ЖУРНАЛ ФІЗИЧНИХ ДОСЛІДЖЕНЬ
т. 11, № 1 (2007) с. 6–21

JOURNAL OF PHYSICAL STUDIES
v. 11, No. 1 (2007) p. 6–21

SYNONYMOUS STRINGS LENGTH AND CONSTITUENTS' PLACEMENT IN ENGLISH DEVERBAL WORD-FORMATION: THE PHYSICIST'S CONTRIBUTION TO DIACHRONIC LEXICOLOGY

M. Bilynskyi
Ivan Franko National University of Lviv, Department of English Philology,
1, Universytetska St., Lviv, UA–79000, Ukraine,
e-mail: bislo@ukrpost.ua
(Received August 3, 2007)

Presented to Professor Ivan Vakarchuk on the occasion of his 60^{th} birthday.

Diachronic onomasiology is concerned with two issues: the availability of lexical resources at specific moments of history and complementation of these resources over time. A peculiar feature of such a complementation is the unevenness of the expansion in various sections of the lexicon. The paper focuses on the heuristic potential for historical onomasiology of the formalism suggested by Professor Vakarchuk to assess the mean values of similarity in the sequential placement of constituents within synonymous strings of specific lengths. The rise of the common-root word-forming families initiated by synonymous verbs makes two kinds of bilateral comparisons plausible: strings of verbs can be compared with strings of deverbatives of a specific categorial affiliation or strings of one class of deverbatives are juxtaposable with those of another class. Both kinds of comparisons can be carried out on the material of the entire lexicon or its various cross-sections providing an insightful application to the heuristic potential of the matrix as a tool of sets analysis in historical lexicology.

Key words: synonymous strings, constituents' sequencing, matrices of temporal similarity, mean similarity calculus, strings' lengths

PACS number(s): 01.90.+g

I. INTRODUCTION

Recovering lists of synonymous words available to the previous generations of speakers allows us to look into the minds of those who spoke the language in the process of its evolution. In reconstructing relations of lexical synonymy our knowledge about the earliest uses of words from the dating of their textual prototypes (first quotations) is combined with the lexical composition of the semantic fields.

In the known attempts at this kind of analysis the semantic fields' composition is accepted as given in *Roget's Thesaurus* [1]. The latter provides onomasiological areas whose constituents then are to be rearranged diachronically in the order of the decreasing age (but increasing year dating) of their earliest attested uses according to the *Oxford English Dictionary* [2].

I suggest that a relevant object of historical semantics is also recoverable from *strings of lexemes* available in the dictionary of synonyms [3]. A sample material for this study was taken from *Webster's New World Thesaurus* [4]. In the latter, as in most dictionaries of this kind, the placement of the constituents is non-alphabetic and, hence, intuitively relevant. The multiplicity of dictionaries of synonyms solidifies the respective epistemological perspective of the developed methodology.

The results of the application of word-formation rules or the respective lexical "consequences" that show up as penetrations of ready-made items of deverbal origin in the conditions of inter-language mix (these would have been motivated "dynamically" could such rules have been applied in a "natural" way to derivational bases) confront us with two problems, namely the *succession of the string's constituents* in diachrony and *the pace of the string's formation* over time.

A study of the temporal positioning of constituents in synonymous strings of parent verbs and/or their common-root derivatives lies at the cross-roads of the synonymic and derivational potentials of the lexicon. It also constitutes a testing ground for statistical tools and even ways of thinking from outside historical linguistics that can be of use in processing diachronic lexical data.

II. CONSTRUING THE HISTORICAL THESAURI OF SYNONYMOUS STRINGS

This study is based on the earliest quotations of verbs and deverbatives according to the *Oxford English Dictionary*(*OED*), of which the second electronic CD-ROM edition, version 3, was used [5]. A derivative is taken as documented when it is attested in the *OED*. Out of the instances of double relatedness of a coinage to the homonymous verb and noun the one that is clearly motivated by the verb was taken into account.

In most cases the date of the earliest attestation, or diachronic textual prototype, of a lexeme is given clearly. However, there are several peculiarities. Dating approximation marked in the *OED* by *a (ante)* or *c (circa)* were accepted as precise. Century dating was replaced by the

next *OED* attestation of the word in question or, failing that, it was substituted by the last year of the century. In the rare cases of period dating the earlier date was accepted.

A string of synonyms from the contemporary thesaurus is an ordered list of lexemes initiated by the headword (string dominant). In the historical thesaurus the constituents of this string are rearranged to make up a chronological sequence.

In cases when two or more constituents are dated in the same year they can be placed in a linear succession only arbitrarily, for instance in the alphabetical order, or in their present-day sequence. We chose the former solution which seems easier although either of the two amounts only to a conventional limitation. Such cases, luckily for the experiment, are pretty marginal statistically for the general set, which may not be the case for its specific partitions (see Fig. 13 in section 5 below).

The diachronic strings initiated by the oldest constituent that is the dominant in one or more (owing to polysemy) contemporary strings and a constituent in one (or more) other strings make up a cluster of strings. Thus in the historical thesaurus there can be a larger number of strings initiated by a given word than in the present-day thesaurus.

Conversely, some words that open up synonymous string(s) in the contemporary thesaurus fail to initiate an historical string of synonyms. It happens when at least one other constituent in such (a) string(s) reveals an earlier dating than the string's present-day dominant and when there is not a single string within which the given word is the oldest constituent.

The reason for this may also be merely technical. Then apart from the word in question the string contains at least one other constituent dated in the same year but preceding it alphabetically.

The basis for referring a deverbative into a category is its paraphrase. In the developed electronic lattice a separate position is allotted to the verb and each of the agreed categories of derivatives: action nouns (N_{action}); action nouns admitting of factitive lexicalisation ($N_{\text{action/...}}$), agent nouns (N_{agent}), patient nous (N_{patient}), adjectives (A), lexicalized present participles (P_{present}), passive modal adjectives ($A_{\text{passive modal}}$), lexicalized past participles (P_{past}).

Each class of adjectives and participles gives rise to secondary deverbal adverbs and nouns, respectively $D < A$; $N < A$; $D < P_{\text{present}}$; $N < P_{\text{present}}$; $D < A_{\text{passive modal}}$; $N < A_{\text{passive modal}}$; $D < P_{\text{past}}$; $N < P_{\text{past}}$.

This basically arbitrary succession of sixteen classes of deverbal coinages is concluded with factitive nouns ($N_{.../\text{factitive}}$). Most of the latter are epidigmatic lexicalisations of the second of the singled out groups of action nouns. But providing at the same time a reason for the said parsing there are suffixal deverbal coinages unrelated to one-word action nouns among this category as well.

Two kinds of variance are to be dealt with opening up new cross-sections of the database and variant applications of the software. Multiple suffixes within a common category of deverbatives can be treated as separate manifestations of the respective category. Variant categories can make up aggregate categories.

A derivative, like its parent verb, concatenates lexemes of the same categorial affiliation provided no constraint was imposed on the respective parent verbs. The derivational constraint as a diachronic entity holds as long as the respective parent verb fails to produce a coinage. The newly coined derivative either *"joins"* the already existing counterpart(s) of identical categorial affiliation derived from the respective verb(s) synonymous with its parent verb or *"begins to wait"* for such counterparts if it happens to be the first coinage of a given categorial, or, eventually, suffixal affiliation.

The necessary and sufficient condition for the chronological sequence of deverbatives occurs when at least two verbs from the parent string are involved in the coining of their common-root coinages. The issue of temporal succession of derivatives holds for the cases when the historical dominant itself does not produce (sometimes only for a time) a coinage but two other constituents of the string do.

The verb that is the earliest constituent in an historical string may produce a coinage that is not the oldest one in the respective string. Thus, the oldest constituents in the strings of varied categorial affiliation should not necessarily be united by the common-root between themselves or with the oldest constituent in the parent string.

In principle, an arbitrary constituent of the parent string may take up the first chronological position in any of the derived strings. Likewise, the ordinal positions in the two compared strings may fall for the convergent or divergent root(s).

When a string of verbs gives rise to more than one string of coinages of varied categorial affiliation a comparison can be drawn not only between the verbs and an arbitrary attested set of such strings but also between each pair of the derived strings themselves on condition that such a pair is made up of the results of the respective word-formation rules (or consequences of contact penetrations) applied (or, respectively, referred) to at least two verbal stems.

III. LOOKING FOR A FORMALIZED FRAMEWORK

I introduce the notion of *chronotropism* of the compared string in respect towards the string that is taken for the reference point in such a comparison. Like similar coinages with the component *-tropism,* chronotropism implies a twist in the structure under the influence of external factor(s). Hence the essence of our problem is the succession of constituents from the reference point string preserved in the compared string.

In order to estimate the (dis)similarity of two strings of lexemes characterised by re-categorisation I examine relative chronological placement of the constituents that share the same roots in these strings. The succession of

the appearance of constituents in the compared strings is presented in a *matrix*[1]. The string that is the reference point of comparison is placed in the matrix row. The compared string is put in the matrix column.

