Original research article
UDC 811.111+811.163.41`367.625
DOI 10.21618/fil2327127j
COBISS.RS-ID 138898945

Jasna M. Jurišić Roljić[1]
University of Banja Luka
Faculty of Philology
English Department

PRODUCTIVITY AND FREQUENCY OF SOME LATINATE VERB PREFIXES IN ENGLISH AND SERBIAN

Abstract: *The Latin language was the Lingua Franca of the ancient world, and the precursor of Romance languages. As such, it had a huge influence on European languages, and we can find elements of Latin both in English and in Serbian. In this paper, four Latinate verb prefixes with spatial and temporal meanings have been examined, and those are: trans-, re-, intra-, and post-, and their frequencies and productivities in both languages have been calculated. The analysis was made on the English corpus BNCweb, and the Serbian corpus srWaC, based on the hypothesis that these prefixes would be more frequent and productive in English, which was more exposed to Latin and Romance languages.*

Key words: *verb prefix, frequency, productivity, morphology, corpus linguistics*

1. Latinate verb prefixes in English and Serbian

As languages change and come into contact with other languages, today 'there are no unfamiliar semantic knots among cultural patterns of the world' (Đuliman, 2019:p.122). This can also be applied to verb prefixation, which means that the presence of foreign prefixes with the same semantics, both in English and in Serbian, mirrors the cultural diversity in those languages because most verbs that contain foreign prefixes are loanwords.

Due to the fact that Romance languages have had a significant influence on English, it is no surprise that English contains a considerable volume of Romance, mainly Latinate, affixes. The list of affixes of Greek origin is a bit shorter. Bauer et al., list the following Latinate prefixes with spatial and temporal meanings: *ante-,*

[1] jasna.jurisic.roljic@flf.unibl.org

circum-, contra-, counter-, inter-, meta-, on-, para-, post-, pre-, retro-, sub-, super-, sur-, trans- (2013:p.334), to which Dikson (2014) adds *re-*.

Most of those prefixes can be found in the Serbian language as well, and most of them have been the result of the influence of German, French, and English. Klajn (2002) notes the following verb prefixes with spatial and temporal meanings: *ko-, pre-, re-* and *trans-* concluding that all of them are unproductive in Serbian. Although he does not mention *inter-* it is by all means present in the Serbian corpus.

2. Semantic analyses of the prefixes and the most frequent verbs

The four prefixes were analysed in the BNCweb and srWaC. In order to obtain relevant results, I chose corpora which are POS-tagged, balanced and lemmatised. Using regular expressions, I determined the most frequent verbs containing the four prefixes. The following is a semantic overview of each affix, which will be followed by the corpus analyses.

Inter- is a Latinate verb prefix that has the spatial semantics of the preposition 'between' (Dixon, 2014:p.139), which is obvious in the following examples *interface, interpose, intertwine.* Nowadays, however, it is more used with the meaning 'having an effect on each other' (*interconnect, interact*). Among the most frequent verbs prefixed with inter-, in the BNCweb there are *interpret, interview, interrupt, interfere, and intervene.*

In Serbian, the most frequent verbs are *interesovati, intervenisati, interpretirati, intervjuisati, and internacionalizovati.* Although we can find hybrid verbs in English (those that have an English base and Latinate prefix), that is not the case in Serbian where *inter-* combines with Latinate bases exclusively.

The **trans-** prefix has the spatial semantics of the preposition 'across' and is commonly used with verb and adjective bases, both Romance and English. The most frequent verbs in English are *transfer, transform, translate, transmit,* and *transport.* The first verb with an English base is *transput* and it is 25th on the frequency list. The most common verbs in Serbian with this affix are *transformisati, transportovati, transferisati, transponovati*, and *transcendirati.* None of the verbs from srWaC has a Serbian base.

Although ***re-*** has the spatial semantics of the adverb 'back', today it is mainly used in its temporal meaning "doing the activity of the base again". The prefix is highly used in English and according to Dixon it has also been attached to English bases since the Middle English period (2014:p.169). Yet, the most frequent verbs containing *re-* are those with Romance bases: *require, remain, remember, receive,*

and *return*. *Re-* is among those rare verb prefixes which occur in multiprefixed verbs such as *reimprison, reinfect, reinspect, rerelease*.

