Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895
https://llrjournal.com/index.php/11

The Evolution of English: A Statistical Analysis of Language Change Tracking Linguistic Change

[1]Ambreen Zehra Rizvi, [2]Nazra Zahid Shaikh

[1]Assistant Professor, Faculty of Engineering, Science & Technology, Hamdard University Main Campus, Karachi, Pakistan. ambreen.rizvi@hamdard.edu.pk
[2]Senior Lecturer, Department of English, Faculty of Social Sciences and Humanities, Hamdard University Main Campus, Karachi, Pakistan. nazra.zahid@hamdard.edu.pk

Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895

Abstract

This paper examines how statistical analysis techniques, such as Markov chaining, regression analysis, and corpus linguistics methods, can be applied to the evolution of English vocabulary and grammar. Using large historical text databases (e.g. Google Ngram, COHA, EEBO) we analyze lexical frequency changes, grammatical simplifications, and interaction with sociolinguistic factors in language change. Among the key findings are that vocabulary growth shows exponential patterns, grammar tends to become more analytic, and external influences such as technology and migration speed up language change. The work shows how quantitative methods can be and need to be combined with traditional philological methods in historical linguistics.

Keywords: Language change, statistical linguistics, Markov models, regression analysis, corpus linguistics, grammaticalization

Introduction

Language is one of the most complex and dynamic cultural products of humanity yet constantly changing due to social, cognitive and technological factors. As a living organism, English has evolved from its Germanic origins in Old English into the world's common language in the digital age. Traditional methods for investigating language change have been largely grounded in qualitative philological methods, which have entailed manual scrutiny of historical texts and comparison to reconstruct linguistic features. But large parallel digitized corpora and advanced computational methods now give researchers a possibility to study language change using sound, data-driven methods. In this paper statistical modeling techniques are used to examine three basic questions about the historical development of English.

Locomotive Lexicon: When and why does the English vocabulary grow, shrink, or change in character over long periods of time? We study trends in the birth of new words (or neologisms), the death of words (or obsolescences), as well as shifting word frequencies in various registers and dialects.

Grammar Simplification/Complexification: Are there tendencies for English grammar to become simpler or more complex throughout time? We examine changes in morphology (e.g., the loss of case systems), syntax (e.g.modified word orders) and new grammar constructions.

Preditive Modelling: To what degree can statistical approaches provide reliable predictions

of the evolution of language? We offer an analysis for the predictive power of different statistical models in the prediction of lexical adaptation rates and grammatical changes.

The paper uses three large historical text corpora, cumulatively covering more than five centuries of English:

· **Google Books Ngram dated Corpus (1800-2019):** This package, in conjunction with the corpus-derived were and was represent the frequency data of 500 billion words across different English dialects, featuring a comprehensive coverage of the lexical and grammatical changes in the printed media.

· **Corpus of Historical American English (COHA) (1810-2009):** Contains a well-balanced 400million-word historical corpus based on a wide range of American print sources, and allows genre-based study of the historical development of American English.

· **Early English Books Online (1473-1700):** Having virtually the entirety of English printed materials produced during this critical time period available in electronic format is crucial for studying the birth of Modern English out of Middle English.

The use of statistics to the datasets is a major progress compared to standard methods in many ways. The ability to sense small and gradual changes that may elude manual observation is one of the key attributes of quantitative assessment. Second, by using computational techniques, textual corpora can be processed at scales that are impossible for human researchers to read in their entirety. Third, statistical modeling yields evidence based measurements of change that can be replicated and tested by other researchers.

This study is located at the confluence of a number of subfields: historical linguistics, where the theoretical apparatus of language change is developed, corpus linguistics, where the data are supplied for analysis, and statistical modeling, which provides techniques for identifying and interpreting patterns in these data. By integrating these methodologies, we hope to generate findings that are rooted in linguistic theory and also supported by large-scale data.

