি-/নী- <i>ni-</i>	252	285	33
6	বি- <i>bi-</i>	116	140	24
Total		2056	2513	457

Table 4.5 Ranking of Negative prefixes in descending order from *Hemkosh*

SI	2006	2016	New words
1	অ- <i>অ-</i>	অ- <i>অ-</i>	অ- <i>অ-</i>
2	নি-/নী- <i>ni-</i>	নি-/নী- <i>ni-</i>	নি-/নী- <i>ni-</i>
3	অপ- <i>অপ-</i>	অপ- <i>অপ-</i>	কু-/কূ- <i>ku-</i>
4	বি- <i>bi-</i>	বি- <i>bi-</i>	বি- <i>bi-</i>
5	দুৰ-/দূৰ- <i>dur-</i>	দুৰ-/দূৰ- <i>dur-</i>	দুৰ-/দূৰ- <i>dur-</i>
6	কু-/কূ- <i>ku-</i>	কু-/কূ- <i>ku-</i>	অপ- <i>অপ-</i>

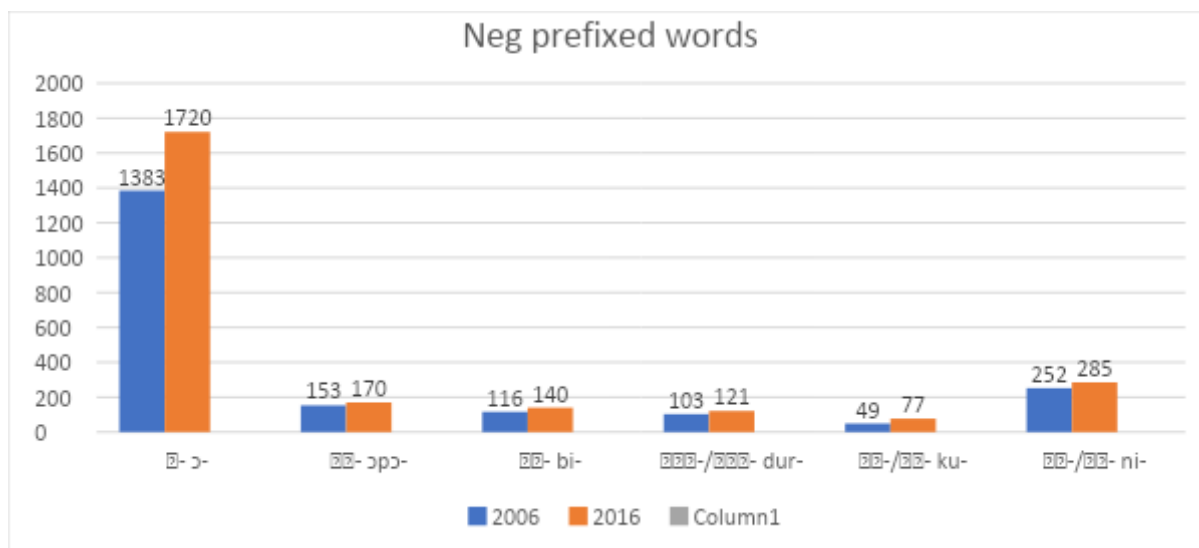


Fig 4.6 Negatively-prefixed words in *Hemkosh* (Ed. 2006 and 2016)

4.3.1 Type frequency V in 2006 and 2016

For the *Type frequency of V*, in both editions, the prefixes have displayed similar type frequencies. ɔ- has the highest number of types in both which is 1383 and 1720 in 2006 and 2016 respectively. ni- is the second highest frequent prefix followed by ɔpɔ-, bi-, dur- and ku- in both editions. However, it is clear from the statistics that compared to the other prefixes, ɔ- has a surprisingly high number of types. It is evident from the graphical representation that if we need to arrange the prefixes into separate categories, ɔ- alone can be placed in opposition to the other categories.

4.3.2 Type frequency V of newly added words

Turning to the words that have been newly added to the dictionary over 10 years, we see that 457 new words were added in the 2016 edition, which were not listed in 2006. After having a look at the kind of new words that are considered for listing, we find that most of the words are not neologisms. Since we know that the listing of words in a dictionary primarily depends on the choices of the lexicographer, and also the gap between two editions is only ten years, it is certain that there must have been several words that went unnoticed in the eyes of the lexicographer in both editions.

Among the new words which are added in the latest edition of the dictionary, almost all are existing words or commonly used words in the language. For example,

ɔsɔlɔnɪjɔ ‘not movable’, *ɔsikitsɔnɪjɔ* ‘not treatable’, *ɔbik^hjat* ‘not famous’, *kusɔritrɔ* ‘bad behaviour’, *nirakaŋk^hja* ‘devoid of desire’, *bizoɔɔk* ‘decomposer’, *bisruti* ‘Perplexed’ etc. to name a few are familiar words of the language, yet they were not added till 2016.