As for Serbian, the most common verbs with this affix are *reagovati, registrovati, rekonstruisati, rezervisati*, and *reflektovati.* Since Serbian does not have an affix with the same temporal meaning, *re-* can be ordinarily found with Serbian verb bases such as *reizbacivati, reobrazovati, reimenovati, reobračunati, reotvarati, reupotrebljavati, reuspostaviti.* Under the influence of English, we can find words in Serbian such as *reskenirati, resortirati, restrimovati*.

The prototypical semantics of the affix ***post-*** is spatial and connected with the preposition 'behind', but it is mostly used in its temporal meaning, which is 'after'. English has a rival prefix *after-* that precedes adjective and noun bases. The most frequent verbs prefixed with *post-* in BNCweb are *postpone, postcode, postmark, postdate*, and *postmultiply*. In srWaC, those verbs are *postproducirati, postponovati, postmodernizovati, postprocesirati.* There are also four tokens with the Serbian base: *postprodati*, from which the noun *postprodaja* has been derived.

3. Calculating productivity of the prefixes

The analysis of the four Latinate prefixes was conducted using the resulting frequency values. The productivity of each prefix has been calculated through two methods: productivity 'in the narrow sense' and the Q-method, after which the results have been compared.

Chitashvili and Baayen (1993) have offered a formula for calculating morphological productivity which includes the number of hapaxes[2], because hapaxes are closely related to the formation of neologisms. According to the authors, productivity 'in the narrow sense' shows one component of morphological productivity, which is the probability of occurrence of new types (p.186):

$$P = n_1 : N$$

In this formula ***P*** stands for productivity, n_1 is the number of hapaxes and N represents the total number of tokens[3] of complex words in a sample.

The Q-method of calculating productivity relies on the ratio of type[4] frequency and token frequency. Chitashvili and Baayen (1993) developed it for cases in which the number of hapaxes is approximately half the total number of lemmas analysed

[2] Hapax legomenon (Old Greek ἅπαξ λεγόμενον), a linguistic unit that is recorded as having been used only once.

[3] Tok**e**n is an individual occurrence of a linguistic unit in a corpus.

[4] Type is an abstract category of linguistic item or unit.

in a corpus which is large enough. This is particularly suitable for analysing affixes, which usually show such distribution of hapaxes. In the following formula:

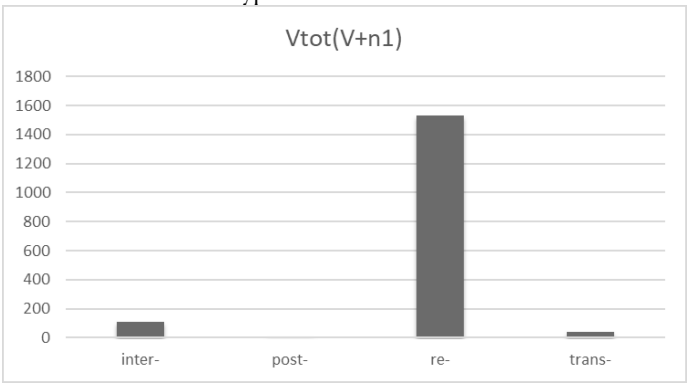
$$Q = V/N$$

Q stands for productivity that is calculated as a ratio of type frequency (*V*) and token frequency (*N*).