The importance of the present research is not limited to academic linguistics. The knowledge of the nature of language change has potentially practical applications in natural language processing, in which historical language data can be used to train machine learning models; in education, as it has implications for the teaching of English (either as a native or foreign language) and in language policy where it can help to shape decisions on language standardization and preservation. In addition, in such digitalized and globalized world, where

English is also evolving naturally, objective and quantitative approaches to monitor such changes become even more important for researchers, teachers and technologists.

Methodology

Markov Chain As a Model For The Observed Lexical Shifts

In this study, we use Markov chain models, a statistical theory that interprets language as a state transition system, to quantify the dynamics of English vocabulary. In this model:

· Words time intervals States are time periods (for example, decades and centuries) in which word frequencies are recorded.

· Transitions between the states are determined by probabilities for a word to become more frequent, less frequent, or will maintain its frequency in the language.

This is a framework that gives us a way to describe not just how words change, but to illustrate the fact that they change. Because future states depend on the current state (but not on the entire past), it is computationally efficient and we show that it captures relevant trends of language change.

Applications Tracing Vocabulary Change

An Example of Innovations in Terms approximately: Modeling the Emergence of TechnologyRelated Terms

· For example, the term "internet" was virtually nonexistent before the 1980s, but its usage skyrocketed after 1990. A Markov model through historical record, can predict the chance of such term changes from being low frequency to high.

· Similarly, it would also be interesting to look at terms such as “smartphone” (which can be post2007) and “selfie” (post-2010) to see how quickly our language for names of new technology stabilizes.

The Finding of Archaic Words in Declining

· The words "thou" and "hath" were also used as the second person singular pronoun and third person singular form of "have," respectively, in Early Modern English which is no longer spoken.

· Modelling their time-dependent transition probabilities allows us to pinpoint when these terms started disappearing and whether they evolved according to a predictable decay pattern.

Benefits of Markov Models in Lexical Analysis

· *Scale*: Can handle large historical corpora (like Google Ngram) well.

· *Interpretability*: Transition matrices tell us which words are getting more or less popular.

· *Predictive Power*: Are able to predict future word likelihoods given the current transition probabilities.

Limitations and Caveats

· ***Memoryless Hypothesis*:** The Markov property does not take long-term historical effects into account, which may over-generalise for the slow drift in language.

· ***Sparseness of Data*:** Probabilities of using unusual words may not have enough historical examples to estimate.

· ***Context Blindness*:** Fails to place a word in the context of considering the semantics or syntax when the variation has occurred.

To address these limitations, we extend Markov models to include hidden Markov models (HMMs) with latent variables (e.g., social factors) and smoothing for sparsity.

This methodological framework offers a solid foundation for a study of the evolution of words, which can highlight mechanisms through which external factors, such as technology and social change, promote or impede lexical change in English.

Regression Analysis of Grammatical Simplification

In order to analyze systematically these long-term simplification trends in English grammar, we use linear (in conjunction with logistic as needed) regression. We can use these statistical methods to measure and model the diachronic development of both the morphology and syntax of the language over a range of centuries.

Linear regression analyses are especially successful at capturing quantitative decay in grammatical items, that is of the kind whereby grammatical phenomena erode gradually (Goldberg, 2006), such as the declining usage of inflectional morphemes. For example, we follow the attrition of strong verb clustering (e.g., “help-holp-holpen” starts to converge towards “helphelped-helped”) over time. The models provide information not only on the speed of this morphological simplification but also on the linearity or discontinuity of the dynamics. Analogously, we use logistic regression to investigate differences between belonging to the category or neutral form, as measured by the total loss of inflectional endings of some change or the stabilization of some previously variable syntactic pattern.

A major component of our analysis is the transition from synthetic to analytic constructions in English grammar. We take a closer look at this through the lens of growing preference for periphrasis (with the simple future “shall” being replaced by “will” in most environments, or the spread of progressive aspect constructions like the one as in “is going”

would have been “goeth”). Many of these changes can be operationalized by the frequencies of entrenchment (cf. Entrenchment and entrenchment theory) of competing constructions in our diachronic corpora and by trajectory modeling.