Again, certain words along with their word-forms within the same word class or different categories are found to be listed in the previous edition and a few others are included in the latest edition. For some, the un-prefixed bases are already registered separately and derivatives created through the process of prefixation are listed separately in 2016. For example, the words *kɔlux* কলুষ ‘filth’ and *kɔluxitɔ* কলুষিত ‘filthy’ are listed in the 2006 edition. The derived form of *kɔlux* কলুষ ‘filth’ is *ɔkɔlux* অকলুষ ‘filthless’ which is also registered in that edition. However, the derivation of *kɔluxitɔ* কলুষিত ‘filthy’ i.e., *ɔkɔluxitɔ* অকলুষিত ‘unfilthy’ is added only in the 2016 edition. Among the newly added words, a sizable number of words belong to this type of derivation which the lexicographer decided to provide a new entry in the later edition. A few other examples are mentioned below.

2006	2016
অকাল <i>ɔkal</i> ‘out of season’, অকালজাত ‘untimely born’	অকালন <i>ɔkalɔn</i> ‘Very young’, অকালজ <i>ɔkalɔz</i> ‘unseasonal’
অজীর <i>ɔziwɔ</i> ‘Lifeless’	অজীরজনন <i>ɔziwɔzɔnɔn</i> ‘Abiogenesis’, অজীরজননি <i>ɔziwɔzɔnɔni</i> ‘Spontaneous generation’, অজীরন <i>ɔziwɔn</i> ‘Lifeless’, অজীরাৎজীরোৎপত্তি <i>ɔziwatziwɔtpɔtti</i> ‘Abiogenesis’, অজীরিক <i>ɔziwik</i> ‘Without earnings’, অজীরিয় <i>ɔziwijɔ</i> ‘Abiotic’
অগণন <i>ɔgɔnɔn</i> ‘countless’, অগণ্য <i>ɔgɔɪnjɔ</i> ‘innumerable’	অগণিত <i>ɔgɔnit</i> ‘incalculable’
অপ্রয়োজনীয় <i>ɔprɔjɔzɔnɪjɔ</i> ‘unnecessary’	অপ্রয়োজনীয়তা <i>ɔprɔjɔzɔnɪjɔta</i> ‘needlessness’
অপ্রবাসী <i>ɔprɔbaxi</i> ‘not living away from home or own country’	অপ্রবাস <i>ɔprɔbax</i> ‘to stay in motherland’
অগতি <i>ɔgɔti</i> ‘One who has no refuge or resources’	অগতিক <i>ɔgɔtik</i> ‘Having no alternative way, having no other course’
অঘটন <i>ɔg^hɔtɔn</i> ‘That which cannot occur or not likely to occur’	অঘটিত <i>ɔg^hɔtit</i> ‘not happened’

অচল <i>asol</i> ‘Not moving, fixed/ mountain’, অচলজ <i>asolz</i> ‘Born in the mountains’	অচলন <i>asoln</i> ‘Obsolescence’, অচলনীয় <i>asolnija</i> ‘Incapable of being introduced or circulated’
অনিয়ন্ত্রণ <i>onijontron</i> ‘Lacking in self-control’	অনিয়ন্ত্রিত <i>onijontrito</i> ‘Uncontrolled’
অপূৰ্ণীয় <i>apuronija</i> ‘Which cannot be filled or compensated’	অপূৰ্ণ <i>apuron</i> ‘Unfulfilled’
অচিকিৎসা <i>asikitsa</i> ‘Lack of proper treatment’	অচিকিৎসনীয় <i>asikitsonija</i> ‘That cannot be treated’

In the first column, the lexemes included in the 2006 edition are listed, and in the second column, the terms included later in the 2016 edition that are related to the previous edition are listed. We note that newly added words are primarily an extension or declension of an already existing lexeme, making it relatively easy to decode their meaning. For example, the words which are formed after *aziwɔ* অজীৰ, all are related to a basic meaning ‘Lifeless, inanimate’. While *aziwɔ* অজীৰ is an adjective, the newly added words belong to both noun and adjective categories. Similarly, *aprobaxi* অপ্রবাসী ‘not living in a foreign country’ is listed in 2006, but the act of not living in a foreign country and living in the motherland, which is termed as *aprobax* অপ্রবাস ‘to stay in one’s motherland’ is added in the 2006 edition. Assuming that this word is not listed in the recent edition of the dictionary, but if it is used somewhere, we will still be able to decode its meaning using the associated word forms that already exist. This suggests that the majority of the newly added words are the outcome of a productive pattern and are not wholly new coinages.