The squares in the matrix are marked with pluses when there is similarity in the strings' constituents succession. It occurs when the ordinal number of the *i-th* constituent from the column is larger than the ordinal number(s) of the row constituent(s) located leftwards of its common-root counterpart and, conversely, when it is smaller than the ordinal number(s) of the constituent(s) located to the right of its common-root counterpart in the row string. If these prerequisites are not present, there is dissimilarity in the strings constituents' succession and the respective matrix squares are marked with a minus (Fig. 1).

The matrix is symmetric. The row and column coming together at the matrix diagonal reveal an identical distribution of pluses and minuses. The upper and lower triangles of the matrix coincide.

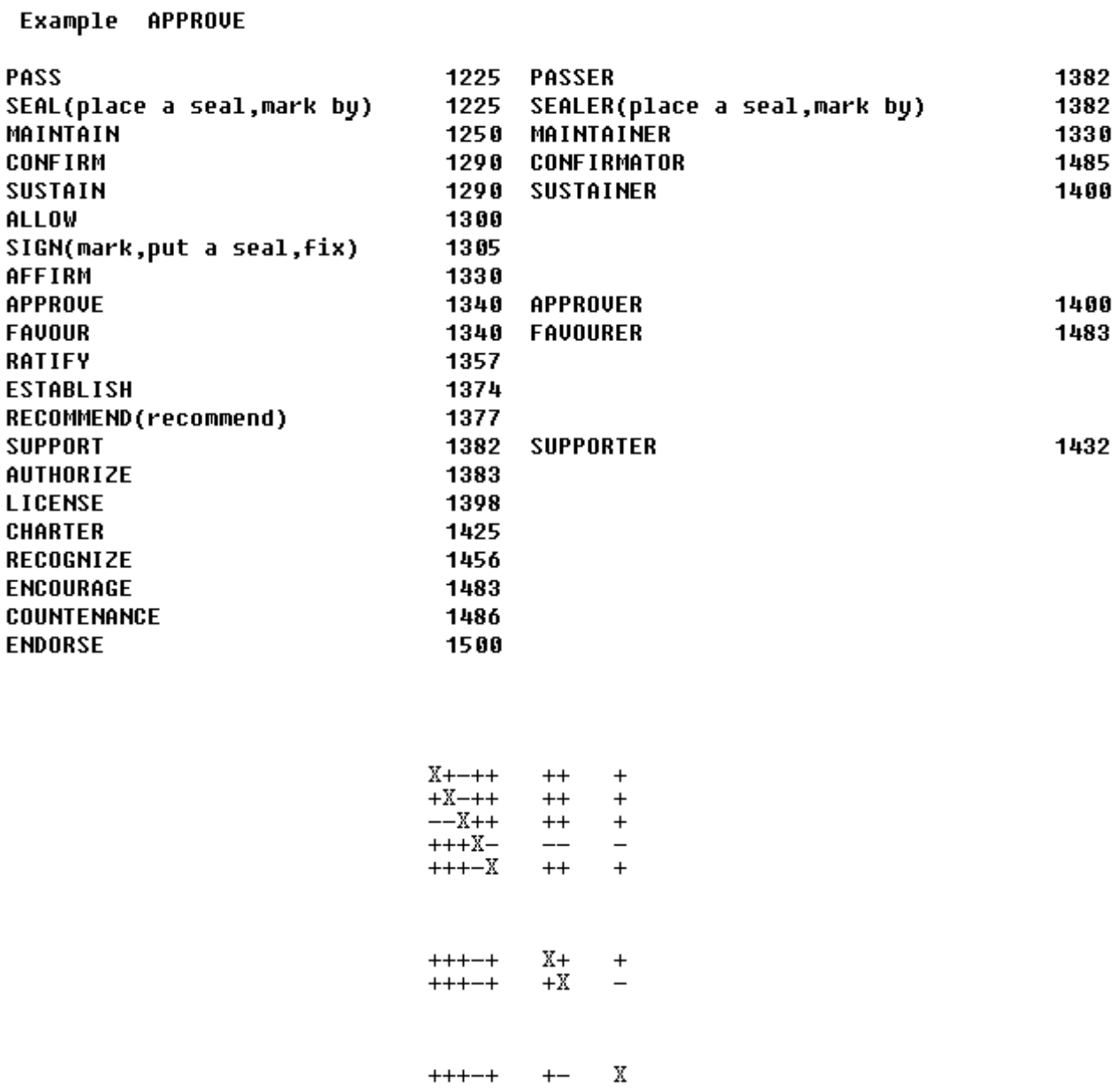


Fig. 1. Temporal similarity of the expansion of two strings in diachrony: matrix row — verbs; matrix column — agent nouns.

[1]To our knowledge, matrix analysis has not yet been used in diachronic lexicology. Cf. a survey of its application to feature analysis in synchronic lexicology in [6].

For reasons of convenience the strings to the construed matrices are written as two vertical lists of words (transforming the directions "to the left" and "to the right" into "up" and "down", respectively) with the date of the earliest *OED* attestation following each constituent. The string that makes up the matrix row is the reference point of the analysis. Hence it is the first to be put down in the illustration as the left hand-side list of words. The string that makes up the matrix column is the compared string. That is why I write it after the row string as the right hand-side list of words. In respect to traditional matrix placement this string is indeed a column but written to the right of the row string.

The described procedure is based on the comparison of the constituents' ordinal positions in the string that is put at the matrix column with the ordinal positions of the constituents of the string written in the matrix row. In the example, the column was attributed to derivatives and the row to verbs. However, it is possible to exchange the location of the respective strings and *invert* the matrix (Fig. 2). In that case, we compare the expansion in the string of verbs (right-hand side list) with that in the string of their derivatives (left-hand side list). Thus the string of coinages in the matrix row is written in the chronological order of its constituents. The string of verbs in the matrix column is written in the order of their common-root deverbatives, i. e. non-chronologically.

Example RECEIVE

RECEIVABLE	1382	RECEIVE	1300
ACCEPTABLE	1386	ACCEPT	1360
ADMITTABLE	1420	ADMIT	1413
TAKEABLE	1449	TAKE	1100
SEIZABLE	1461	SEIZE	1290
INHERITABLE	1470	INHERIT	1388

```
X++--+
+X+--+
++X---
---X++
---+X+
++-++X
```

Fig. 2. Temporal similarity of the expansion of two strings in diachrony: matrix row — passive modal adjectives; matrix column — verbs.

As the placement of the strings in the matrix row and column is exchangeable the quotas of pluses and minuses remain intact but their location in the matrix changes depending on which of the strings is put in the matrix row or column. Although in either case the comparison is bilateral only one of the strings can be a reference point.

The length of the string in the row of the matrix determines the matrix dimension. When we compare strings with a motivational relationship between their common-root constituents it is more convenient to place the motivating string in the position of the matrix column. In this way we avoid empty rows/columns in the matrix since there are no attested coinages with unattested common-root bases in the evolution of lexicon at large. This, however, would be too strong an assumption when applied to specific moments in lexical history in view of back-derivation when a coinage subsequently reveals an inverse transpositional time in respect towards its motivating base.

The placement of the motivating string in the column though shortens its representation in the construed matrix as only those bases that give rise to the respective coinage are forming the column. For the purpose of establishing the extent of imitative logic in sequential ordering of the column string to that in the row string this problem is irrelevant but for the problem of assessing the scope of that logic against the background of varied lengths of row strings it is of consequence.

When both the compared strings are deverbal at the same or adjacent tiers of derivation (with the exception of adjectives/participles and their derived adverbs or nouns where the previous logic holds) strings' inversion has no effect on whether there are empty positions in the matrix.

The constituents of the column string reveal a correlative non-motivational relationship with their common-root counterparts in the row string but the row string may produce empty columns owing to its extra length arising due to the constituents that do not happen to demonstrate such a relationship. In the split of the material to chronological layers this situation is sometimes just temporary.

However, the filling of all the rows/columns in the matrix may be unrelated to matrix inversion. This situation holds when there are no *derivational* constraints imposed on any of the constituents of the parent string. Then all

the filled-in positions in the strings of primary deverbal derivatives compared with strings of verbs and all the filled-in positions in strings of secondary deverbal coinages when compared with their motivating strings of adjectives/participles and ultimately verbs are refutable only by intermediate diachronic evidence.

Finally, there are no empty rows/columns in the matrix when in the compared strings of deverbatives with a non-motivational relationship between them derivational constraints affect the same stems.