Our regression models include a number of thoughtfully chosen predictors that are designed to represent the intricate dynamics of grammatical change:

Dependent Variables

· Frequency of some particular grammatical forms (e.g. use of strong verb forms per 10,000 words)

Proportion of analytic vs. synthetic constructions in similar syntactic contexts · The presence/absence of specific morphological marks in given periods of time

Independent Variables

Temporal Factors

· Continuous Variable of Cultivation (year of text generation period)

Periodization variables (EModE vs. LModE etc.)

Sociolinguistic Factors

Urbanization rates (a measure of dialect contact and innovation diffusion)

· Literacy rates (indicating standardization impulses) · Population movement metrics (language contact situations)

Text-Type Variables

· Genre variegations (for example, formal-classical vs. informal-modern registers)

· Characteristics of the authors (if provided)

The regression models are organized hierarchically to accommodate the structure of our data (words nested within texts; texts nested with periods). We use mixed-effects models to account for the inherent variance of historical texts, but are still able to observe larger patterns. Model diagnostics encompass screening for multicollinearity between predictors, residual examination to confirm model assumption and cross-validation to evaluate predictive precision. To help mitigate the limitations imposed on us by the historical nature of the data, we employ a number of methodological controls:

· Text normalization to handle spelling variants

· Sampling methods for equally representing over time

Bootstrapping to get confidence intervals for our coefficients.

· Sensitivity analysis to check the stability of the results to a variety of model specifications

The findings of these regression analyses help to confirm statistically a number of the crucial features of grammatical simplification in English, including:

-The gradual loss of case-marking inflectippns

-Verb paradigm regularization

-The spread of analytic formations on the account of syntheticalержоп.

-The varying rates of change in grammatical subsystems

By measuring such changes and by determining their sociotemporal correlates, our regressionbased approach provides new insights into the mechanisms and pathways of grammatical simplification in the history of English. The statistical models corroborate patterns found in traditional philological studies and uncover hitherto unnoticed regularities in the dynamics of grammatical change.

Corpus-Based Trend Detection

The study of language evolution aims at systematically studying large patterns of change over time, using methods of time-series analysis. The analytical method also allows to detect and measure two major types of changes: lemmas' replacements and grammaticalization. The quantitative approach allows to find patterns of language change that would have been difficult to identify with traditional qualitative methods of analysis.

Lexical replacement analysis captures the competition between words in a given semantic field over a long period of time. Analysis of the evolution of technical terms, beginning with “wireless” and “radio,” and continuing with “Wi-Fi” reveals the gradual replacement of early terminology with newer language. Time-series models do not only represent the eventual replacement results, but also intermediate states involving lexeme competition, and concurrent lexeme usages. Change-point detection methods reveal historical tipping points that mark the emergence of new terms to the fore and the decline of terms in established use, and survival analysis methods model the longevity of legacy vocabulary in specialist discourse.

Grammaticalization research focuses on the process by which words develop over time into grammatical markers. The curves that fit the patterns of change, such as the one from going to to gonna, use curve fitting to depict changes that are characteristic of the S-curve in grammaticalization. Quantitative measurements follow development of phonological reductions and semantic weakening, represented, respectively, with phonetic length, collocational frequency and syntactic distribution which are treated as parameters of progress

of a grammaticalized form.

It includes several advanced time-series techniques in the analytical framework:

-ARIMA models Balance the predictability of short-term lexical usage with a longer-term trend.

-Methods of spectral analysis demonstrate cyclicous patterns in language utilisation.

- Dynamic time warping (DTW) algorithms map similar linguistic developments from different time periods

-Significant changes in usage patterns are detected by the Bayesian change-point detection techniques

Validation protocols that contribute to analytical robustness are:

· Cross-corpus validation across multiple ancient populations

· Bootstrap resampling is used to provide estimates of confidence interval for trends.