In the case of *apɔ-* also, most of the words are already in the language but are used less frequently. The opposite of *zati* ‘caste’ already exists as *azati* which means ‘out of caste’, and by creating the word *apazati*, it is tried to suggest that caste that is inferior instead of ‘out of caste’. Again, we can find analogous creations such as *apɔbadɔk* অপবাদক ‘one who reviles’. *badɔk* বাদক usually means someone who plays an instrument, such as *bahibadɔk* বাঁহীবাদক ‘flutist’. *apɔbadɔk* অপবাদক is created in line with *bahibadɔk*. In the case of *dur-* and *ku-* also, most of the words are analogous creations. The new words recorded for *ni-* are basically creative words such as *nirbinijozon* ‘disinvestment’,

nikʰonizkəron ‘demineralization’, *nirbixaktəkəron* ‘detoxification’ etc. Again, for *bi-* these words are both existing as well as creative words.

Now, coming to *Type frequency V* of newly added words, we arrive at the following result:

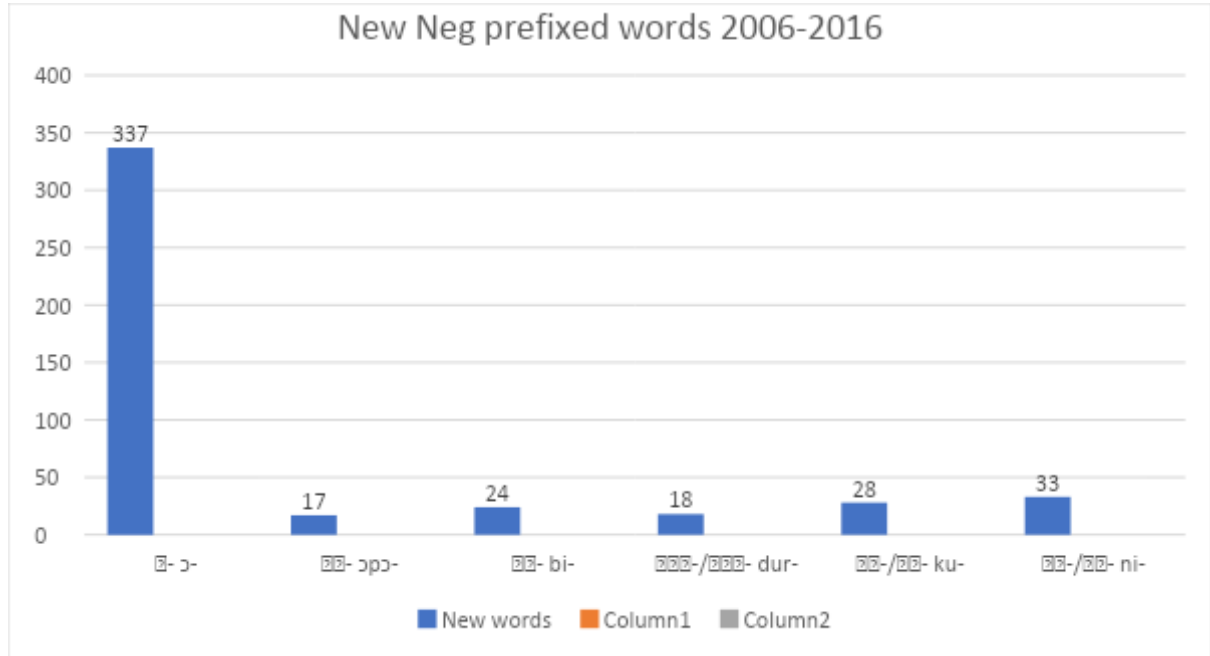


Fig 4.7 Number of negatively-prefixed words in 10 years

Among the six prefixes, *ɔ-* does not only have the highest number of type frequency, it also has the highest number of new words. 337 words have been added for *ɔ-* in the 2016 edition of Hemkosh. Here also, we can see that the number is significantly larger than the rest. While the other prefixes show similar tendencies in new word addition, *ɔ-* alone stands out from the rest in terms of types. After *ɔ-*, *ni-* shows the second highest token frequency for new words followed by *ku-*, *bi-*, *dur-* and *ku-*. In terms of new words also, as we can see in Fig.4.7, the prefixes can be divided into two groups, where *ɔ-* alone can be placed against the other prefixes.