4. Analysis results

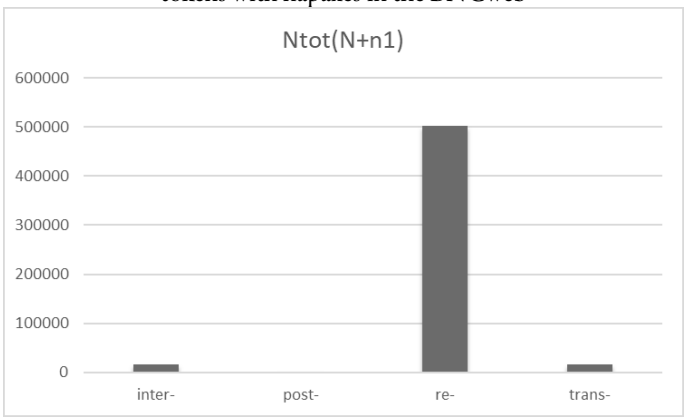
After extracting data from BNCweb using the *NoSketchEngine*[5], and from srWaC using regular expressions, I obtained frequency lists for each prefix in which verbs were grouped according to lemmas. The lists were saved in the CSV format and thoroughly cleaned in order to obtain reliable data which included types, tokens, and hapaxes. Applying the formulae for 'productivity in the narrow sense' and the Q-method, the productivity of each prefix was calculated. Finally, all the results are visually presented in the following graphs and tables.

Graph 1: Comparing prefixes inter-, post-, re-, and trans- in terms of total number of types in the BNCweb.



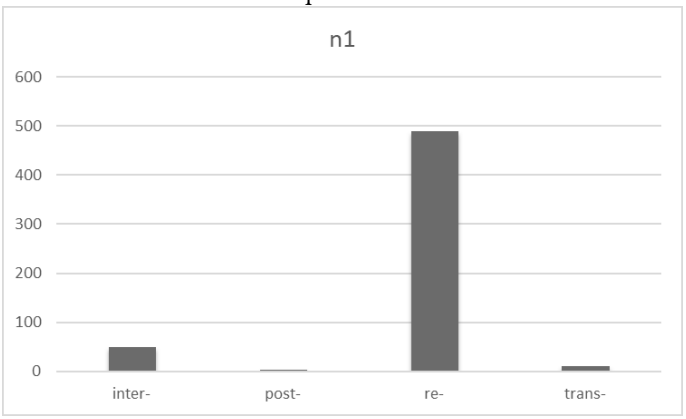
[5] Link to *NoSketchEngine*: https://nlp.fi.muni.cz/trac/noske

Graph 2: Comparing prefixes inter-, post-, re-, and trans- according to the number of tokens with hapaxes in the BNCweb



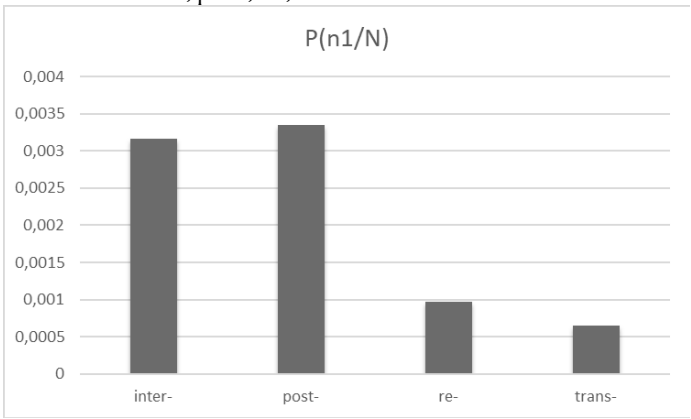
Graphs 1 and 2 show that the prefix re- is considerably dominant in frequency compared to the rest of the prefixes analysed when observing both types and tokens. The prefix *inter-* has a higher frequency of types, while *inter-* and *trans-* are similarly frequent.

Graph 3: Comparing prefixes inter-, post-, re-, and trans- according to the number of tokens with hapaxes in the BNCweb.



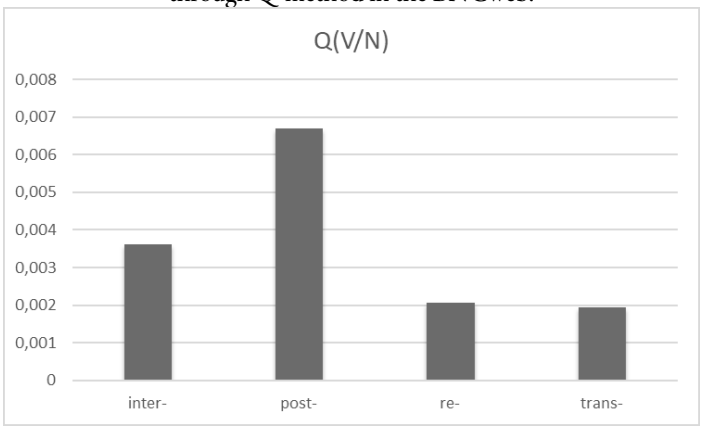
Graph 3 clearly shows the relationship of frequency and hapax number: a prefix that occurs multiple times in the corpus will have a higher number of hapaxes, which are an indicator of productivity.