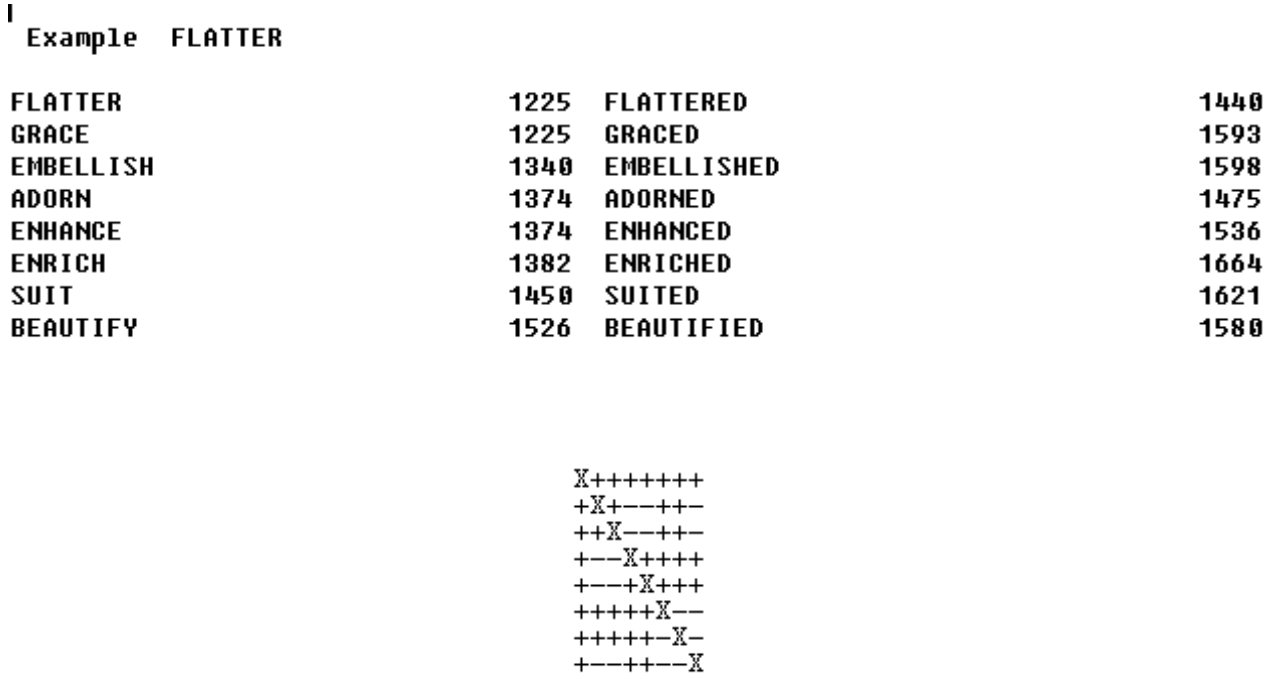


Fig. 3. Temporal similarity of the expansion of two strings in diachrony with same year attestations in the matrix row string.

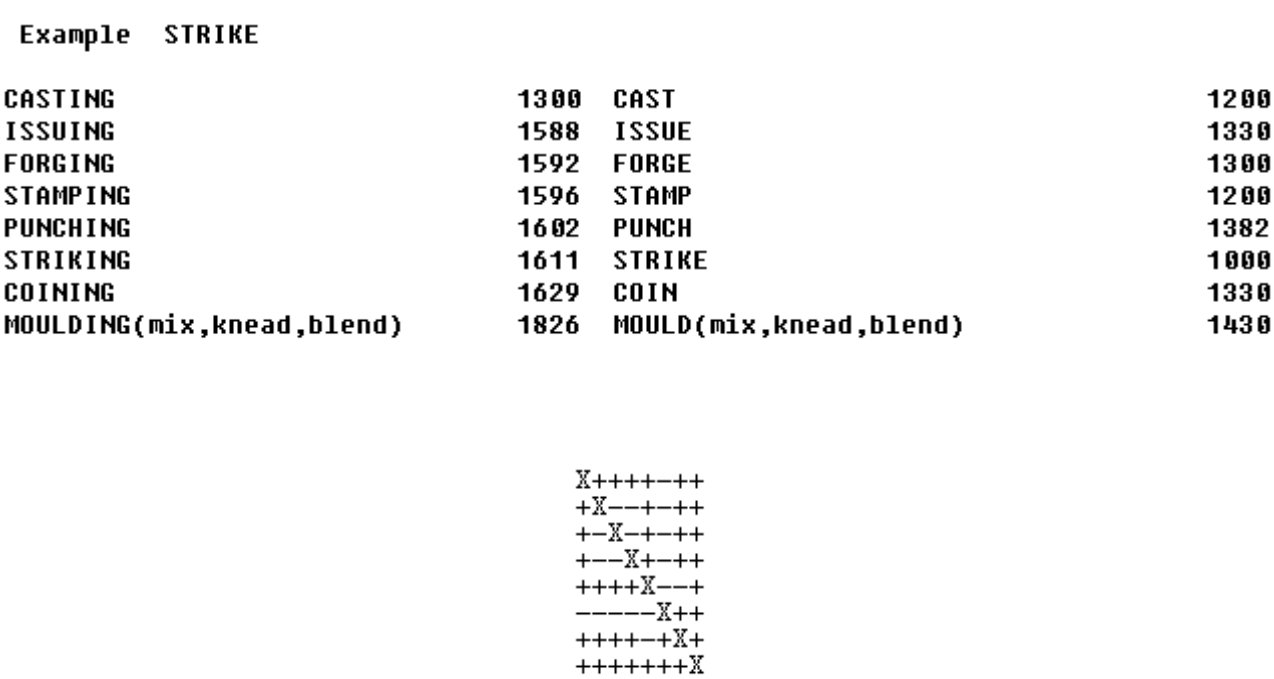


Fig. 4. Temporal similarity of the expansion of two strings in diachrony with same year attestations in the matrix column string.

The issue of the placement of same year constituents within the row string is resolved by putting them in the alphabetical sequence (Fig. 3). As the constituents positioning in the column string follows that from the row string the placement of the same year constituents in it follows the placement of their counterparts in the row string (cf. the difference in the succesion of counterparts of constituents 1 and 4 on one hand and 2 and 7 on the other in the matrix column, i. e. the left hand-side list on Fig. 4 of which only the first pair is alphabebtcal). The placement of the constituents attested in the same year in the column string is generally free of the alphabetical succession unless the respective counterparts of these in the row string reveal same year attestations and thus are placed alphabetically.

The comparisons of strings of two categorial affiliations for their common-root constituents' mutual placement forms a respective cross-categorial domain. For the taxonomy of such domains each of the eighteen categories of lexemes from the deverbal word-forming families, including the verb, is juxtaposed with the remaining seventeen classes of words.

This gives 306 outcomes. They are filled by the total of over thirty-five thousand synonymous strings. All told cross-categorial domains within deverbal derivational families yield several hundred thousand matrices of chronotropism. The strings' constituents mutual placement is visible in well over a million positional comparisons manifested in the filled-in matrix squares.

These numbers will be even several fold higher when we take into account the splitting of the lexicon of deverbal derivational families into respective sub-bases obtainable from its multiple stratification on chronological, etymological or thematic grounds. The given statistics is based on a single chosen dictionary opening up a multiplication prospect by the number of dictionaries of synonymous strings available for English and yet another version of their chronological rearrangement for the period part-base recoverable from the Middle English Dictionary. In one word, the empirical reality behind such comparisons is really vast.

IV. WHAT TO DO WITH THESE MATRICES?

A corpus of matrices for a specific cross-categorial domain is subjected to an n-partite division by the degree of the column string sequential similarity with the row string. The numeric power for each section of this division is established with the possibility of downloading individual strings illustrations, complete sets of examples for specific sequential similarity values intervals as well as their matrices of chronotropism.

The composition of the construed corpora of matrices will be the same only at identical lengths of both strings. Varied lengths at the row-column interchange produce groups of matrices whose composition is not identical. The choice of cross-categorial domains in the applied multiple queries depend on the linguistic logic of the juxtaposition of strings of the respective categorial affiliation.

The distribution of the corresponding quotas of matrices builds up a descending scale where growth of sequential dissimilarity is accompanied with numeric representation decrease. This visualises the *diachronic optimality* implication that generally there are more chronotropic pairs of strings than the non-chronotropic ones or rather that more chronotropic strings outnumber less chronotropic ones.

Each section of the scale of sequential similarity of the column string expansion to that of the row string at its n-partite division has a chance of getting filled depending on the number of the involved string constituents. At the ten-partite division, for example, this is true starting with the numeric threshold of five constituents in the matrix column string. Two-member strings reveal only extreme values of temporal similarity. The extreme values of temporal dissimilarity (all minuses in the matrix) are typically unattested in strings exceeding two members. Complete sequential dissimilarity occurs at the matrix column string of just two or, eventually three, constituents at which, respectively, the intermediate values of chronotropism are absent or scarce.

The distribution of the quotas of the similarity of the sequential expansion of the matrix column strings can be mirrored in that of the dissimilarity quotas with a slight optional variation of the actual numbers of matrices in the neighbouring areas of the n-partite division owing to understandable fluctuations in interval values (cf. the values in rows 3 and 4 on the left hand-side table on Fig. 5 with those in rows 4 and 5 on the right hand-side table). The total number of matrices in both pairs of rows from the tables is identical.

The same category reveals somewhat varied quotas of the degrees of sequential imitation in the arising strings of other categories. It seems natural that strings of varied lengths should be characterised by fluctuating quotas of chronotropism as well. If so, that quantity may be averaged for each length. It would be interesting then to see whether the distinctions of mean values for the respective cross-categorial domains as regards constituents sequential (dis)similarity continue to hold for different lengths of the row string. Conversely, possible leaps or falls in the value of chronotropism at the respective lengths may return to the overall curve's pattern at some time in history as well as at the partition and/or attestation manipulation.