4.4 Interpretation

From the above discussions, we find that in terms of type frequency V in 2006 and 2016, type frequency V of new words, *ɔ-* stands out as the most frequent prefix than the rest in sample B. While discussing the type of new words added in 2016 for *ɔ-* in

section 4.3.2, we have observed how un-prefixed word forms are derived into new words through the prefixation process and they are provided new entries. This can be observed in both editions. The inclusion of different word forms after prefixation helps in exceeding the total number of words for ৩-. Although, other prefixes also have such types of words, the ratio is relatively higher for the prefix ৩-. Adding a few more examples:

অচলন *asalon* ‘Obsolescence’, অচলনীয় *asalonij* ‘Incapable of being introduced or circulated’
 অচলিত *asalito* ‘Not in use or circulation’, অচল *asal* ‘Not moving, fixed’, অচলজ *asalaz* ‘Born in the mountains’ অচলা *asala* ‘Firm, unshakable’

অচপল *asapol* ‘Sobriety, mild and serious’, অচপলা *asapola* ‘calmly, not disturbed by passion’

অক্রিয় *akrija* ‘Inactive, devoid of work’, অক্রিয়া *akrija* ‘Improper action, evil acts’

অখণ্ড *akhand* ‘unbroken’, অখণ্ডন *akhand* ‘That which cannot be split into parts’, অখণ্ডনীয় *akhandnij* ‘Inevitable, irrefutable’, অখণ্ডিত *akhandito* ‘not split up into parts’

অকার্য *akarja* ‘An improper, wrong or vicious act’, অকার্যকৰ *akarjakor* ‘Impractical, futile, serving no purpose’

The varying degree of usage of ৩- has enabled the suffix to utilize the suffix in new situations, including the deliberate invention of new words to define concepts in addition to word forms. This merely serves to confirm that the suffix ৩- is productive since it is a less difficult choice for a negative prefix than others because it is assumed that the meaning will be obvious to most people, as it can express a number of senses. A case in point can be seen in the year 2016, when Nilim Kumar, a prominent Assamese poet invited a wide range of criticism because of his appeal to ‘save poetry’ from *okobita* ‘non-poem’ and *okobi* ‘non-poet’. Although these are not commonly used or widespread in the language, they are not neologisms as well. The two words should have been created relatively recently, as we can see that in Bengali, *okobi* was first recorded in ‘*Bangla Academy Byabharik Bangla Abhidhan*’ June 2011, edited by Muhammad Anamuk Hauk⁷. This happens because, although the bases and their meanings are not

⁷ <https://bn.wiktionary.org/wiki/%E0%A6%85%E0%A6%95%E0%A6%AC%E0%A6%BF>

new, these concepts have been developed to describe situations that have previously been conceptualized in some other ways in the language. *Okobi* means a poet devoid of poetic sentiment or poetic verse. Generally, one who writes poems is considered a poet. Although some poems may have more artistic value and some may lack certain standards, speakers often do not dismiss a poem outright by using a ‘harsh’ or overtly negative term, even when they may criticize it. Because literary works are subjective and something which appears to be unworthy to one person could be valuable to another. However, in recent times, such new terms were coined to distinguish the ‘good’ poems and poets from the ‘bad’ ones. *okobi* was created to label the poets who were devoid of poetic verse. Similar is the case with *okobita*. *Okobita* is not included in the 2016 edition, although there is a chance that it will be in the editions to follow. Other such kinds of words are *obrahmainya* ‘enemy of Brahman’⁸, *Oseujikaran* ‘degreening’ which were not listed in 2006 but added in 2016. *Oseujikaran* seems to be an analogous creation to *abananikaran* ‘deforestation’.

Now, coming to the ranking of the other prefixes, we see that *ni-*, *ɔpɔ-*, *bi-*, *dur-* and *ku-* remain unchanged in both editions. However, in terms of new words, *ɔpɔ-* and *ku-* have exchanged their position, *ɔpɔ-* taking the bottom place of the ranking. In the discussion of the semantic relevance of the prefixes in the previous chapter, we have mentioned that *ku-*, *ɔpɔ-* and *dur-* are adverse prefixes. Although *ku-* precedes *ɔpɔ-* only by 11, the greater number of new word additions for *ku-* than *ɔpɔ-* hints that *ku-* has been used in more words as a negative prefix over ten years. Again, the unchanged position of *dur-* in 2006, in 2016 and in terms of new words tells that it is moderately productive. Between *bi-* and *ni-*, which are discussed as formal prefixes in the previous chapter, *ni-* turns out to be more frequent as a greater number of words are recorded for the same.

4.5 General analysis of Prefixes: Sample A and B

The only suitable method in sample B, which consists of data gathered from two editions of a dictionary, is Type frequency. Consequently, if we compare Type frequency V between samples A and B, we arrive at a few inferences (Table 4.3 and Table 4.5).

