Research Article

*Interdisciplinary British and American Studies Journal*
December 2024; Volume 1, Issue 1

https://doi.org/10.5281/zenodo.14602681

# Remodelling Verbal Synonymy for the History of English: Sources and Methods*

https://orcid.org/0009-0005-1736-0904 Michael Bilynsky[1]

[1] Professor, Ivan Franko Lviv National University, Ukraine

## *abstract*

A synonymous placement of words in present-day thesauri can be reconstructed to become a relevant object of diachronic onomasiology. In a historical sense present-day synonyms were attested at specified moments in the string's diachrony and at a given succession over time on the imposed condition of inventory sameness. A possible method of studying this problem may lie in the sum total of deductions between the ordinal positions of elements in the contemporary string and its historical counterpart. As English provides a dated diachronic textual prototype for any lexical item in its first attested, as well as other, meaning(s), owing to the textual corpus of the OED, a better fitting model for this task is a vector space model. It distinguishes the *relative* and *absolute* chronologies of diachronically reshuffled present-day strings. The similarity measure between the present-day string and its diachronic permutation then consists in the difference between the lengths of the obtained vectors. These reflect contemporary (fragmentary, or, eventually, all) and the corresponding period, if set so, weight placements of the given constituents. The developed framework allows for data-driven exemplification as well as distributional visualization. The methodology is deemed to be applicable to multiple lexicographic sources of verbal synonymy. The data is combined with the OED-obtained composition of de-verbal families.

***Keywords***: verbs, de-verbal coinages, synonymous strings, diachronic permutation, similarity

Correspondence: Michael Bilynsky, e-mail: icome7.lviv@gmail.com

Received: 09.09.2024 | Accepted: 15.11.2024

**This article builds upon and expands upon the findings presented in Bilynskyi, M. (2009). Synonymous Stringing as a Diachronic Reconstruction: The OED Middle English Evidence for Verbs and Deverbal Coinages. ІНОЗЕМНА ФІЛОЛОГІЯ (INOZEMNA PHILOLOGIA), 121, 20-42.*

## 1. INTRODUCTORY REMARKS

The developed methodology of assessing sequential similarity between present-day strings of synonyms and their historical reconstructions was presented for the first time in my talk given at the Sixth International Conference on Middle English held in Cambridge (Bilynsky 2008). The mathematical side of this methodology, which, we hope is usable for diachronic lexicology, is our development of a simple idea. It was suggested by the renowned quantitative linguist Professor Gabriel Altmann in our personal correspondence and then extended into a more subtle vector space solution described in detail in our co-authored paper (Bilynsky, Pereymybida and Altman 2009).

In the present work, apart from the application issues, I juxtapose the ME and the post-ME sample reconstructions on select variables. A ME reconstruction was presented in an extended version of my post-ICOME 8 paper in 2009. The post-ME evidence analysed along the suggested methodological principles is presented for the first time, as well as some lines of comparison between the set periods and inside select variables modelling.

Multiple thesauri of verbs, which for present-day English are over twenty in the book or online formats (and sometimes in both), have been processed. The present paper, however, focuses more on the methodology of carrying out a particular task than on the epistemology and heuristics of databases which, when combined with the OED earliest quotations, turn into objects of historical lexicology.

Verbs initiate single or multiple synonymous strings depending on their polysemy. For verbs that have more than one meaning the dominant is accompanied with the respective meaning specification. If no such formula is provided by the thesaurus, we can distinguish the strings by their compositions, which typically differ already at the first synonym after the dominant.

In the present framework, the object of our study is a documented string of synonyms. The extents of similarity that it reveals when reshaped into a historical sequence in the reconstruction rely on the recreation of the appearance of its constituents over time.

More complex strings provide a larger quantity of synonyms and less complex ones adduce smaller numbers of synonyms. This property of a string is also traceable diachronically on the premise that its constituents are characterized by a given age. They were molding the string ‘in turn’, and by the produced cumulative effects they were building the string gradually until, after a few intermediate stages, it became what it is like today in the inventory sense.

The string is diachronically representable by its earliest and latest constituents, given that they do not share this status with any other lexeme(s) of identical dating. This is too harsh a condition as same-year attestations of two or even more constituents within a string occasionally occur. Also, present-day strings sometimes reveal coincident historical sequencing.

The suggested modeling takes place as an interaction and mutual enrichment of synonymy dictionaries and historical texts illustrating the respective lemmas in the OED. The dated extracts from the OED are according to Antoinette Renouf “chronologically held texts” (2007, 37). The bank of these earliest quotations (precedent usage texts) constitutes a diachronic textual corpus (Hoffmann 2004; Alexander and Dallachy 2020, 182). We have used the 3rd version of the 2nd

The architecture of our queries allows for the updating of data in line with the ongoing OED3 editing. The lexemes from all the queries are provided with the date labels that stand for the corresponding OED earliest quotations.

The union of the OED evidence of diachronic textual prototypes of words in their meanings, even senses, and a fabric of synonymy according to *Roget's Thesaurus* was achieved in the *Historical Thesaurus of English*, 2010 (cf. Kay and Wotherspoon 2002) that is now being integrated into OED3.