A way of verifying these hypotheses was suggested by Professor Vakarchuk on the 20$^{\text{th}}$ of March 2004 in the following formalism.

Let the corpus of matrices for an arbitrary cross-categorial domain

$$\hat{A} = \begin{pmatrix} A_{11} & A_{12} & A_{13} & \dots & A_{1N} \\ A_{21} & A_{22} & A_{23} & \dots & A_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{N1} & A_{N2} & A_{N3} & \dots & A_{NN} \end{pmatrix}$$

be $A_{ij} = 0, 1$.

Then the measure of chronotropic reflection within this corpus of matrices can be obtained from the equation

$$n(N) = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij}/N(N-1) \qquad (1)$$

where i does not equal j.

The mean value of this reflection $n_k(N)$ at specific length M of the string $k = 1, 2, 3, \ldots, M$ in the matrix row can be received from the equation

$$\overline{n(N)} = \frac{1}{M} \sum_{k=1}^{M} n_{k=1}(N) \qquad (2)$$

Equation (2) stipulates the building of a curve for each cross-categorial domain of the historical thesaurus. The horizontal axis represents the lengths of the strings in the rows of respective matrices and the vertical one does the chronotropic reflection mean values valid for the strings in the columns of the same matrices.

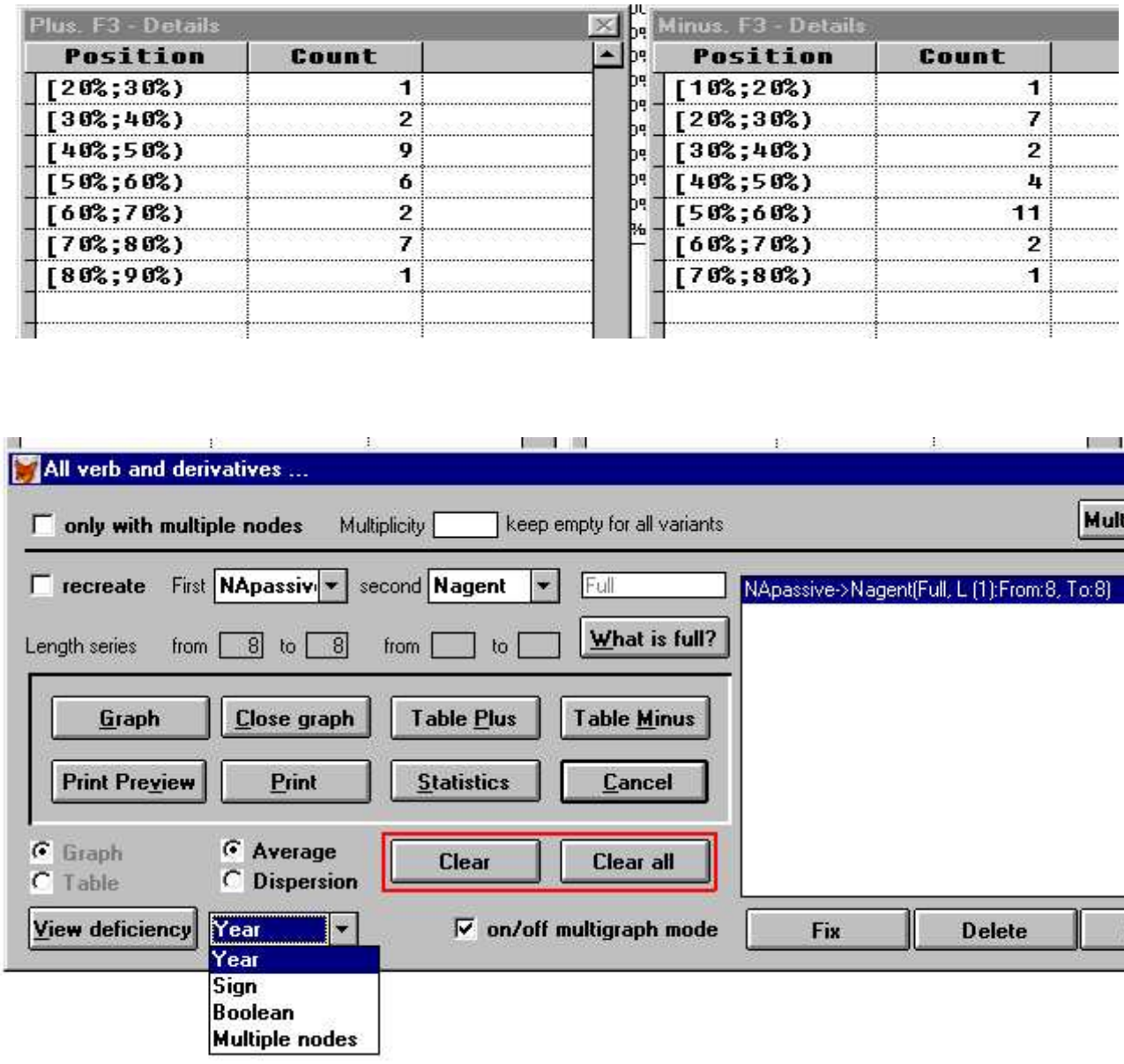


Fig. 5. Exemplification of the electronic query for the ten-partite division of quotas of temporal similarity values with the row string eight constituents matrices in pairs of secondary deadjectival (passive modal) nouns and their shared-root deverbal agent nouns.

Example ASSEMBLE

ASSEMBLE	1250	ASSEMBLER	1635
JOIN(put together)	1297	JOINER(put together)	1483
GLUE	1382	GLUER	1483
ERECT	1417	ERECTOR	1538
SOLDER	1420	SOLDERER	1530
MOULD(mix,knead,blend)	1430	MOULDER(mix,knead,blend)	1440
UNITE	1432	UNITER	1587
WELD	1599	WELDER	1828
MODEL	1604	MODELLER	1603
CONSTRUCT	1610	CONSTRUCTOR	1751

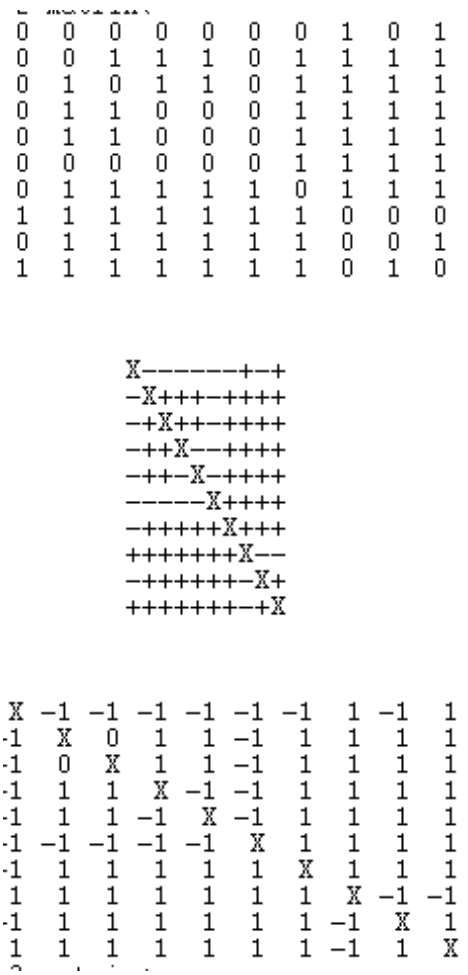


Fig. 6. Variant quantification of the temporal similarity of the expansion of two strings in diachrony: upper matrix — Boolean representation; lower matrix — precedence/follow-up balanced representation with the nullifying of the comparisons involving matrix column string's constituents with same year attestations.

The mean value of the sequential similarity of strings at specific lengths required a quantification of the imitative logic reflected in the construed matrices.

Four versions of such a quantification are plausible within the above-suggested formalism (cf. the lowest subwindow for electronic queries on Fig. 5):

(1) The plusses in the squares of chronotropism matrices can be ascribed the numeric value of plus one and the minuses that of zero (see the upper matrix on Fig. 6);

(2) The positive and negative outcomes of the multiple comparisons of the matrix column string constituents with the constituents of the string in the matrix row can be imputed the numeric values of plus one and minus one, respectively. At this kind of quantification of matrices the construed curve is lower in comparison with that obtained when applying the method given under 1 by the ratio of the doubled quantity of the negative outcomes of sequential comparisons;

(3) The kind of quantification described under 2 can be made more subtle when we take into account same year attestations of constituents in the column string. Then their comparisons to the respective common-root counterpart across the domain are to be nullified (see the lower matrix on Fig. 6). The matrix of the procedure described under 2 has no place for the zero value ascribed to its squares as there are no same year attestations among the column string constituents (Fig. 7). However, when instead of assessment procedure 3 we apply assessment procedure 2 for the matrix row string that contains same year attestations we unduly ascribe diachronic relevance to the string's constituents sequential placement that is basically non-historical (cf. Fig. 6 where the values of zero in the lower matrix are ascribed the values of pluses in the respective squares of the medium matrix). It would be curious to see to what extent this distortion is observable in the various cross-categorial domains although the intuition is that the induced curve correction is negligible.