⁸ A member of the highest of the four major Hindu castes of traditional Indian society

Although the nature of samples A and B is totally different and there is little similarity between the two, yet both are taken for comparison to see if we spot any findings. While sample B contains words that are established words of the language at some point in time, sample A contains words that are in use as the data are collected from contemporary digital texts. The comparison of them highlights the similarities or differences of frequencies in both.

Now, we see that in both the samples other than the prefixes *ɔpɔ-* and *dur-*, all remain at the same ranking. The ranking of *ɔpɔ-* and *dur-* gets changed from 5th and 3rd in sample A to 3rd and 5th in sample B respectively. In terms of the highest frequency, *ɔ-* occupies the same rank in both samples. Similarly, *ku-* is also found as the least frequent prefix in both.

From here, we can draw two inferences. First, the same measuring techniques or methods have more likelihood of providing the same sort of results irrespective of different data sets.

Second, although the same method is applied in sample A and sample B, in terms of the nature of the samples, both are completely different. Even then, in both the samples, *ɔ-* has the highest number of words. Moreover, *ɔ-* has the highest number of words which is significantly larger than the others in both the samples. In sample B, the 2006 edition has 1383 types and 2016 edition registers 1720 types. The sample A has 208 types for *ɔ-*. It is clear from the comparison that there is a substantial gap between the number of types recorded for *ɔ-* and the other prefixes. This implies that the prefix *ɔ-* must, at least up to the present, be a productive one. Similarly, in both the samples, *ku-* registers the lowest number of types indicating it as the lowest productive prefix.

4.6 Productivity and Role of Semantics

Through various measuring techniques, we currently know where the prefixes stand in terms of productivity from two samples. Now that we have examined the factors that affect the productivity of the prefixes, we can draw a conclusion.

Prefixes largely have semantic significance, which was already covered in the previous chapter. The first four prefixes *ɔ-*, *ɔpɔ-*, *ku-* and *dur-* among the six prefixes listed can also be referred to as adverse prefixes because they indicate the general unfavorability of a circumstance or state. They also signify several and numerous facets of adversity at the same time.

We see that the prefix *ɔ-*, which is more frequent than the others in terms of statistics, does not limit itself to a single sense. It refers to concepts such as dissimilar to the qualities mentioned by the root, the absence of qualities- free from a situation, lacking the thing mentioned by the root, unsuitable from an ethical standpoint, having the quality of being bad, difficulty, inability, unfavorability, etc. *ɔpɔ-* primarily means adversity of a situation or action which is ethically improper. In essence, the prefix *ku-* which denotes an ‘immoral or dishonourable quality’ is the polar opposite of *xu-*. *dur-* means the quality of being bad with the sense of incomprehensibility, which invokes the sense of incapability but can theoretically be attained with the requisite audacity. These are potentially competing prefixes as they are adverse prefixes. There is always a chance that one or two of the competing prefixes will become more productive than the others in the presence of more than one prefixes (Booij, 2012). In this case, *ɔ-* and *ɔpɔ-* seem to be more productive than the others.

The prefix *ku-*, as we have seen, is a polar negative prefix of *xu-*. It appears in a situation that is directly opposite to that indicated by *xu-*, i.e., necessarily negative. In case of *dur-*, the type of negativity is associated with ‘audacity’ of doing the action, again a marked condition like *ku-*. This could be one of the reasons why it is used only in specific conditions and consequently has a lesser frequency of usage compared to *ɔ-* and *ɔpɔ-*. Although *ɔpɔ-* and *ku-* nearly mean the same. i.e., ‘the quality of being bad’, as we have said *ku-* is more critical and direct than *ɔpɔ-*. At times, as a politeness strategy, speakers choose to use indirect methods to save the face of the hearer. In this context, it is more likely that *ɔpɔ-* is a preferred choice to mean the same without appearing too harsh.

The prefixes which are mostly used in formal contexts such as *ni-* and *bi-* have a more creative usage. In many cases, they go through morpho-phonemic changes when they are attached to bases. Therefore, it appears that they are moderately being used in

new word formation. However, amongst these two, we find that the productivity of *ni-* ranges from top to middle in the methods, while in the case of *bi-*, it ranges from middle to bottom, which means that *ni-* is more productive than *bi-*.

It can be established that *ɔ-* is and has the optimum chance of remaining as the highest productive prefix followed by *ɔpɔ-*. The reason is primarily due to the dynamic usage of *ɔ-*. It carries several senses which make it easier for the users to use it for new coinages by avoiding the semantic restrictions that the other prefixes involve. Besides, *ɔ-* is monomorphemic and it does not go through morphophonemic changes while attaching to the bases. This is another probable reason for choosing this prefix over others to avoid bottlenecks, restrictions and impositions.