Our inventory framework also entails strings data contained in *Roget's Thesaurus*. It includes a number of other dictionaries of synonyms as well. Altogether we can speak of an extended (but, also, partitioned) base of approximately 100,000 strings of synonymous verbs and over half a million of strings of their shared-root coinages. For present-day criteria of representativeness and validation in lexicography and corpus linguistics cf., for instance, de Schryver (2003) and de Schryver et al. (2023).

The diachronically reshuffled string starts the first lexeme and continues (sometimes instantaneously) into a rise of a precedent diachronic pair that is taken for the prototype of the respective string. Then such a pair was complemented by at least one, or possibly more words until the entire composition of the string was attested historically.

The exemplification of the queries and distribution visualizations are data-driven from the framework.

The framework can be used for studying the evolution of present-day synonymous strings over time. It can be also split for more specific tasks of select variables reconstruction.

## 2. RECONSTRUCTING PERIOD THESAURI OF VERBAL SYNONYMS

The formation of the string is traceable by the first and the last elements of its composition. The string is understood as at least one additional element added to a pair of synonyms. The attestation of the initial and concluding lags of the string relevant for a specific period is determined by the respective chronological boundaries.

Alongside of the notion of a synonymous string for period reconstruction we may also need the notion of a synonymous sub-string. Sometimes strings figure alongside of sub-strings. Hence a handy notation appears as a (sub-)string. The situation resembles that with synonyms, near-synonyms and (near-)synonymy. The first element of the string or its prototypical pair in the ME thesaurus could be dated with the start of the ME period. Significantly, a lot of stringing of elements revealed on the basis of the OED earliest attestations during the ME period has the first element of the string, its prototypical pair or even a larger portion of the composition inside the OE records attested before the $12^{th}$ c.

Thus, in some cases the ME thesaurus contained only the elements dated in ME, but in others they were dated in both ME and OE, and in some cases the OE elements had no ME extensions, but were there in the ME texts as could be seen from the OED attestations after their diachronic textual prototypes.

Hence the respective ME chronological layer in the historical thesaurus refers to ME stringing alone, ME stringing and an OE element, a ME element after OE stringing, a ME element after a single OE element, or even OE stringing that was not continued in ME, but could, although not necessarily, be continued in post-ME.

The chronological compositions of synonyms relevant for the ME cross-section of the thesaurus were the very strings from the entire thesaurus or segments of these strings, i.e., sub-strings, with a possible, pre-ME recovery or/and post-ME complementation.

The post-ME cross-section of the thesaurus contains the elements registered after 1500. Understandably, some of them were the strings registered for the whole thesaurus while most were its (sub-)strings initiated during the (OE+)ME cross-section.

Following the above assumptions a thesaurus is split into the age wise extreme values elements corresponding to the chronological confines of the period of language history and the element(s) inside these confines. At the same time a period string may be a real string from the thesaurus. Conversely, the contents of a real string from the thesaurus may belong to more than one chronological period.

The dimensional organization of strings exhibits adaptability to both exact measurements and range parameters regarding the quantity of constituents. The temporal boundaries of each string are established through OED's chronological documentation, identifying its earliest and most recent constituent(s). Within the sophisticated electronic interface, temporal parameters can be configured independently, either as specific chronological points or extended periods within the textual sequence.

Navigation through historical strings within the compiled corpus is maintained via contemporary reference points. The system framework generates comprehensive age-set classifications and string listings, with variable length specifications, complementing the completely reorganized historical thesaurus.

Analysis of contemporary strings within the ME thesaurus reveals two distinct patterns. The first scenario presents no constituents of subsequent temporal classification within the present-day string, resulting in identical string compositions across both the comprehensive historical thesaurus and its ME component. The alternative scenario involves certain constituent(s) in the string from the broader collection dating beyond 1500, while maintaining a minimum of two constituents documented prior to 1500, forming the ME *sub-string*.

The sign '⊂' functions as a demarcation symbol, distinguishing the string's dominant from its other compositional elements: e.g. ***revere*** *'⊂'. venerate, regard, respect*. When there is a floating dominant in the historical reconstruction this sign will appear inside the string.

The general set encompasses 1,198 strings that align with the ME reconstruction. Notably, 422 strings, comprising approximately one-third of the total, maintain identical lexemes in both the present-day dominant position and the string's earliest constituent. To denote this characteristic, we employ specific notational conventions: an asterisk (*) indicating historical reconstruction and brackets ([]) signifying conditional present-day symbols * ⊂ [⊂], both appearing after the string's dominant. When the dominant remains consistent across both versions of the string, it is classified as historically *intact.*

The strings exhibit two primary classifications: those with historically *intact* dominants and those with *floating* dominants. Their constituents demonstrate correlation patterns with specific historical periods in the OED textual prototypes. The distribution is as follows:

• ME correlation: 163 strings with *intact* dominants and 308 with *floating* dominants, as exemplified in (1) and (2)
• Exclusive OE correlation: A smaller subset of 15 strings with *intact* dominants and 17 with *floating* dominants, as shown in (3) and (4)
• Combined OE and ME correlation: 229 strings with *intact* dominants and 466 with *floating* dominants, as illustrated in (5) and (6)