(4) The signs "plus" and "minus" from the squares of matrices can be converted into the negative and positive (note the reverse order) values of the age differential of the respective *OED* textual prototypes of the compared lexemes. This measure is to be expressed in years. Then the mean value of sequential similarity can be expressed as that of the average age differential of the earliest quotations.

When the mean value of chronotropism for a specific length of the strings in the matrices rows approaches zero it means that the cases of the strings constituents sequential imitation (the negative values of age differential) roughly equal those of sequential dissimilarity (the positive values of age differential) revealing basically convergent total values of the opposite age differentials. The nature of the obtained zero mean values would be in principle indentical under procedure 2 but impossible under procedure 1.

Owing to this peculiarity of the obtained values of the age differential the construed curves would typically appear in the zone of negative values on the vertical axis and run below the horizontal axis. This prediction complies with the *diachronic optimality* prediction expressed in the third paragraph of the current section of the paper. In this version of quantification same year attestations of the constituents in the matrix column string are being taken into account 'naturally' as part of the calculus.

Example VANISH

DEPARTINGLY*	1388	DEPARTABLE*	1377
DYINGLY(cease to live)	1435		
DISPERSINGLY	1594	DISPERSABLE	1827
ESCAPINGLY	1631	ESCAPABLE	1864
GOINGLY*	1651	GOABLE	1573
PERISHINGLY	1698	PERISHABLE	1611
DISSOLVINGLY	1822	DISSOLVABLE	1541
RETIRINGLY	1861		
VANISHINGLY	1870		

X	1	1	1	1	1
1	X	1	−1	−1	−1
1	1	X	−1	−1	−1
1	−1	−1	X	1	−1
1	−1	−1	1	X	−1
1	−1	−1	−1	−1	X

Fig. 7. Temporal similarity of the expansion of two strings in diachrony in precedence/follow-up balanced representation of comparisons in the matrix.

V. DISCUSSION

The construed curves help us to better understand the dynamics of constituents expansion in different sections of the lexicon over time.

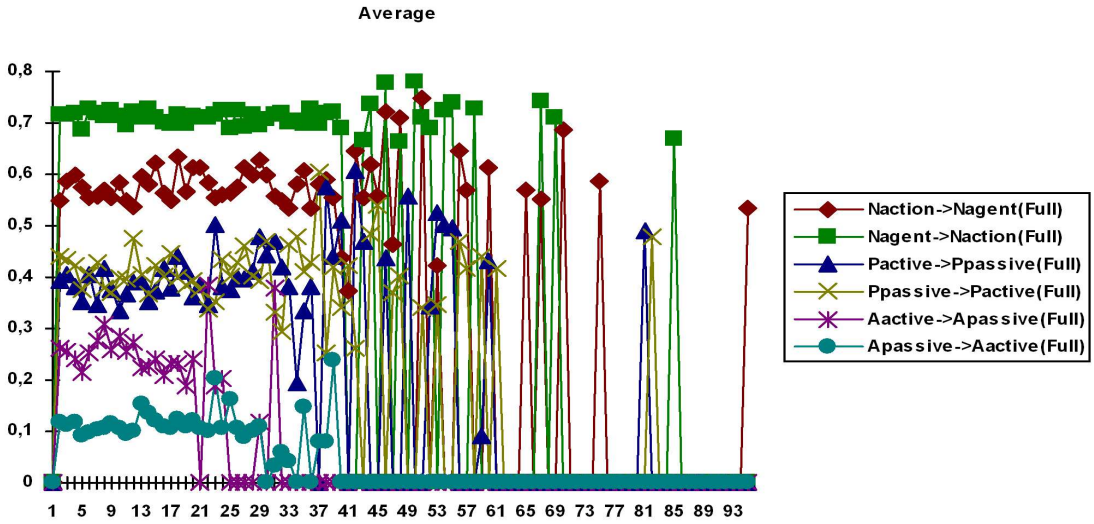
Behind each of the curves there is statistics serving as a foundation for the mean values of expansion similarity of the column string constituents with that in the row string on condition of at least two attested manifestations of a specific length. The quantification procedure becomes uninformative at a single manifestation of the matrix.

A drop of the curve to the zero value on axis y for a specific string length on axis x may though testify to the calculus outcome equalling zero. Then the cases of sequential similarity and dissimilarity fully coincide. This is easier in the precedence/follow-up representation than in the one basing on the age differentials although in

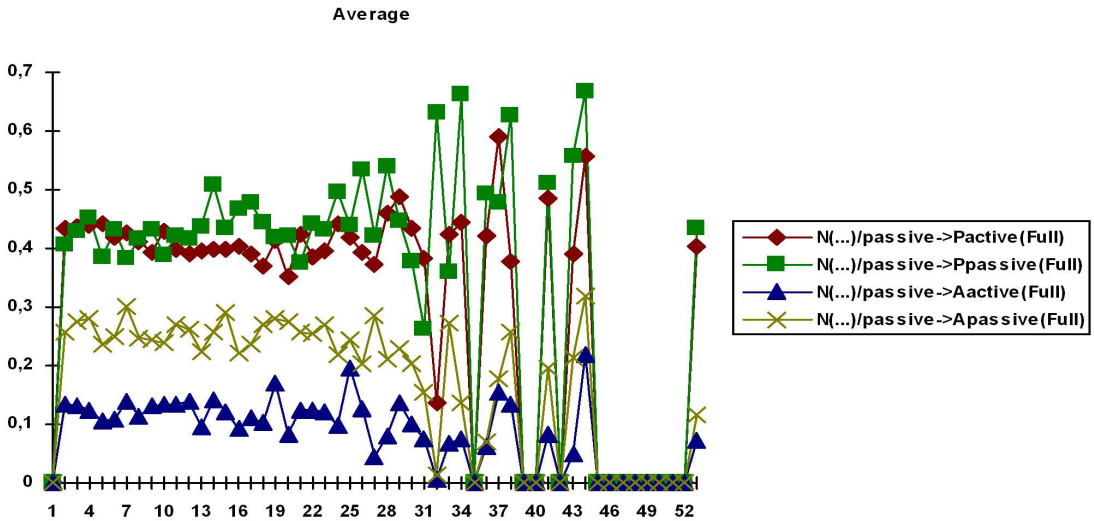
either of them the effect can become observable.

The zero values on axis x may have a purely technical though historically relevant reason. They are present when the stage of lexicon evolution as yet cannot yield a specific length to the strings of the given categorial affiliation, a process the respective query is capable of visualizing fully.

Pairs of sequential similarity curves for the respective cross-categorial domains within deverbal word-forming families are variedly convergent/divergent. Various deverbal categories optionally fall different by way of reflecting the expansion of the arbitrarily chosen reference category. For instance, the mutual imitation of the constituents' placement in the expansion of strings of participles (curves 3 and 4 on Fig. 8,a) are considerably more convergent than that in strings of adjectives or nouns. Likewise, participles resemble the expansion of factitive nouns rather uniformly whereas adjectives do so pretty variedly (Fig. 8,b). The mean values of chronotropism for different lengths in some cross-categorial domains are more scattered than in others (cf. the strings within pairs of curves on Fig. 8) opening up a prospect of curvature analysis of a multitude of such distributions in the future.



a) exemplification of varied part-of-speech affiliation of the compared strings.



b) exemplification of the same part-of-speech affiliation of the compared strings.

Fig. 8. Imitation of constituents expansion in classes of deverbatives in Boolean representation: axis y — mean values of similarity; axis x — lengths of the row strings as indicated in the left hand-side category in the table lines (to be repeated in subsequent charts. In curves 1 and 2 on Fig. b the notation N_{action} stands for an aggregate count of action and factitive nouns).

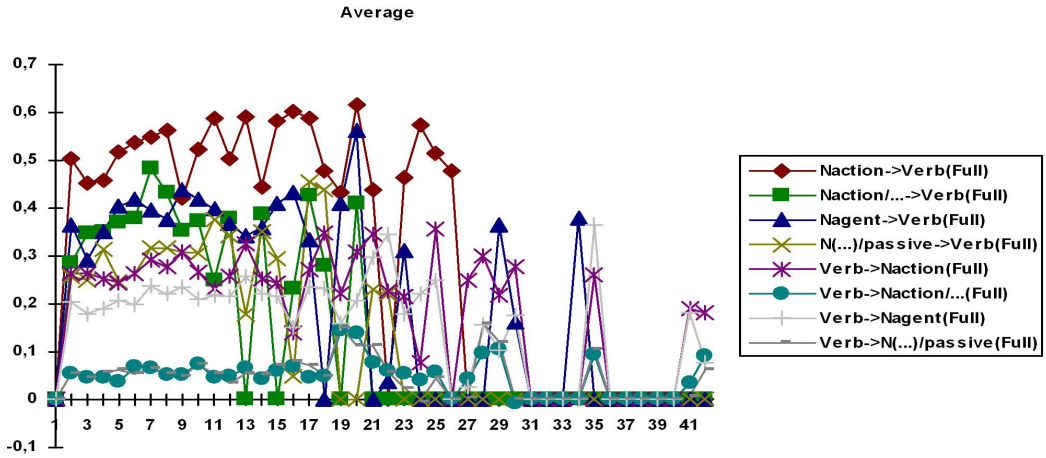


Fig. 9. Mutual imitation of constituents expansion in verbal-substantival/substantival-verbal cross-categorial domains for Old English and Middle English (888–1500) in Boolean representation.