4.7 Clustering of the prefixes in R

In the earlier sections, a few statistical methods are used to assess the productivity of the prefixes in samples A and B. There, we gained an understanding of the productivity rate for each measuring method for each prefix and learned which prefixes are more productive and how likely they are to be used in future word building. In this part, we will examine how R's clustering function groups prefixes based on the shared characteristics between them. The machine learning algorithm divides them into several clusters using the numerical values of all the measuring methods.

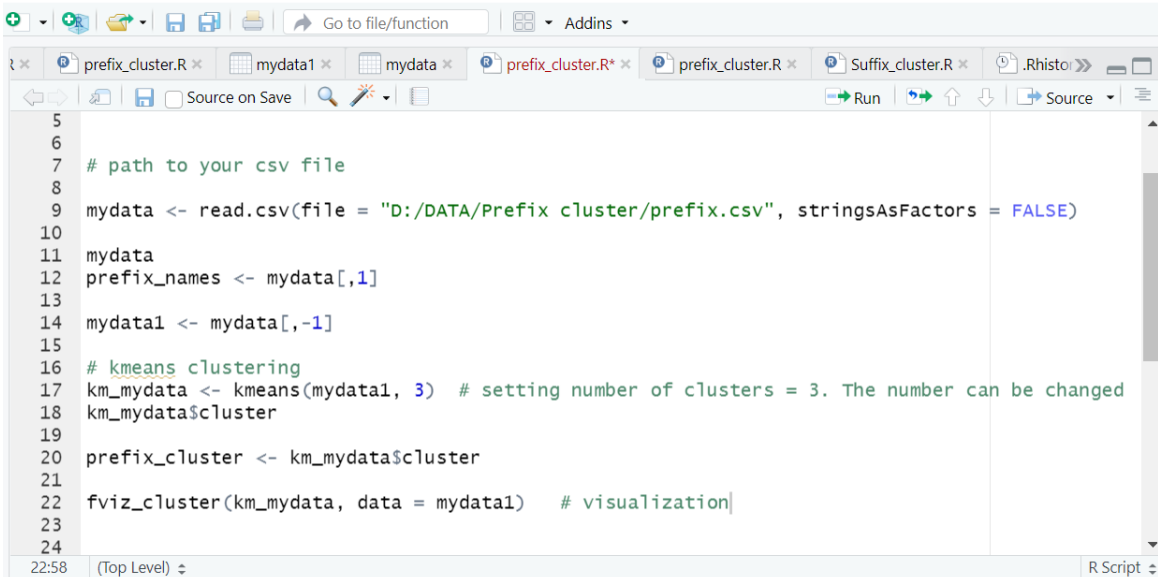
There are variables and observations in R's Clustering analysis. Here, the prefixes are 'variables' or 'clustering variables,'⁹ while the measurements are referred to as 'observations'¹⁰ or 'objects.' Each prefix that is identified as 'variable' contains all values that measure the same underlying attribute, i.e., values measured by a particular method. The measuring values are identified as 'observation' or 'object' against each 'variables' i.e., for each prefix.

However, it is only an experimental test of R in productivity analysis. As clustering in R requires large unstructured data as much as possible, we have excluded sample B from experimentation from it. Based on the values produced by the methods in

⁹ Quantitative characteristics of the objects which is called 'clustering variables'

¹⁰ an observation is a case of the collected data or all values that were measured for the same unit.

example A, we attempt to cluster the prefixes. Fig 4.8 is an image of R scripts that bears the commands to arrive at the results. Fig 4.9 and Fig 4.10 are the visualisations of the clustering of the prefixes in 3 and 4 clusters respectively in the cluster plot.



```
5
6
7 # path to your csv file
8
9 mydata <- read.csv(file = "D:/DATA/Prefix cluster/prefix.csv", stringsAsFactors = FALSE)
10
11 mydata
12 prefix_names <- mydata[,1]
13
14 mydata1 <- mydata[,-1]
15
16 # kmeans clustering
17 km_mydata <- kmeans(mydata1, 3) # setting number of clusters = 3. The number can be changed
18 km_mydata$cluster
19
20 prefix_cluster <- km_mydata$cluster
21
22 fviz_cluster(km_mydata, data = mydata1) # visualization
23
24
```

Fig 4.8 The R-script for the prefixes' clustering¹¹

Below, the prefixes are grouped as per the outcome of R clustering in 3¹²:

Cluster 1: *o-*

Cluster 2: *ni-*

Cluster 3: *opo-, dur-, ku-, bi-*

¹¹ The script is written by Tonmoya Sarmah, a research scholar from the department of Computer Science and Engineering, Tezpur University, Assam, India

¹² The assigned cluster number may change in each running of the command, but grouping is same every time.