· Sensitivity analyses explore the influence of differing smoothing parameters

· Correlating historical events look at outside forces on language change

Time-series analysis has several advantages for linguistic study:

-Distinguishing actual linguistic change from temporary variations in the use of language

-Quantifying characteristic rates of change in linguistic features

-Modeling of the propagation of change over spoken registers and dialects

Framework For Testing Mechanisms Of Language Change

Some Important Patterns Emerge From Preliminary Analyses

· Lexical substitution undergoes regular logistic dynamics for semantic categories

· There are clearly different stages in the process of grammaticalization

· Social variables like urbanisation promote subtypes of language change

· Distinct grammatical subsystems change at their own rates

Time- series methods combined with auxiliary analyses form a whole approach of English language evolution examination. In addition to complementing the traditional philological approach, this quantitative view also offers new ways to study the nature and causes of linguistic change. The method allows to subject theoretical claims regarding the pattern of language development across historical periods to rigorous empirical testing.

Key Findings

Vocabulary Enrichment has an Exponential Growth Rate in its Size and they could be identified as following:

Neologisms: First appearance of a word or phrase, in any section of the dictionary or in any location of the published or online text, since 1990.

Semantisation: Old words that then receive new meanings to fit to new cultural and technological realities

Loanwords: Adoption of loanwords from contact languages, with frequency and source languages depending on historical conditions

Temporal analysis further suggests that word growth is through waves and not linear. Growth curves are punctuated by large-scale historical events, such as wars, migrations, and technological transformations. Will they do the same in 21st-century English?♀♀♀♀The two World Wars of the 20th century, for example, created a large uptick in both technical vocabulary and loanwords from languages spoken among combatants.

The power law model fits better for content words (nouns, verbs and adjectives); the function words (prepositions and conjunctions) are characterized by a much more stable frequency rate over time. This contrastive growth displays a crucial factor to the rise in lexical density in English texts throughout the centuries, being observed in corpus analyses of Old English and Modern English texts.

Methodologically, these results derive from the meshing of various methods of analysis:

· Markov behavior models the likely trails of word adoption

· Regression indicates important predictors of the vocabulary increase

· Long-term patterns and inflection points are identified through time-series analysis

· Comparative historical period analysis is used to bring history to bear on period rc effects

The consequences of these findings do not, however, spare historical linguistics, lexicography, natural language processing, or teaching practice. The observed patterns of vocabulary growth receive empirical support based on theories of language change and have predictive power in terms of future lexical evolution.

Analyticization – Grammatical Change Towards Analytic Structures

The diachrony of English grammaticalization is unambiguously leading towards increased analyticity, that is, increasing use of noninflectional morphology and periphrasis. Comparisons between Old English and Modern English reveal an impressive simplification of the morphology, and especially of the nominal system. Old English had five cases of nouns (nominative, accusative, genitive, dative, and instrumental), whereas the present-day English case system has fallen back on only two cases (subjective and objective) and uses

prepositional words and strict word order to convey the other aspects of grammar. This is one of the most radical structural changes in the history of the language.

The verbal system exhibits a parallel development to that of analytic expression. The use of auxiliary verbs has increased fivefold in English since the 18th century, especially in "do-support" constructions in questions and negatives. This development of analytical verb forms parallels the erosion of inflectional verb endings of earlier periods of the language. Modernisation reveals, in addition, extensive regularisation of irregular verb paradigms, many of them undergone no longer or the weak inflection type.

There is a striking loss of the passive voice in formal writing registers, because the COHA-based counts showed that over the past 200 years, the frequency of the passive decreased by about 40%. This change seems especially robust in academic and journalistic registers, and may reflect stylistic preferences as much as grammatical change. Decline in passive use synchronizes with the ascendancy of the prescriptive grammar tradition assigning greater "directness" and "vigor" to "active voice".

A number of observables describe this analytical deformation:

Morphological Simplification

· Reduction of nominal case endings from five to two (common and genitive)

· Elimination of sexual gender implications of grammatical gender differences

· Verb paradigms Regularization

Periphrastic Expansion

· More frequent use of modals (will, would, can, could)

· Increase of the progressives format.effects of the progressives format.