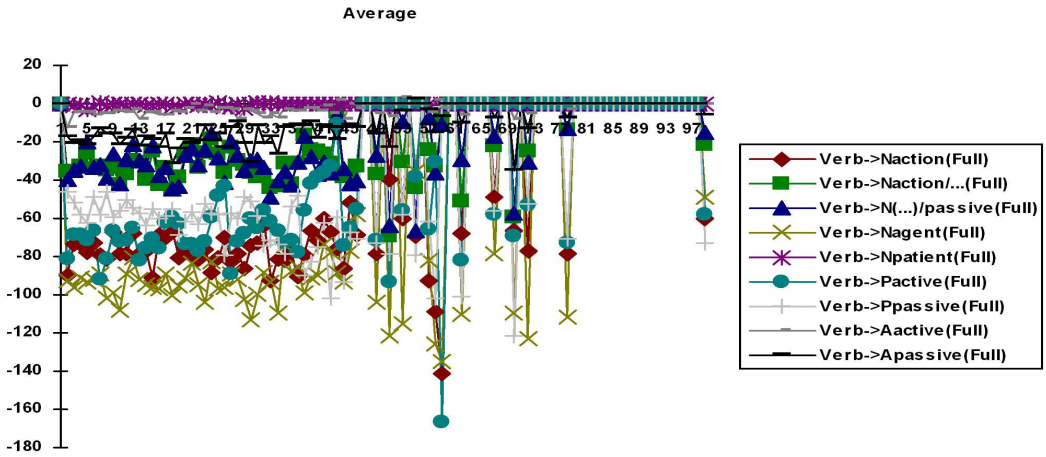


Fig. 10. Imitation of the sequential logic of strings of deverbatives in the expansion of strings of verbs in the age differential representation.

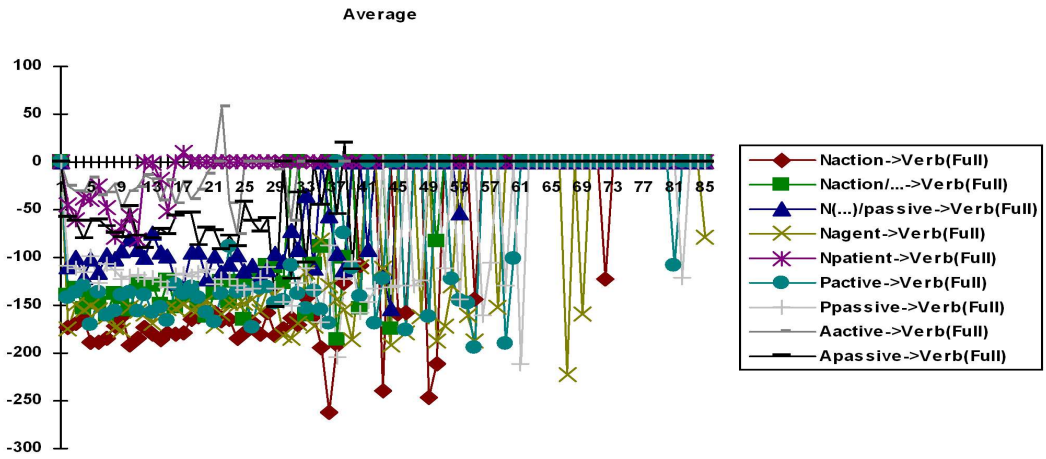
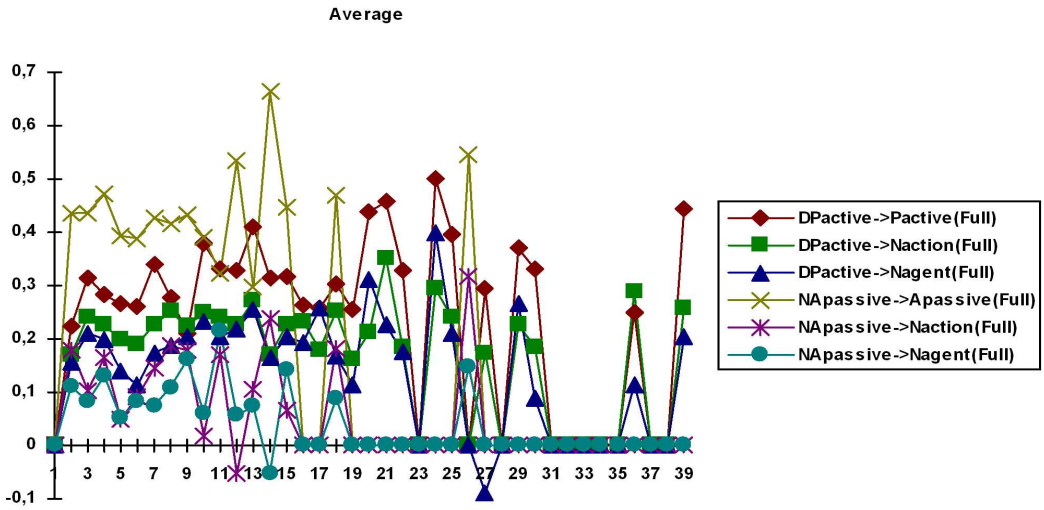


Fig. 11. Imitation of the sequential logic of strings of parent verbs in the expansion of strings of deverbatives in the age differential representation.

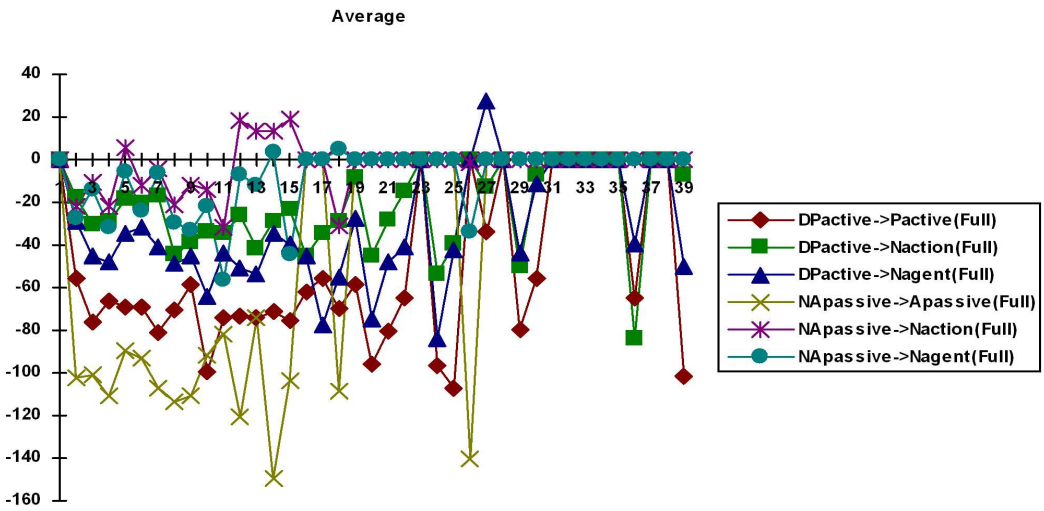
The mean values of chronotropism of the compared string to the reference point string placed in the matrix row takes into account the number of the constituents in the latter (the value of the parameter M from (1.2)). Some of these do not produce common-root constituents to fill in the positions in the column string. That is why although the matrices can be inverted the strings placement inside them is to a point expediency determined.

At the initial stages of lexicon evolution, for instance up till the end of Middle English, when the stringing of verbs largely precedes that of deverbatives there are "too many" verbs with an empty correlative deverbal position of the given categorial affiliation. Hence it is expedient to place them in the position of the matrix column string (cf. the first four and the last four curves on fig. 9). Similar logic holds for the immediate (adjectival/participial) and ultimate (verbal) bases of secondary deverbatives.