Graph 4: Productivity 'in the narrow sense' of prefixes inter-, post-, re-, and trans- in the BNCweb



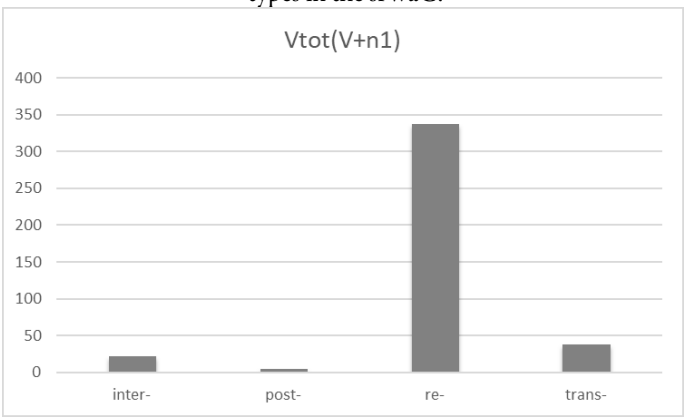
When calculating productivity of the morphological process of verb prefixation through the two formulae, both methods prove that the prefix post- is the most productive in the BNCweb corpus. It is followed by *inter-*, while *re-* is slightly more productive than *trans-*.

Graph 5: Productivity of prefixes inter-, post-, re-, and trans- through Q-method in the BNCweb.



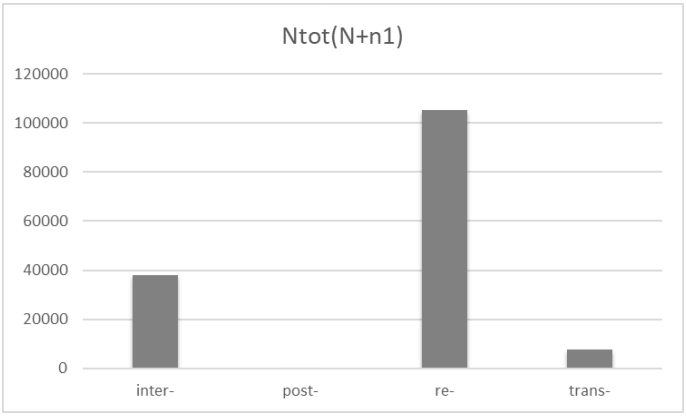
The graphic presentation of the Serbian corpus analysis shows the following results:

Graph 6: Comparing prefixes inter-, post-, re-, and trans- in terms of total number of types in the srWaC.

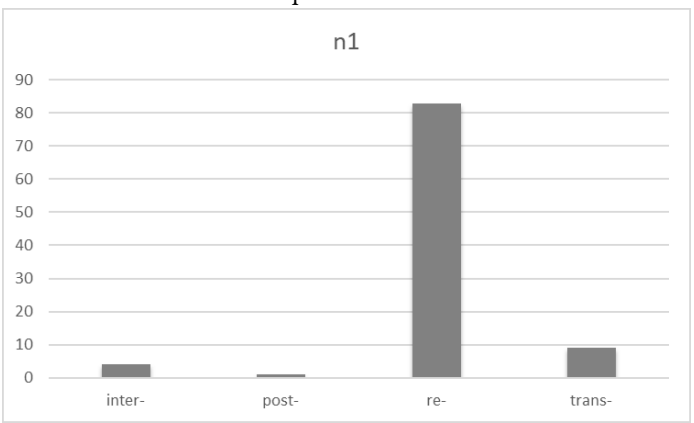


If we observe the total number of types and tokens of those prefixes in Serbian, *re-* is far more frequent than *trans-, inter-*, and *post-*. As for the total number of tokens, *inter-* is more frequent than *trans-*, while values for the prefix *post-* are so low that they cannot be graphically presented because there are only 5 tokens in the Serbian corpus.