· Creation of elaborate tense-aspect systems involving auxiliaries

Syntactic Restructuring

· Maintenance of the SVO (Subject-Verb-Object) word order

· More use of prepositions to establish relationship between words (example: The fight was between two male calves).

Grammaticalization of phrasal compounds (such as “be going to” future)

This transition to analytic structures demonstrates systematic patterns within the grammatical subsystems. The earliest and most comprehensive changes are in nouns, the loss of most of the noun case system probably being complete by Early Modern English. The verbal system has undergone a similar, but less drastic, evolution with some remnants of the

synthetic forms (for instance the 3rd person singular -s) remaining in Modern English. The auxiliary is still evolving, and newer contractions (like “gonna”) are the most recent in a history that has been moving in this analytic direction.

Quantitative analyses of such changes have shown that some registers and dialects change at different implicit rates. Formal writing in English eschews most of the few remaining synthetic forms, like the nail (third person singular present indicative) in autocatalytic networks. Taken as a whole, the trend is clearly towards more analyticity, thereby confirming predictions that English is a language that is in the process of crossing over from normatively characterized as synthetic to normatively characterized as analytic.

Social Determinants of Linguistic Change

English lexis and syntax has evolved, not independently, but in response to a whole range of external sociocultural and technical influences. Our quantitative analysis shows that there are three major external forces that have caused a significant acceleration in language change over the last few centuries: technological innovation, the effects of globalization, and the rise of digital communication media.

The most rapid linguistic changes have been made due to technological advancements. The transformation of “text” into a verb is a case in point, and corpus data provide clear evidence for a 1,200% rise in verbal use from 2000 to the present. This same process seems to be at work with other technology-mediated conversions such as google as a verb and friend as a verb in the context of social media. And these changes are a symptom of the language’s ability to adapt itself to new technological realities in its grammatical structure.

The globalization has greatly affected the lexical stock of English, as the increased language contact has provided a better opportunity for this language to adopt new loans. PATTERNS OF LOANWORD INTEGRATION patterns of loanword integration follow strong historical waues for periods of cultural exchange:

· Hindi loanword from colonial times (“shampoo,” “bungalow”)

· Post-war Japanese adoptions ("emoji," "karaoke")

· Words for post-20th-century technology from many languages ("algorithm" from Arabic by way of Latin).

The digital revolution through social media platforms has brought new changes to patterns of linguistic innovation. Acronyms and shortened forms ("LOL," "DM") show typical logistic curve growth patterns in the corpus data, progressing toward saturation within

5-7 years of initial introduction. These digital forms of communication have a number of attributes:

· Dramatic phonological reduction (‘because’ → ‘bc’)

- Platform-specific semantic narrowing (“tweet” as noun and verb)

· Shorter periods of obsolescence than traditional knowledge of vocabulary

Demographic evidence uncovers further extra- and intralinguistic language changes:

Urbanization Effects: The condition of an entity that at more dense places of aggregation of people leads to a higher speed of coining neologisms

Educational Expansion: Standardization pressures versus colloquial innovation

Patterns Of Migration: the diaspora as channel for loanwords

Media Influence: The rapid spread of inventions by TV and film

This interaction of external pressures and the internal gen-erative language structures exhibits regular mathematical behav-ior. Terms from technology are often characterized by a quick initial expansion, which stabilizes rather quickly, while loan words often take a longer time to adapt phonologically and morphologically. Social-media innovations show the ste… run-up in adoption curves but plummet the quickest to fatality in the general lexicon.

These forces from 'non-linguistic' spaces conspire together to refashion English in several ways:

· Lexical growth through the introduction of new terminology requirements

· The influence of the contrastive factor in language contact situations

· Orthography in innovation in digital communication arenas

· Semantic transfer by technological resemiotization

The quantifiable effect of these external motivators represents a strong argument for the role of the sociocultural environment in language change. Quantitative models that take into account both linguistic and extralinguistic factors make it possible to obtain a significantly better predictive accuracy in the case of the trajectory of a language change, compared to purely intrinsic models of linguistic evolution.