> (1) ME strings of verbs containing ME textual prototypes with the diachronically intact dominant, e.g. ***conjure*** *(1290) * ⊂ [⊂] [To appeal to] entreat (1340), adjure (1382), implore (1500)*
> (2) ME strings of verbs containing ME textual prototypes with the diachronically floating dominant, e.g. ***amend*** *(1220) * ⊂ redress (1325), reform (1340), rectify [⊂] (1400)*
> (3) ME strings of verbs containing OE textual prototypes with the diachronically intact dominant, e.g. ***spare*** *(825) *⊂ [⊂] forbear (888), forgive (900)*
> (4) ME strings of verbs containing OE textual prototypes with the diachronically floating dominant, e.g. ***arise*** *(825) * ⊂ stir (888), awake [⊂] (1000)*
> (5) ME strings of verbs containing OE and ME textual prototypes with the diachronically intact dominant, e.g. ***sail*** *(893) *⊂ [⊂] [To fly] glide (1000), float (1100), soar (1374), skim (1420)*
> (6) ME strings of verbs containing OE and ME textual prototypes with the diachronically floating dominant, e.g. ***arise*** *(825) *⊂ uprise (1300), mount [⊂] (1362), ascend (1382)* (Bilynsky 2009).

A substantial part of ME verbal synonymy contains sub-strings from the general set series, the latter entailing one or more constituents dated after the expiry of the ME period (post-1500 earliest quotations). The selection of almost 5,000 strings drawn here from *Webster's New World Thesaurus* (the synonymy from other thesauri can be subjected to the same treatment) split virtually to equal measure (2,339 and 2,495 ME (sub-)strings initiated, respectively, by OE-ME and ME elements).

The same headword in a contemporary string and in its historical reconstruction solidifies the perception of a lexical group whose present-day and historical ontologies may differ only as regards the constituents placements 'held together' by the headword. This holds true for a mere one fifth of the set groups, respectively, around 500 examples started by the OE (e.g. under 7 and 7.1) or a ME (e.g. under 8) headword. In the remainder of the strings the reshuffling affected the present-day dominant.

The quotas with the diachronically intact headword rest on barely 20 per cent of the selection in (7), (7.1) and (8). Also cf. fig, 2 and 3 in the next section.

Change of the headword during a historical reshuffling of the string elements is more common for ME sub-strings (with chronologically extending elements outside the period), as in e.g. (9/9.1) and (10), rather than for ME strings (where all elements stay within the time confines of ME). Likewise, the sub-sets with the constituents with the OE earliest quotations as well as both ME and post-ME elements, as in e.g. (7.1) and (9.1), were more common than those with just post-ME complementation and no additions during the ME period, as in e.g. (7) and (9):

(7) ME sub-strings of verbs containing OE (and eventually post-ME) textual prototypes with the diachronically intact dominant, e.g. ***grow*** *(725) *⊂ [⊂] [To begin] arise (825), start (1000),* originate (1653)

(7.1) ME sub-strings of verbs containing OE, ME (and eventually post-ME) textual prototypes with the diachronically intact dominant, e.g. ***bow*** *(893) *⊂ [⊂] [To submit] bend (1000), surrender (1466), capitulate (1580), acquiesce (1620)*

(8) ME sub-strings of verbs containing ME (and eventually post-ME) textual prototypes with the diachronically intact dominant, e.g. ***approach*** *(1305) *⊂ [⊂] [To approach personally] propose (1340), address (1374), corner (1387), request (1533), accost (1578), button-hole (1828)*

(9) ME sub-strings of verbs containing OE (and eventually post-ME) textual prototypes with the diachronically floating dominant, e.g. ***wet*** *(825) *⊂ overflow [⊂] (893), water (897), inundate (1623)*

(9.1) ME sub-strings of verbs containing OE, ME (and eventually post-ME) textual prototypes with the diachronically floating dominant, e.g. ***crave*** *(1000) *⊂ want [⊂] (1200), covet (1225), require (1375), aspire (1460), fancy (1545)*

(10) ME sub-strings of verbs containing ME (and eventually post-ME) textual prototypes with the diachronically floating dominant, e.g. ***glare*** *(1250) *⊂ pout (1325), scowl (1340), frown [⊂] (1386), gloom (1399), glower (1500), lower (1606), grimace (1762), sulk (1781)*

Post-ME complementation of a ME sub-string fell on varied lengths: e.g. ***swing*** *(725) *⊂ [⊂] wield (825), wave (1000), whirl (1290), flourish (1300), brandish (1325), twirl (1598)* vs. ***cackle*** *(1225) *⊂ [⊂] cluck (1481), giggle (1509), gabble (1577), chuckle (1598), quack (1617), titter (1619), snicker (1694), snigger (1706)* (Bilynsky 2009).

The inventory of the present-day synonymous strings typically contains elements of older (before 1500) vocabulary. Curiously, only about two hundred strings in Webster's New World Dictionary, which is quite an insignificant number, consist of lexemes with the earliest attestations after the end of the ME period (11). A much higher number of post-ME sub-strings (almost nine hundred) are anchored by one element attested before 1500 (12).