Graph 7: Comparing prefixes inter-, post-, re-, and trans- in terms of the number of tokens with hapaxes in the srWaC.



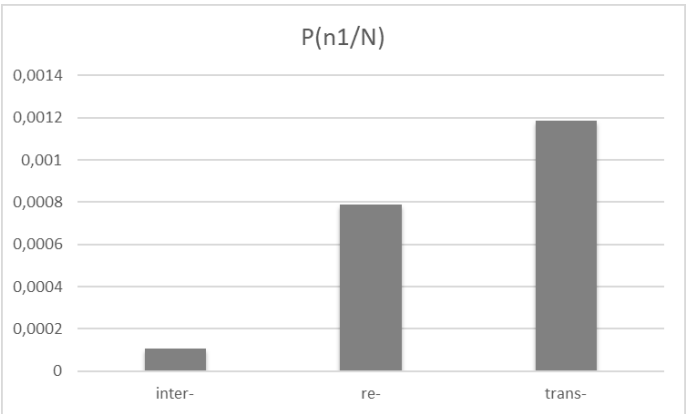
Graph 8: Comparing prefixes inter-, post-, re-, and trans- in terms of number of tokens with hapaxes in the srWaC



Since the prefix *re-* is the most frequent, it has the largest number of hapaxes.

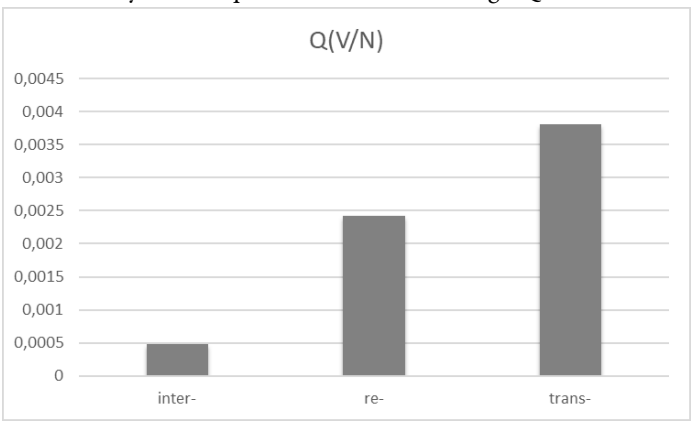
The values for the affix *post-* significantly deviate from the rest of the data in the srWaC, which is why this affix is considered to be a statistical outlier, and it has not been included in productivity calculations. Analysis of the productivity 'in the narrow sense' shows that the most productive prefix in the Serbian corpus is *trans-*.

Graph 9: Productivity 'in the narrow sense' of inter-, re-, and trans- in the srWaC.



The same conclusion is reached when calculating productivity using the Q-method, which does not include hapaxes (Graph 10).

Graph 10: Productivity of inter-, post-, re-, and trans- through Q-method in the BNCweb.



As all the previous graphs have shown, in the morphological process of verb derivation with the four prefixes, the most productive affix is *post-* in the BNCweb, while in the srWaC it is the affix *trans-*. The prefix *re-* is the most frequent one in both corpora.

Table 1: Overall results of analysis of inter-, post-, re-, and trans- in the BNCweb

Prefiks	Vtot(V+n1)	V	Ntot(N+n1)	N	n1	P(n1/N)	Q(V/N)
inter-	107	57	15851	15801	50	0.00316436	0.00360737
post-	12	8	1198	1194	4	0.00335008	0.00670017
re-	1530	1041	502946	502457	489	0.00097322	0.00207182
trans-	40	30	15434	15424	10	0.00064834	0.00194502

Table 2: Overall results of analysis of inter-, post-, re-, and trans- in the srWaC.