Verbal-deverbal matrix inversion can otherwise be quite informative about the extent of positive/negative comparison outcomes for specific lengths of the matrix row strings (cf. the mean values on axis y on Fig. 10 and 11 and also the mutual placement of curves 1 and 4 as well as 2 and 3 on the said figures). The same effect of a change in the mutual placement of the curves brought about by the distribution of matrix row strings' lengths may be hidden in the very quantification (cf. curves 2 and 3 on Fig. 12, a and b). Likewise, for earlier stages of the lexicon expansion same time attestations of matrix row constituents tend to modify the mean similarity values distribution at specific lengths (cf. Figs. 13 a and b).



a) precedence/follow-up representation

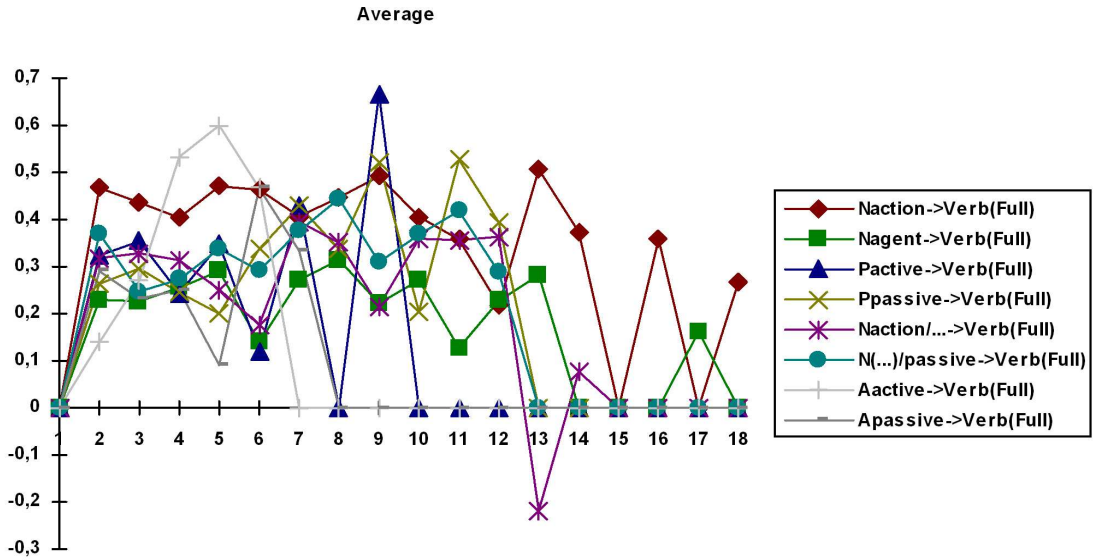


b) age differentials representation

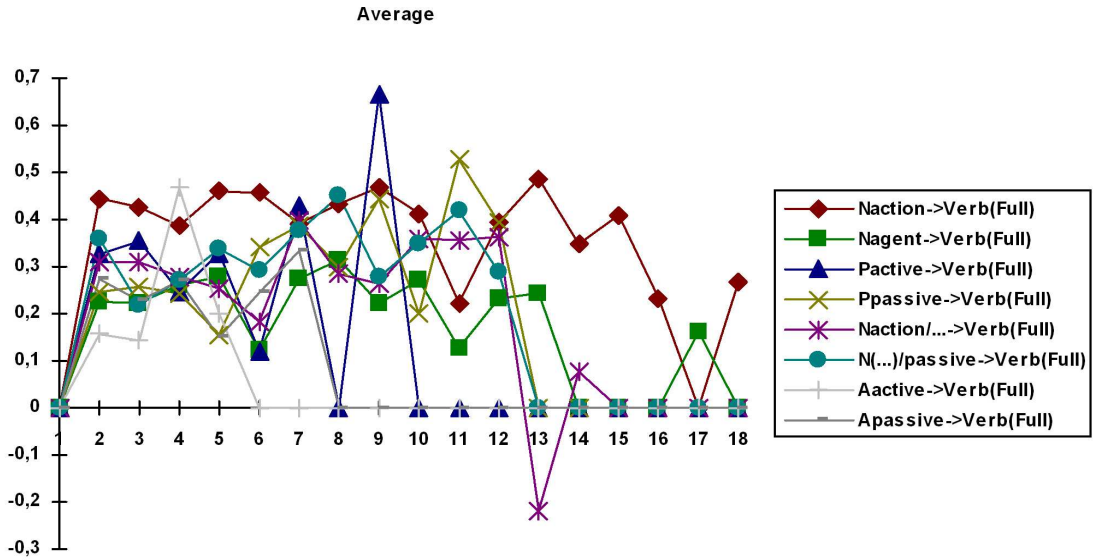
Fig. 12. Exemplification of mode of representation relevance in the cross-tier categorial domains.

The extent of sequential resemblance of secondary deverbatives to that of their motivating bases should not necessarily be identical. Hence there is larger affinity between strings of secondary deverbal nouns and their motivating passive modal adjectives than between strings of adverbs and those of their motivating bases. The inter-tier cross-categorial domains of the two types of secondary deverbatives mentioned and strings of action and agent nouns reveal an opposite tendency. At most lengths of the matrix row string the expansion of strings of adverbs coined from present participles is more imitative of the expansion of either category of the nounal strings (Fig 12). These disproportions will disappear if the secondary deverbal strings mentioned are placed in the columns of the respective matrices of chronotropism.

A heuristic extension of the developed framework is feasible upon streaming the derivational bases into respective sub-corpora on etymological and/or thematic grounds (cf. Fig. 14 a with Fig. 11 and Fig 14 b with Fig. 10).

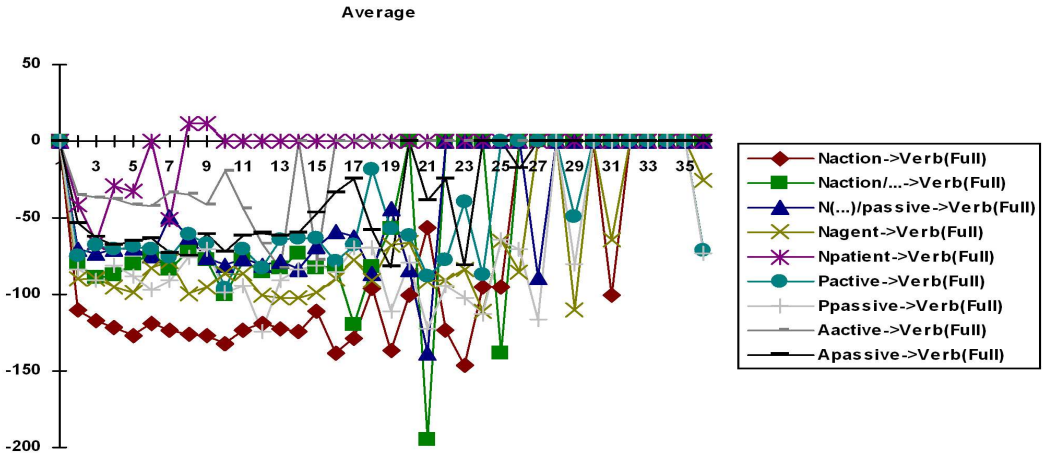


a) same year attestations taken as sequential

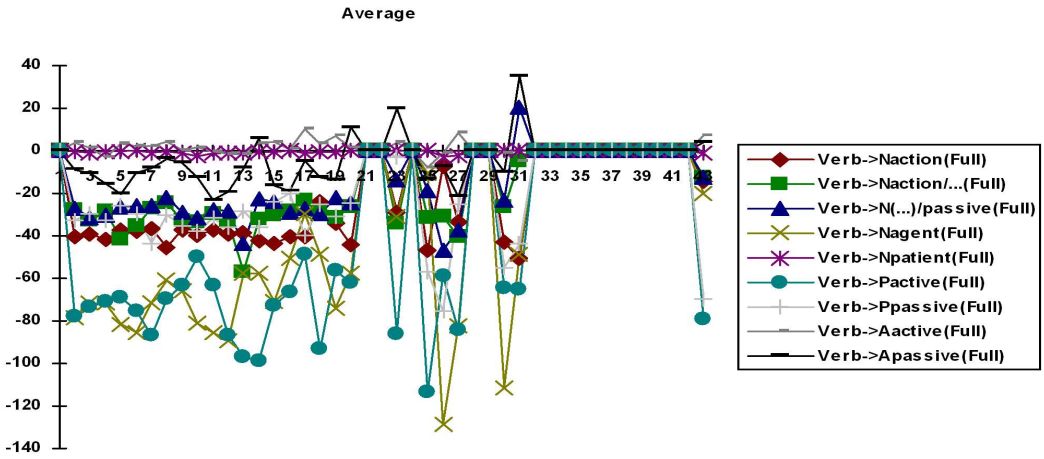


b) same year attestations nullified

Fig. 13. Exemplification of mode of representation relevance in comparing the ordinal placement of verbs with the positions of their common-root deverbatives in Middle Englishin.



a) the case of verbs of French lineage as strings' constituents in the position of matrix rows



b) the case of verbs denoting abstract relations (by [1]) as strings' constituents in the position of matrix columns

Fig. 14. Exemplification of distributional fluctuations in samples from the universal corpus set.