(11) Post-ME strings containing only after-1500 diachronic textual prototypes, e.g. ***annihilate*** *⊂ demolish, exterminate, obliterate;* ***picket*** *⊂ [To strike] blockade, boycott;* ***lampoon*** *⊂ satirize, caricature, parody.*

(12) Post-ME sub-strings anchored in ME by one pre-1500 constituent, e.g. ***stylize*** *⊂ conventionalize, formalize, accord (1123);* ***dishearten*** *⊂ dampen, dismay (1297);* ***bustle*** *(1362) ⊂ hasten, hustle;* ***embitter*** *⊂ acidulate, sour (1340);* ***quibble*** *⊂ dodge, avoid (1300);* ***sidle*** *⊂ veer, tilt (1399);* ***hypnotize*** *⊂ mesmerize, entrance, stupefy, drug, narcotize, soothe (950), psychologize, anaesthetize* (Bilynsky 2009).

Other contributions to the post-ME reconstructions of the thesaurus came from strings that contained at least two elements dated before 1500 and two or more elements falling on the post-ME period.

Within the general set, a subset of 4,902 strings exhibits multiple constituents (specifically, two or more) with textual prototypes that have documented evidence predating 1500.

Post-ME additions to ME sub-strings could take up diverse composition quotas, which fell short of the ME composition or exceeded it, cf. e.g.

***swing*** *(725) *⊂ [⊂] wield (825), wave (1000), whirl (1290), flourish (1300), brandish (1325), twirl (1598)* and ***cackle*** *(1225) *⊂ [⊂] cluck (1481), giggle (1509), gabble (1577), chuckle (1598), quack (1617), titter (1619), snicker (1694), snigger (1706)* (Bilynsky 2009).

A single string constituent with the OE or ME textual prototype could be complemented with (a) post-ME constituent(s). This situation pertains to cases with the unchanged or reshuffled present-day headword, e.g.

> ***bury*** *(1000) *⊂ [⊆] entomb (1576), enshrine (1583), inhume (1616), mummify (1628);* ***shun*** *(950) *⊂[⊆] evade (1513), neglect (1529), dodge (1568), ignore (1611) vs.* ***shout*** *(1374) * ⊂ exclaim[⊆] (1570), blurt (1573), ejaculate (1578), assert (1604), vociferate (1623), emit (1626);* ***blast*** *(1300) * ⊂ shell (1562), bombard (1598), bomb [⊆] (1688), torpedo (1771), raid (1865), napalm (1950)* (Bilynsky 2009).

In the period recovery of such elements their relevance pertains only to post-ME reconstructions .

The chronological attestation of a lexeme serving as the present-day dominant in a string might occur after 1500, despite the existence of either an associated sub-string or, in certain instances, a solitary constituent whose documented evidence extends before 1500:

> e.g. ***exchange*** *(1300) *⊂ interchange (1374), relieve (1374), substitute (1532), alternate [⊆] (1595)* vs. ***copy*** *(1387) *⊂ illustrate (1526), film (1602), reproduce (1611), photograph [⊆] (1839), snap-shot (1894), microfilm (1940)* (Bilynsky 2009).

The historical restructuring process involves replacing the present-day dominant with its historically earliest counterpart, as determined by textual prototype dating. The lexeme in question may simultaneously serve as the dominant in separate string(s) where it maintains chronological precedence over all other constituents. This systematic arrangement results in the formation of string clusters within the historical thesaurus, with each cluster anchored by its temporally primary constituent.

> e.g. ***glare*** *(1250) * ⊂ [⊆] pierce (1297), menace (1303), scowl (1340), goggle (1380), gaze (1386), fix (1489), glower (1500) vs.* ***glare*** *(1250) * ⊂ glitter (1399), beam [⊆] (1430);* ***glare*** *(1250) * ⊂ pout (1325), scowl (1340), frown [⊆] (1386), gloom (1399), glower (1500);* ***glare*** *(1250) * ⊂ scowl (1340), glower [⊆] (1500)* (Bilynsky 2009).

## 3. COMPARING MODERN-DAY SYNONYM STRUCTURES WITH THEIR HISTORICAL COUNTERPARTS

Permutation, defined as the systematic reordering of elements within a complete system, emerges as a valuable analytical framework for examining the diachronic development of synonymy.

The prerequisites for the study of permutation in reconstructive synonymy is a change of ordinal placements of constituents and a difference in the dates of their OED textual prototypes.

The consequence of the reshuffling of a diachronic remake of a synonymous sequence of lexemes can rest on the assessment of each constituent's weight in the string's contemporary and historical headword.

Within the string's present-day dominant, the weight value (*wi*) of each constituent is calculated using a mathematical formalism that considers two key variables: the constituent's ordinal placement (*i*) and the comprehensive string length (*n*) set by the following formalism:

$$w_i = \frac{n - i + 1}{n}$$

(1)

A string of synonyms may be represented as a vector $\{i\}_{i=1}^{n}$ with its magnitude determined by the cumulative sum of weight factors across all constituents $\{w_i\}_{i=1}^{n}$ .