Discussion

Combining Statistical And Classic Linguistic Techniques

The quantification of language data has led to good agreement with traditional philological results and unearthed new patterns of language change. Statistical results confirm a number of wellknown tendencies of English history in terms of numbers of vowels in a language,

such as the Great Vowel Shift(1350- 1600). The time-series data on vowel latitude for the historical texts provide evidence for the gradual, flag-systematic nature of these phonetic shifts, and the spectral data show that vowel movements are timed collectively, which is as one critical pattern within this pivotal change in English phonology.

Computational methods have not only validated traditional approaches to the study of language evolution, they have made important new discoveries about the processes that drive grammatical evolution. Optimal estimation of the origins of syntactic or morphological innovations suggests S-shaped adoption curves rather than straight linear transitions. This paralleling of language change across grammatical subsystems is indicative of language change as a process that occurs via social diffusion like other types of cultural change. The S-curve depositional model is most clearly expressed in:

Morphological Decay: The second person singular "-est" ending of verbs (e.g."thou goest") shows the typical pattern of slow initial decline, fast mid-period decay and slow late stabilization.

Micro-Variation And Syntactic Change: The case of do-support in questions and negatives, which shows an astonishingly universal S-shaped development across dialects

Lexical Substitutions: Rival forms (e.g., hath/has) show predictable periods of coexistence prior to tipping points.

The conjunct methodology provides complementary advantages, compensating for the deficiencies of single approaches:

· Traditional philology can achieve an even deeper context sensitivity for historical change
· Can make systematic comparison across various variants using statistical analysis
· Patterns not apparent to manual analysis can be exposed by computational methods
· Quantitative approaches permit the testing of opposing theoretical models

This "unification" has successfully obviated many controversial arguments in historical linguistics. For example, the statistics heavily favor a wave model of diffusion of linguistic change over a family tree model for nearly all the common grammatical innovations for English. Regression analyses also highlight the crucial influence of population density and social network structure in shaping the rate and direction of linguistic change.

The methodological synthesis also provokes important theoretical questions for future investigation:

- On a relationship between the slope of S-curves and importance of innovations
- The guess at alphabet size and population size
- The hand in hand written standardization and spoken innovation
- The modeling of competing changes in subcorpus grammars

These results validate the value of combining statistical methods with traditional linguistic analysis in revealing a more complete and nuanced picture of language change. The quantitative approach also yields quantifiable evidence for patterns that have long been hypothesized, but formerly described only qualitatively, and it reveals novel systematicities in linguistic process for which evidence was previously lacking. This blending of verification and discovery is the most important advance in the scientific study of language change.

Methodological and Data Caveats

Their research results have to be interpreted in the context of some important methodological limitations that may limit the generalization of their results. The most relevant restriction is the biases of intrinsic corpora, in particular the fact that written forms of the language are much more likely than spoken form to be represented in the available historical datum. This bias towards the Written Language results in contrastsive empirical ‘voids’: much of the phonological and syntactic apparatus of the colloquial has not been built up from the text. Written historical corpora are based on records of formal registers, educated usage, and standard varieties, which filter analyses of language change through a conservative representation.

Data sparsity is another significant challenge, especially in earlier historical periods. Textual sources for the English language grow fewer and more fragmented in the transition from Old to Middle English, and by the time of the Norman Conquest in 1066, the language's superstrate was Norman, a variety of Old French. The most important development in the history of the English language between the mid-14th century and the late 15th century was the Great Vowel Shift, which radically transformed the pronunciation of English, and created a number of habits that were increasingly deviant from European norms.

- Gaps of representation: in early periods, many kinds of texts and many social registers are not documented.
- Statistical reliability: Rare, low-frequency phenomena may seem to be exceptionally rare simply because of small sample sizes
- Regional variation: Some dialects and local varieties are under-attested in remaining

texts

- Temporal resolution: Taking exact time for an event is something that is decreasing the further back we research.