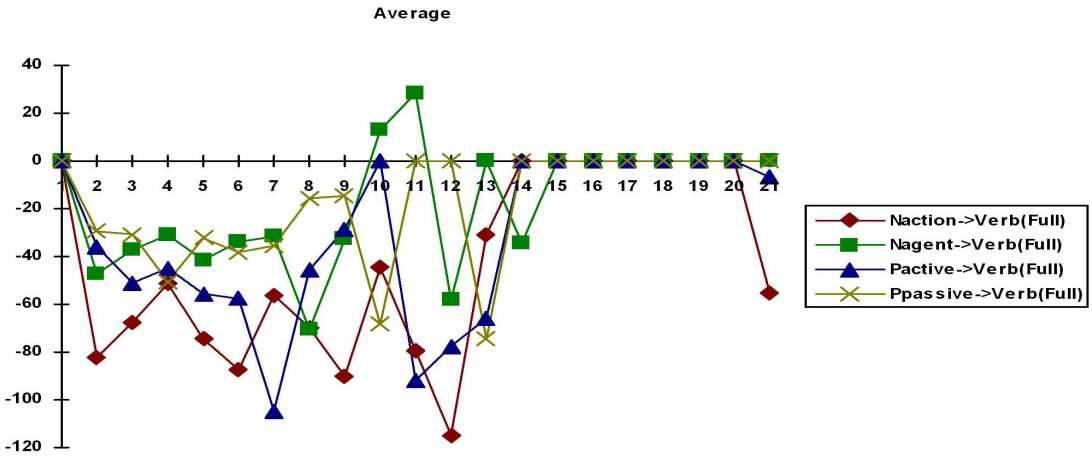
We conclude by illustrating that the suggested approach is capable of carrying out queries at the level of detail that might seem like a barely accomplishable whim to a historical lexicologist working in the pre-electronic age (Fig. 15–16). Note that the mean values above axis x are shown to occur at some lengths on Fig. 15,a but never on Fig 15,b proceeding from the natural morphology of base-suffix etymology-wise ordering of the underlying queries.

CONCLUDING REMARKS

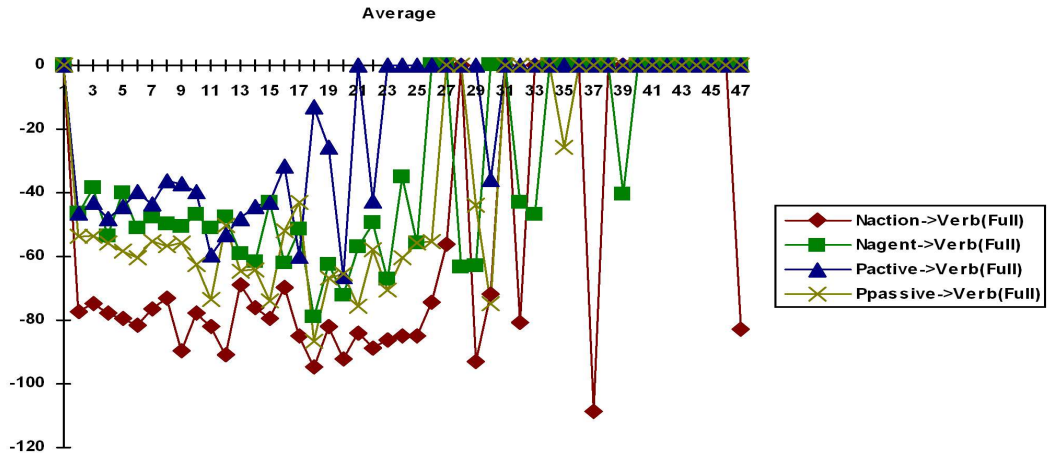
The developed framework surely provides us with more information than we can safely digest for the present. Yet hopefully it conforms to two fundamental methodological principles of linguistics as a science by offering a better and better description of data and pinpointing the things that fit in its chaos.

The epistemological help in putting these principles into practice may come from outside linguistics. I am profoundly grateful to Professor Vakarchuk for his response as a physicist to a linguistic problem and wish him long and prolific life of a scholar who is not afraid of the deep water of intercultural exchange between natural sciences and humanities capable of enriching them both.

The age of applying "the flexible use of dictionaries, such as the *OED* or the *MED*, in addition to, and complementation of, special corpora, whether these are *self-compiled* or not" [7] is capable of bringing to light new tasks that might look like 'exotic' problems of statistical physics. The same holds true for the impetus at a more in-depth penetration into traditional lexicological concerns like sequences length and their constituents' succession that we have addressed here proceeding from the grounds which have not at all been common in historical lexicology and which we can only owe to statistical physics.



a) the case of native stems as strings' constituents in the position of matrix columns



b) the case of borrowed (Latin and/or French) stems as strings' constituents in the position of matrix columns

Fig. 15. Mean values of sequential similarity in the cross-categorial domains of the late Middle English and early New English (1300–1700) chronological layers with aggregate action nouns (factitive couterparts included) and agent nouns coined with the help of the *-ing/-er* native suffixes.

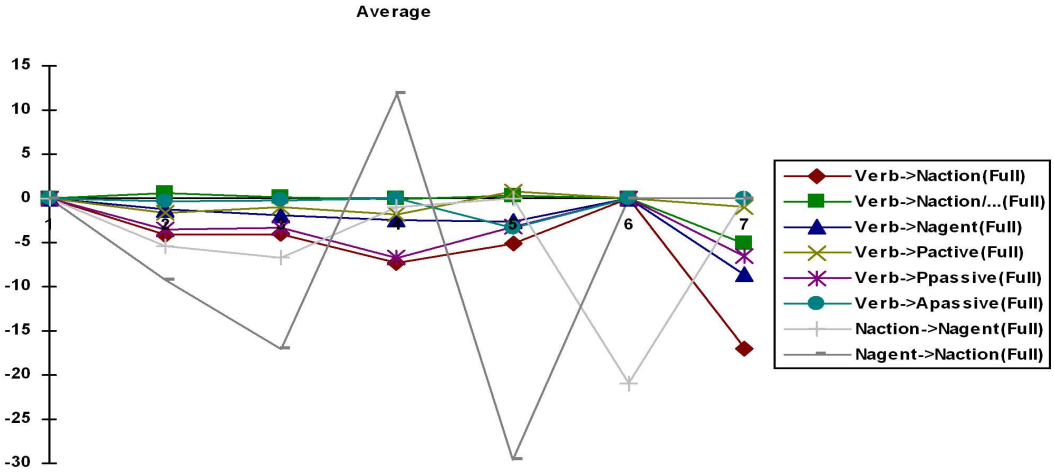


Fig. 16. Similarity of synonymic expansion in the 19th c. *OED* registered neologisms in the age differential representation.

The construed *Historical Thesaurus of Deverbal Families in English* as a new lexicographic object is a reconstruction feasible on condition of appropriate epistemic procedures and simulative tools of an electronic corpus-based framework of relevant queries.

The distributions of values of temporal similarity expansion forming respective curves in their visualization constitute a new empirical reality in historical lexicology. This reality is quite ample as the chronotropic reflection of 306 cross-categorial domains representing a succession of mean values in four representational modifications mentioned yield over a thousand curves. The quantity of these curves is multipliable by the number of partitions separated on the grounds of agreed relevance (randomly exemplified above) as well as by the number of the respective thesauri from the dictionaries of synonymous strings taken for the database of such an experiment. Hopefully, they present a prolific ground for speculation still awaiting for historical lexicologists of English (and possibly some other languages with compatible empirical base in diachronic lexicography and/or textual corpora) in the future.

Acknowledgements

I am very grateful to Professor Volodymyr Tkachuk and Associate Professor Andrij Rovenchak from the Department of Theoretical Physics of the Ivan Franko National University in Lviv for discussing the statistical framework suggested by Professor Ivan Vakarchuk and also to Associate Professor Andriy Pereymybida from the Faculty of Computer Science of the Ivan Franko National University in Lviv for writing appropriate software.

[1] *Roget's Thesaurus of English Words and Phrases*, new edition prepared by S. M. Lloyd (Longman, 1982).
[2] Ch. J. Kay, A. W. I. Wotherspoon, in *A Changing World of Words. Studies in English Historical Lexicography, Lexicvology and Semantics*, (Rodopi, Amsterdam, 2002), p. 109.
[3] M. Bilynsky, in *Medieval English ans its Heritage. Strucvture, Meaning and Mechanism of Change* (Peter Lang, Frankfurt am Main, 2006), p. 77.
[4] *Webster's New World Thesaurus*, prepared by Ch. Laird, (Prentice Hall Press, New York, 1986).
[5] *OED Oxford English Dictionary. 2nd Ed. on CD-ROM Version 3.0* (Oxford, Oxford University Press, 2002).
[6] В. Г. Вилюман, *Английская синонимика* (Высшая школа, Москва, 1980).
[7] M. Markus, http://drops.dagstuhl.de/opus/volltexte/2007/1052/

ДОВЖИНА СИНОНІМНИХ РЯДІВ ТА РОЗТАШУВАННЯ ЇХНІХ КОНСТИТУЕНТІВ У ВІДДІЄСЛІВНОМУ СЛОВОТВОРЕННІ АНГЛІЙСЬКОЇ МОВИ: ВНЕСОК СТАТИСТИЧНОЇ ФІЗИКИ ДО ІСТОРИЧНОЇ ЛЕКСИКОЛОГІЇ

М. Білинський
Львівський національний університет імені Івана Франка, кафедра англійської філології
вул. Університетська, 1, Львів, 79000, Україна

Розглянуто можливості вивчення проблеми конституентного складу історичного тезауруса з позицій статистичної фізики. Основну увагу звернено на перспективу застосування до задач історичної лексикології способів оцінювання міри послідовної подібности при розгортанні деривaційно співвіднесених синонімних рядів.