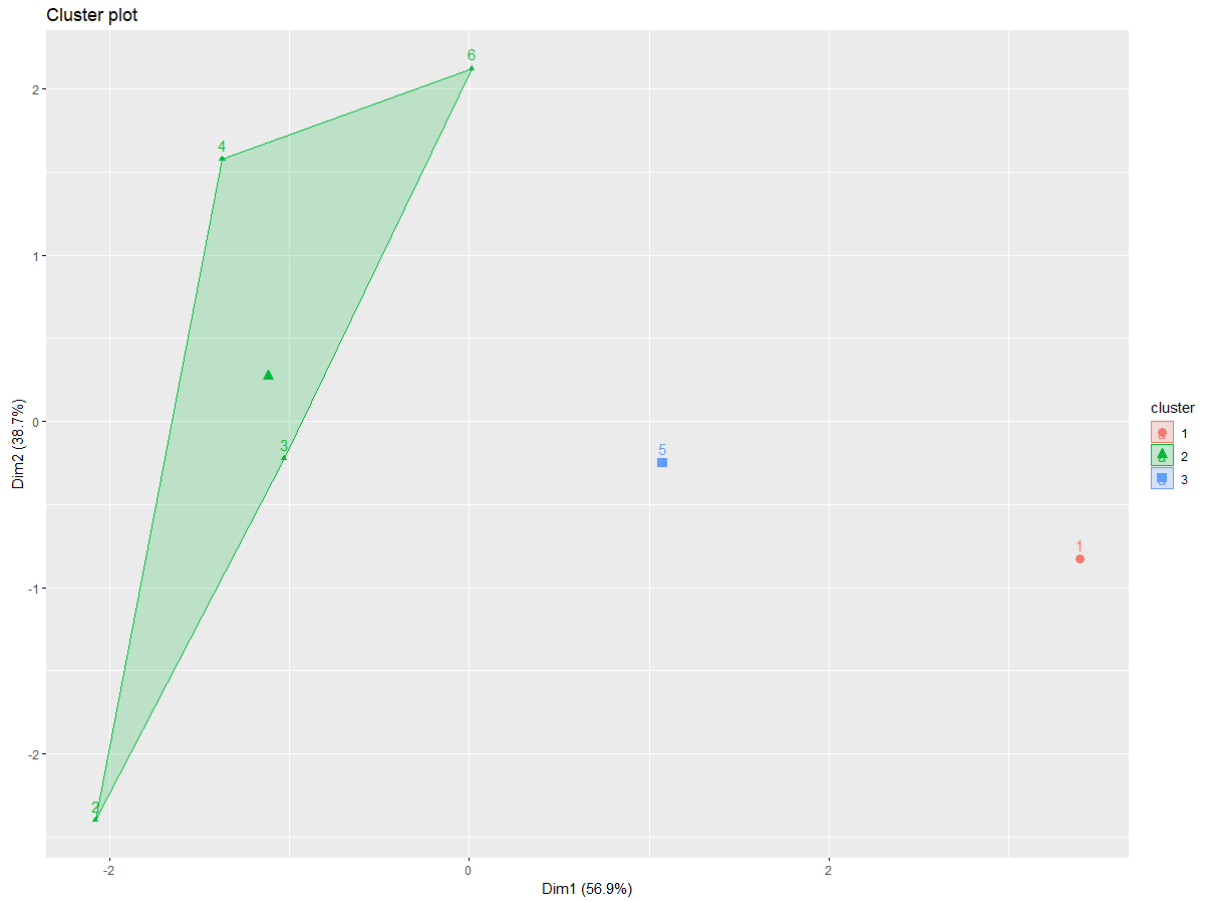


Fig. 4.9 Clustering of the prefixes in 3 clusters by k-means clustering

Now, coming to 4 clusters, the prefixes are clustered the following way:

Cluster 1: *ni-*

Cluster 2: *ɔpɔ-, dur-, ku-*

Cluster 3: *bi-*

Cluster 4: *ɔ-*

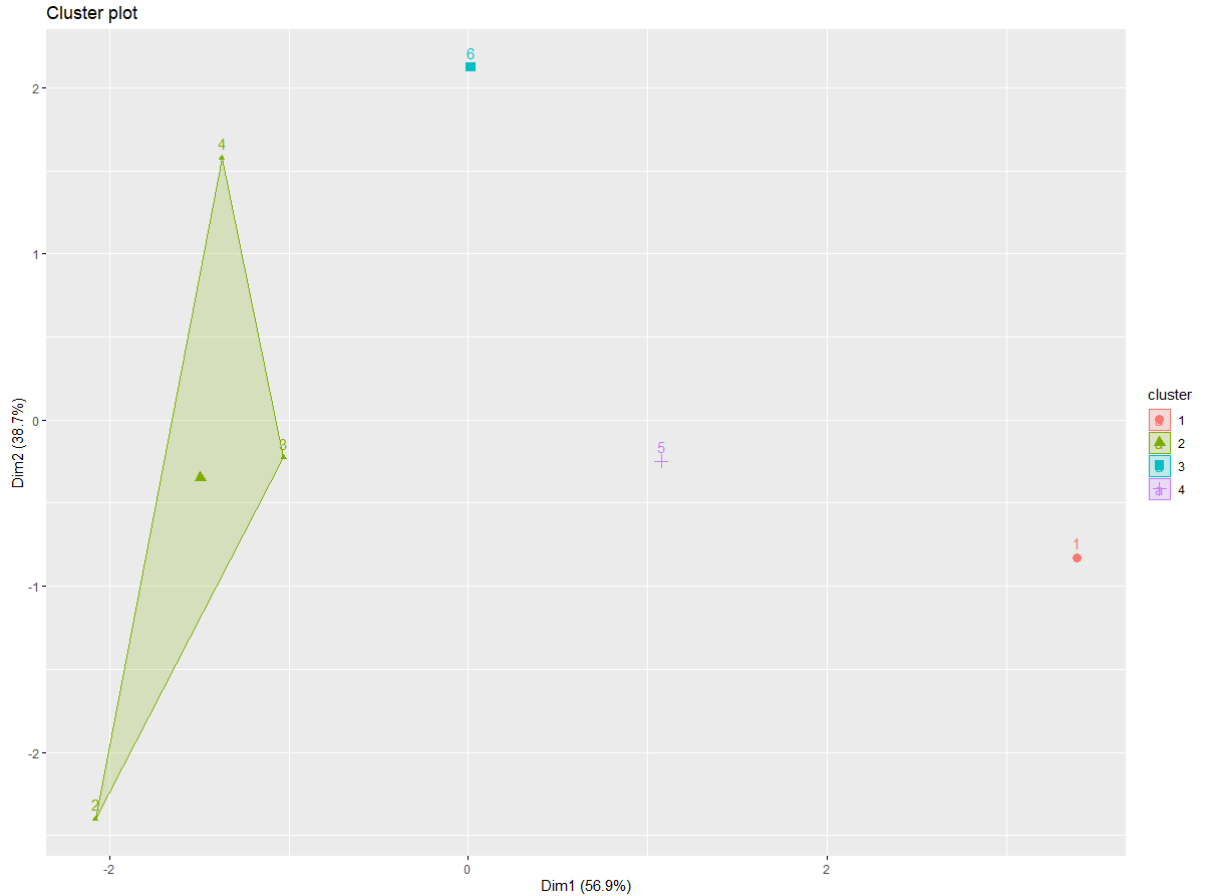


Fig 4.10 Clustering of the prefixes in 4 clusters by k-means clustering

In Fig 4.9, which is clustered in 3, we see that the prefixes *ɔpɔ-*, *dur-*, *ku-*, *bi-* are clustered in one, while *ɔ-* and *ni-* are given two different clusters.

In 4 clusters (Fig 4.10), we see that while the prefixes *ɔpɔ-*, *dur-*, *ku-* form one cluster, and *ɔ-*, *ni-* and *bi-* form three different clusters.

In 3 clustering, where four prefixes are grouped in one, three of them (*ɔpɔ-*, *dur-*, *ku-*) are adverse prefixes and *bi-* is a formal negative prefix. When we proceeded to 4 clustering, we saw that *bi-* is separated from the previous cluster and given a different cluster, while *ɔ-* and *ɔni-* are already in two different clusters.

Now from these two, we find that Fig 4.10 where the prefixes are clustered in 4 can be associated more with our previous findings or analysis. The three prefixes *ɔpɔ-*, *dur-*, *ku-* which are grouped together have semantic similarity as this trio denotes

different aspects of adversity. This undoubtedly had an impact on their numerical values, which in turn led R to group them together based on the features they shared. Again, in the case of *o-*, we see that it is a dynamic and the most productive prefix of all. It includes many aspects of negativity, hence used in various and in more contexts. As a result, it increased the overall number of words providing it the status of a high frequency prefix. This by itself can provide evidence that it owes another cluster. The frequency rates for *ni-* and *bi-* are similar and both have creative usage, which is reflected in their different clustering.