Further restrictions stem from the nature of historical language data:

- Standardization effects: Stereotyping effects of prescriptive grammars and dictionaries may hide LS variation
- Text survival bias: the random preservation of some texts and loss of others introduces sampling bias.
- Digitization artifacts: OCR errors and transcription conventions affect machine-readability of early texts
- Metadata limitations: Incomplete information about authorship, provenance, and sociolinguistic context

The interplay of these restrictions poses a complicated problem on the interpretation of statistical information. For example, supposed changes in grammatical patterns could be manifestations of changes in text types rather than in language itself. In the same vein, lexical coinages could seem more instantaneous in the record than they did in life, because the evidence of language is gappy.

These limitations lead to several important caveats for the current results:

· Rates of change reported here may be underestimates of spoken language flux

· Earlier period estimates may represent elite usage rather than overall tendency

· "Drastic changes" in apparent time may actually be resulting from data lapses rather than linguistic shifts

· You almost must be underrepresenting regional variation in the models

Future study needs in respect to these shortcomings are:

- Construction of more balanced historical corpora
- Better Statistical Methods for Sparse Data
- Integration of sociolinguistic metadata
- Multimethods combining qualitative with quantitative

Although these restrictions restrict some interpretations, the main results survive even under controlled-methodological caution analysis. The patterns are given further belief through their survival under a wide range of analytic techniques, and their correlation with independent records of language change processes.

Future Research Directions

The present study suggests a number of exciting directions in which quantitative historical linguistic studies may develop new methods and broaden the scope of comparison. More relevantly, neural language architectures are especially transformational in breaking new ground for the study of finer-grained language change patterns. Pretrained models based on the Transformer architecture, trained on large corpora of historical text, would be able to represent minute syntactic and semantic developments that elude statistical methods, and attention indictators functions could uncover hitherto undetected patterns of language contact and change spread through time.

Multilingual comparative approaches are also important subjects for future research, and a comparison between closely related Germanic languages that have undergone different historical paths is one of the most vital next steps. A comparative overview of English and German grammatical development might contribute to illuminating the following issues:

- Inequality of loss of inflectional morphology: though both languages ultimately stem from the same Proto-Germanic origin, modern German preserves considerably more morphology of case and gender agreement.
- Nonparallelisms in the Learnability of (Missing) Verb-Second Across Germanic Dialects: Evidence from Language Change Differing in Synchrony and Diachrony
- Degree of Analyticization – How close analytic languages are to synthetic ones, and how quickly each has been changing into analytic ones.

Other interesting research directions are the following:

Computational Historical Sociolinguistics represented by:

· The unification of social network analysis and language change models

· Quantitative analysis of changes in diffusion across various social groups

· Geospatial approach to patterns of migration and dialect contact

Enhanced Corpus Development

· Constituting balanced historical corpora covering various registers

· Enhanced representation of spoken language in plays, letters and transcripts

· Growth of Pre-1800 Collections to address sparsity of data

Interdisciplinary Methods

· Use of models of biological evolution to study language change

· Modifying physics-based diffusion equations for lexical spread · Application of

econometric methods to the analysis of rates of change

Predictive Modeling

- Development of forecasting models for language change
- Selection of features in language which are most prone to change
- Simulation of the potential evolutionary paths under alternative social conditions

All these future developments together should help overcome certain methodological shortcomings and enrich the theoretical and empirical base of quantitative historical linguistics. Combining neural with traditional statistical methods may be especially useful for modeling complex, non-linear change trends that proved resistant to traditional analytic methods. In the same vein, systematic multilingual comparison may serve in future work to unravel crosslinguistic developments from general typological shifts in grammar evolution.

In practice, the research directions described will demand the following:

· Jointly developed computational resources for shared use

· Annotation practices should be standardized across languages

· Development of interoperable historical language datasets

· Education programs using both linguistics and data science methods

The higher objective of such future research efforts will be to establish more comprehensive, more empirically sound models of language change that can reflect internal linguistic as well as external social correlates on various timescales and in multiple subsystems of linguistic structure. Modeling along these lines would help us better understand one of the most basic cultural evolutionary processes of humanity, which is the changing of language through time.