Prefiks	Vtot(V+n1)	V	Ntot(N+n1)	N	n1	P(n1/N)	Q(V/N)
inter-	22	18	37840	37836	4	0.000105719	0.000475737
post-	5	4	11	10	1	~~0.1~~	~~0.4~~
re-	337	254	105199	105116	83	0.000789604	0.002416378
trans-	38	29	7607	7598	9	0.001184522	0.003816794

5. Conclusion

The results prove that the hypotheses was correct: the number of Latinate prefixes with spatial and temporal meaning is larger in the British corpus than it is in Serbian. Prefixes *inter-, trans-, re-*, and *post-* are more frequent and more productive in the English language. The analysis of the English corpus proved that the Latinate prefixes analysed in this research tend to attach to English bases, which is not surprising considering the fact that English was under the strong influence of Latin, Norman and French.

In the Serbian corpus, only the prefix *re-* combines with Serbian bases in a statistically significant quantity. Also, a modern trend in Serbian indicates the increase of Latinate prefixes as a result of the influence of English, and this trend would surely be a very interesting topic for future research.

References

1. Bauer, L., Lieber, R. & Plag, I. (2013) *The Oxford Reference Guide to English Morphology*. Oxford, Oxford University Press.
2. Chitashvili, R. J. & Baayen, H. (1993) Word Frequency Distributions. In Hrebicek, L. & Altmann, G. (eds.) *Quantitative Text Analysis.* Trier, WVT, 54–135.
3. Dixon, R. M. W. (2014) *Making New Words: Morphological Derivation in English*. Oxford, Oxford University Press.
4. Đuliman, S. (2019) Semantička transpozicija humora u prevodu TV serije "Blackadder" na bosanski jezik. *DHS Časopis Filozofskog fakulteta u Tuzli.* 3, 115–130.
5. Клајн, И. (2002) *Творба речи у савременом српском језику. Део 1, Слагање и префиксација.* Београд, Завод за уџбенике и наставна средства, Институт за српски језик САНУ – Нови Сад, Матица српска. / Klajn I. (2022) *Tvorba reči u savremenom srpskom jeziku. Deo I, Slaganje i prefiksacija*. Beograd, Zavod za udžbenike i nastavna sredstva, Institut za srpski jezik SANU – Novi Sad, Matica srpska.

Corpora

1. British National Corpus http://bncweb.lancs.ac.uk/ [Accessed in February 2023]. http://www.natcorp.ox.ac.uk/
2. Ljubešić, N. & Klubička, F. (2016) *Serbian web corpus srWaC 1.1.* Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1063 [Accessed in February 2023].

Jasna Jurišić Roljić
Univerzitet u Banjoj Luci
Filološki fakultet
Katedra za anglistiku

PRODUKTIVNOST I FREKVENTNOST NEKIH GLAGOLSKIH PREFIKSA LATINSKOG PORIJEKLA U ENGLESKOM I SRPSKOM JEZIKU

Rezime

Latinski jezik bio je jezik sporazumijevanja starog svijeta, ali i preteča današnjih romanskih jezika. Kao takav, imao je veliki uticaj na sve evropske jezike, pa elemente latinskog nalazimo i u engleskom i u srpskom jeziku. U ovom radu ispituju se četiri glagolska prefiksa sa značenjem vremena i prostora koja su latinskog porijekla: *trans-*, *re-*, *intra-* i *post-*, izračunava se njihova frekventnost i produktivnost. Ispitivanje se vrši na korpusu engleskog jezika BNCweb i korpusu srpskog jezika srWaC, a polazi se od pretpostavke da će ovi prefiksi biti brojniji i produktivniji u engleskom jeziku, koji je u većoj mjeri bio izložen latinskom i romanskim jezicima. Korištenjem formula za *produktivnost u užem smislu* i *Q-metodu*, izračunata je produktivnost navedenih prefiksa, a zaključci o frekventnosti izvedeni su nakon temeljitog prečišćavanja podataka ekstrahovanih iz korpusa. Analiza prefiksa potvrđuje početnu pretpostavku, a u oba korpusa izdvaja se prefix *re-* kao dominantno najučestaliji.

▸ ***Ključne riječi:*** glagolski prefiks, frekventnost, produktivnost, morfologija, korpusna lingvistika.

Preuzeto: 23. 3. 2023.
Korekcije: 9. 5. 2023.
Prihvaćeno: 12. 5. 2023.