In the sequence based analyses we have a succession of the dated OED quotations of the constituents $\{j_i\}_{i=1}^{n}$ or their ordinal positions within the historical string $\{y_i\}_{i=1}^{n}$, including one shared position for adjacent constituents dated identically.

The weight formalism for the relative chronological scale (2 ) is $\left\{\overline{w}_j = \frac{n-j_i+1}{n}\right\}$

whereas that for the absolute one (3) is $= \frac{y_i - y_{\min}}{y_{\max} - y_{\min}}\left(\frac{1}{n+1} - 1\right) + 1$

"Then both contemporary and diachronic versions of the string can be presented in terms of lengths of the respective vectors. The difference between the two vectors is taken as permutation factor which is a measure of correspondence between the contemporary string and its historical counterpart" (Bilynsky, Pereymtbida and Altmann 2008).

$$\left\|\vec{w} - \overrightarrow{w^{(y)}}\right\|^2 = \sum_{i=1}^{n}\left(w_i - w_i^{(y)}\right)^2$$

(4)

When constituents maintain equivalent ordinal positions in both historical (relative chronology) and contemporary strings, a perfect alignment occurs between contemporary weight factor values and relative historical weight factor values, resulting in a permutation factor of 0. For absolute historical weight factor values to satisfy this condition, the age differential must remain constant between each consecutive pair of constituents within the string - a circumstance that rarely manifests in practice.

Hence, formalism (3) should give more precise values in an uneven distribution than formalism (2). In both (2) and (3) the oldest lexeme will have the historical weight =1. The words that have this characteristic close to the oldest word will have the weight value nearing 1. Younger words will have this value closer to 0.

Formalism (3) is also more responsive in the treatment of same-year OED textual prototypes that fall on the first or, more often, subsequent positions in a chronological reshuffling of some strings, e.g.

> ***shimmer*** *(1100) *⊂ sparkle (1200), blink [⊂] (1300), glimmer (1399), glitter (1399);* ***lean*** *(950) *⊂ dip (975), shift (1000), turn (1000), sway (1399), tilt[⊂](1399), tip (1399).*

Identically dated textual prototypes may fall on the oldest string constituents: e.g.

> ***handle*** *(1000) *⊂ settle (1000), receive (1300), manage (1561), collect [⊂](1573);* ***ferry*** *(1000) *⊂ pull (1000), tow [⊂] (1000), tug (1225), lug (1375), drag (1440)* (Bilynsky 2009).

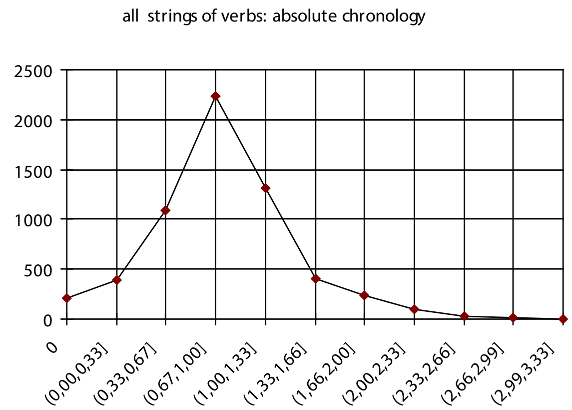
Strings with more constituents are more likely to have the same-year textual prototypes than shorter strings. For this reason as well as due to denser chronological constituent placements in longer strings formalism (3) appears more sensitive than formalism (2).

"The permutation factor can be calculated for each synonymous string that is subjected to a diachronic reconstruction on the basis of the OED textual prototypes of its constituents. The strings tend to group on the strength of this value" (Bilynsky 2009). Individual permutation factor values upon a query to the framework give access to all available examples.

The permutation factor quantifies the diachronic reshaping of the present-day synonymy of a lexeme. In principle, a period selection contains a portion of strings which are not complemented by any elements that are younger or/and older than the set chronological affiliation. At the same time a period selection has a portion of sequences which are sub-strings from the point of view of the overall thesaurus. Such sequences are complemented outside the confines of chronology set for period synonymy.

The graphs for the chronological layers of evidence are based on the insufficiently described notion of the (sub-)strings of synonyms that are obtainable in their diverse variant lengths and chronological affiliation expression.

In the charts shown on the horizontal axes are the adjustably split internals of values of the permutation factor of the present-day string's constituents and their historical reflexes with regard to the numbers of the attested cases of synonymous (sub-)strings shown on the vertical axes ( fig. 1 and, also, fig. 2-4 below). We start with the overall assessment of strings without separating the cases of the diachronically intact and historically floating dominants.



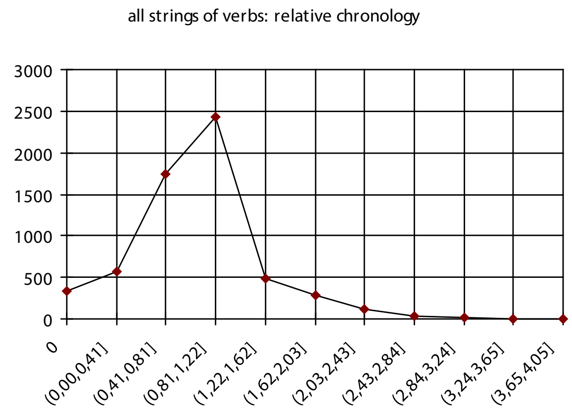
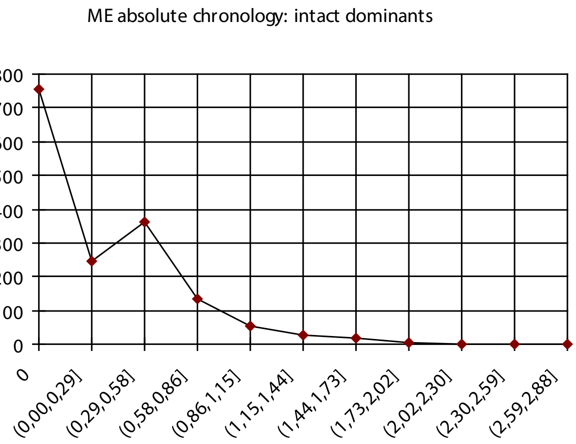




**Figure 1.** Distribution of the permutation factor values in synonymous strings of verbs: axes x – range of the vectors differential values established according to (3) on the upper chart and according to (2) on the lower chart; axes y – the numbers of relevant examples (same axes notifications on fig 2-4 below)

Absolute chronology is more precise than relative chronology at large that is seen by a flatter curve on the upper chart and subsequently better representation of individual permutation factor values.

The threshold between short and long strings may run at the numeric value of the so-called depth hypothesis that is also known as the Miller-Yngwe hypothesis (Hutchins 2012; cf., also, Saaty and Özdemir 2003). It postulates the length of the optimally stored lexical group at seven plus or minus two words. The lower threshold of this interval of lengths forms a border line between shorter and numerically predominant (in the one used for examples for the present study, but not necessarily all thesauri) strings of four lexemes as opposed to longer strings.

The developed framework allows to run queries for smaller segments than the Miller-Yngve depth hypothesis and even for individual lengths of the strings. The obtainable curvatures can be both predictable and unexpected depending on different sets of inventory, chronological or/and etymological variables.



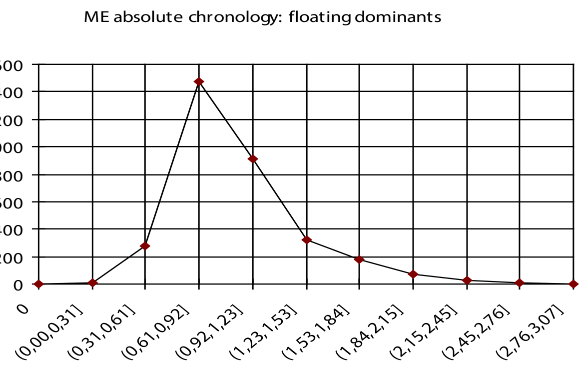
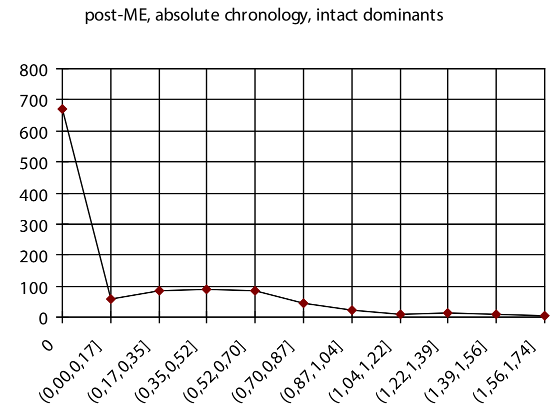




**Figure 2.** Distribution of the permutation factor values in ME synonymy of verbs: upper chart – coincidence in present-day and historical headwords; lower chart – floating headwords (axes notifications as on fig.1)

The charts for the ME (sub-)strings could be found in my previous work (Bilynsky 2009). The charts for the post-ME (sub-)strings have been drawn on the same variables for the present paper. The reflexes of contemporary synonymy to ME synonymy numerically surpass those to post-ME synonymy. In both examined chronological layers shorter strings outnumbered longer ones.

Better representation of the intervals of the said values is characteristic of contemporary synonymy reflected in ME rather than after ME. Longer strings tended to reveal larger values of dissimilarity in the present-day and historical sequencing of constituents than shorter strings. Absolute chronology is more informative than relative chronology at large, but not so much with regard to longer strings In shorter synonymous strings there are more apparent distribution differences in the permutation factor values established while using relative chronology data, especially in the left hand-side part of the obtained curves. In the overall distribution of the permutation factor values medium-range ones prevail over extreme values with some differences between shorter and longer strings that might serve as orientation in a study of strings composition over time (cf., also, fig. 2 and 3 ).



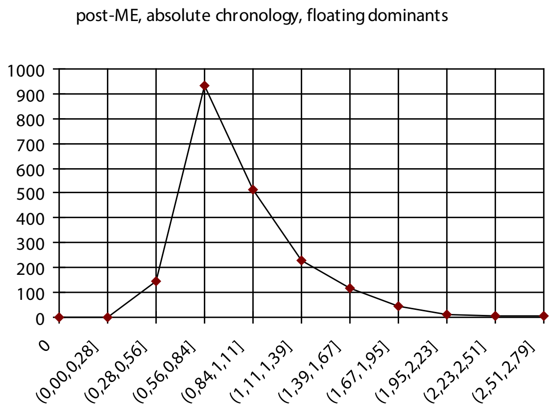




**Figure 3.** Distribution of the permutation factor values in post-ME synonymy of verbs: upper chart – coincidence in present-day and historical headwords; lower chart – floating headwords (axes notifications as on fig.1)

The conducted research seems to be in line with the quantitative turn and statistics as the new normal in present-day linguistic analysis (Buschfeld et al. 2024). This quantitative linguistics context holds true for research on historical reconstructions as well. Its heuristic value lies in the feasibility to make predictions.

This appears to be the case with regards to the suggested tools of analysis for the study of reconstructed lexical relatedness making use of the information side of relative and absolute chronology of the attestation of lexemes in two kinds of vector spaces for the study of language change .over time. Our prediction also lies in the applicability of the framework to diachronic reconstructions of relatedness of de-verbal lexemes modified by derivational constraints as well as diverse stratifications of the entire (de-)verbal database in all the available lexicographic versions and corpora backgrounds as a potential Large Language Model.

The framework can model arbitrary chronological layers of verbal or de-verbal thesauri against a set of diverse variables in selections of arbitrary or justified lengths of strings and de-verbal coinages (cf. sampling fig. 4 and 5) .

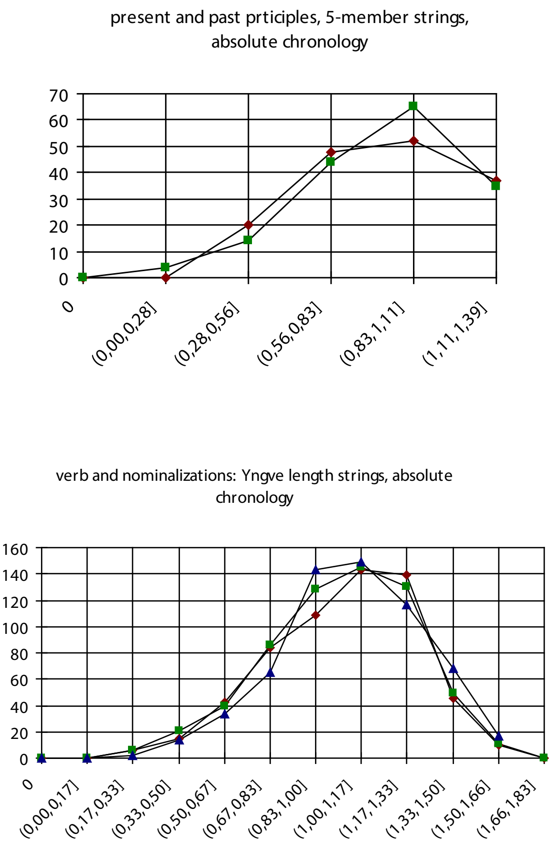
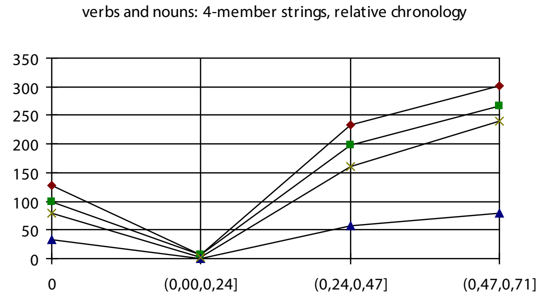



**Figure 4.** Sample queries for the distribution of the permutation factor values from inventory reconstructions of present-day synonymy over time. Upper chart: Five-member strings, rhombus – present participles, square – past participle. Lower chart: Five-to-nine member strings; rhombus – verbs, square – process nominalizations; triangle – lexicalized nominalizations (axes notifications as on fig.1)



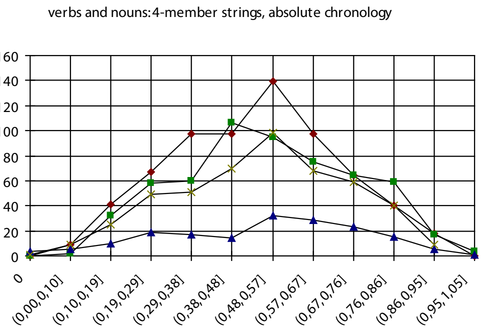




**Figure.5.** Heuristics of chronology modes in temporal permutations of present-day four-member strings: rhombus – verbs, square – nominalizations, triangle – lexicalized nominalizations; cross – agent nouns (axes notifications as on fig.1)

The present framework provides more 'thesaural fodder' for the 'subsistence' of de-verbal morphology in the lexicon. In the future, it may also narrow the gap between the distributional foci on the collected evidence and evidence-based word group analyses.

## 5. CONCLUDING REMARKS

The developed framework allows for data-driven exemplification/curve visualization of the said reconstruction for lengths of strings tested against the composition of the present-day English synonymy according to *Webster's New World Thesaurus*. The procedure is extendible to any of the processed thesauri for English, or their combination(s), and two chronological cross-sections in (OE+)ME and post-ME (de-)verbal (near-)synonymy postulating relevant (dis)similarities between the diachronically permutated and contemporary sequential placements of string constituents. The imposed condition of inventory sameness can be checked with the principle of attested composition in lexicographic sources where synonymy for de-verbal coinages is attested.

Our approach is a result of the fruitful discussion we were privileged to have with the late Professor Altman on the possibilities of lexical reconstructions for the study of synonymy over time (Bilynsky, Pereymybida and Altmann 2009). In fact, the current vector space model was prompted by his ingenious idea concerning the use of the permutation principle in reconstructions of synonymy which could be made informative by the sum total of the differences between the ordinal positions of elements in contemporary and historical sequences, even on limited evidence. This seems to be an unmentioned context (cf. Köhler and Kelih 2021) of his many-sided legacy.

## REFERENCES


Alexander, Mark, and Fraser Dallachy. “Lexis” *in The Routledge Handbook of English Language and Digital Humanities*, eds. Svenja Adolphs and Dawn Knight P. 164-184. NY: Routledge (2020)

Bilynsky, Michael. “Constituents’ permutations in synonymous strings as a diachronic reconstruction problem: the OED Evidence for ME and ENE verbs and deverbal coinages” *6th International Conference on Middle English (ICOME 6) Cambridge* 24-26 July 2008 http://www.english.cam.ac.uk/icome6/abstracts/bilynsky.htm

Bilynsky, Michael, Andriy Pereymybida, Gabriel Altmann. “Weight Factor Formalisms in the Study of Lexical Growth: The Case of Textually Modelled Strings of English Verbs,” *Methods of Text Analysis*. P. 13-39. Chernivtsi, (2009)

Bilynsky, Michael. “Constituents replacement in synonymous stringing as a diachronic reconstruction: the OED Middle English evidence for verbs and deverbal coinages” *Іноземна філологія (Inozemna Philologia)*. Вип. 121, P. 20-42. (2009) http://publications.lnu.edu.ua/collections/index.php/foreighnphilology/article/view/488

Buschfeld, Sarah, Sven Leuckert, Claus Weighs and Andreas Weilinghoff. “How Real is the Quantitative Turn? Investigating Statistics as the New Normal in Linguistics” *ICAME Journal* 48(1):1-21 (2024) https://doi.org/10.2478/icame-2024-0001

de Schryver, Gilles-Maurice. “Lexicographers’ Dreams in the Electronic-Dictionary Age” *International Journal of Lexicography*. 16 (2): 143-199. (2003)

de Schryver Gilles-Maurice, Michael Rundell, Arvi Tavast, Pavel Rychlý, Marko Kokol (panellists) and Simon Krek (moderator). 2023. Round Table on “Large Language Models and AI in Lexicography”. Panel at the 8th Electronic Lexicography in the 21 st Century Conference. Brno, Czech Republic, 27-29 June 2023. https://www.youtube.com/watch?v=dNkksTDYa_s

Hoffmann, Sebastian. “Using the OED Quotations Database as a Corpus – a Linguistic Appraisal” *ICAME Journal* No. 28(-):17-30 (2004)

Hutchins, W. John. “Victor H. Yngve.” *Computational Linguistics*, 38 (3): 461-467 (2012) https://aclanthology.org/J12-3001.pdf

Kay, Christian, Irené Wotherspoon. “Turning the Dictionary Inside Out: Some Issues in the Compilation of a Historical Thesaurus” *A Changing World of Words: Studies in English Historical Lexicography, Lexicology and Semantics* eds. Diaz Vera and Javier E. (Amsterdam: Rodopi). P. 109-135 (2002)

Köhler, Reinhard and Emmerich Kelih. “Gabriel Altmann (1931–2020) Obituary” *Journal of Quantitative Linguistics* 28 (2): 187–193, (2021) https://www.tandfonline.com/doi/full/10.1080/09296174.2021.1902057?scroll=top&needAccess=true

Laird, Charlton Grant, *Webster’s New World Thesaurus*. Prepared by Ch. Laird, updated by D. Lutz. – New York: Prentice Hall Press, – p. 854 (1985)

Proffitt, Michael. The Oxford English Dictionary. (http://www.oed.com), (2018)

Renouf, Antoinette. “Corpus development 25 years on: from super-corpus to cyber-corpus” *Corpus Linguistics 25 Years on,* ed. Roberta Facchinetti New York: Rodopi, P. 27-49 (2007)

Saaty, Thomas L., Müjgan S. Özdemir. “Why the magic number seven plus or minus two,” Mathematical and Computer Modelling, 38: 233-244 (2003) https://doi.org/10.1016/S0895-7177(03)90083-5

The Historical Thesaurus of English. https://ht.ac.uk/ (2010)