Conclusion

The open application of these statistical techniques to historical linguistic data has provided valuable results with respect to the systematic nature of the evolution of the English language. Quantitative analysis reveals that these changes proceed along observable and often predictable pathways on multiple timescales and subsystems of language. This examination has three key results:

In the first place, the expansion of the vocabulary is an exponential growth in which the lexical innovation rate increases with new technologies and new culture. This pattern of adaptive lexical expansion continues into the present day, with the proliferation of domain-specific lexicon in domains such as artificial intelligence and biotechnology. A mathematical

model of these expansion patterns offers a form of prediction for future vocabulary development, as well as the possible paths of AI driven neologisms and digital communication expressions.

Secondly, grammatical evolution exhibits a robust tendency in a particular direction - towards analytic constructions - supported by a number of independent lines of quantitative evidence. The longitudinal investigation will show that there is a systematic loss of inflectional morphology, an increase in the use of periphrasis, and escalating use of word order and function words. These patterns are consistent with S-curve adoption patterns found for other grammatical features and historical stages. Recognition of these patterns allows us to more adequately model developing phenomena in grammar and to provide an accurate record of endangered constructions before they are lost from a live language.

Third, Turkic contact-induced language change follows observable and sometimes even predictable sociocultural conditioning. Statistical analysis verifies that the technological progress, processes towards globalization and the transformations in media, are the main drivers for linguistic change. The latter couple with external forces to give rise to intricate but measurable patterns of change.

These observations have implications outside of academic linguistics, for use in various settings:

Language Teaching: These observed patterns can be used to help formulate better pedagogical strategies that align with typical learning trajectories.

Natural Language Processing: Historical change models can contribute to the improvement of machine learning systems by including an evolutionary dimension

Policy For Language: Quantifying evidence for decision making on standardization, language communities and preservation efforts

Lexicography: Predictive models can help in the selection of emerging words for future inclusion in dictionaries

Research implications

The recommended avenues for future research based on the conclusions drawn were:

· Further refinement of prediction models that include linguistic and social factors

· Quantitative methods applied to other languages (and families of languages)

· Combining neural language models with classical statistical ones

· Developed Historical Corpora Data collection can be improved responsive to current gaps

in available information

The success of statistical methods in the study of language change is therefore confirmed young historical linguist. When combined with traditional philological procedures, the other techniques may contribute toward a fuller and more empirically based knowledge of language change. By no means limited to historical practices, this interdiscursive methodology sheds light on how English has evolved, enabling scholars and stakeholders to predict and influence its future course. The systematic patterns we see here show that language change, although a very complex phenomenon, does follow principles that are formalizable and that can be scientifically dealt with.

References

Michel, J.-B., et al. (2011). "Quantitative analysis of culture using millions of digitized books." *Science*, 331(6014), 176–182.

Davies, M. (2012). "The Corpus of Historical American English (COHA): 400 million words, 1810–2009." *ICAME Journal*, 36, 9–12.

Bybee, J. (2015). *Language change*. Cambridge University Press.

Lieberman, E., et al. (2007). "Quantifying the evolutionary dynamics of language." *Nature*, 449(7163), 713–716.

Bergs, A., & Brinton, L. J. (Eds.). (2012). English historical linguistics: An international handbook (Vol. 1). De Gruyter Mouton. https://doi.org/10.1515/9783110251593

Biber, D., & Gray, B. (2016). Grammatical complexity in academic English: Linguistic change in writing. Cambridge University Press. https://doi.org/10.1017/CBO9780511920776

Blank, A. (1999). Why do new meanings occur? A cognitive typology of the motivations for lexical semantic change. In A. Blank & P. Koch (Eds.), Historical semantics and cognition (pp. 61-90). De Gruyter Mouton. https://doi.org/10.1515/9783110804195.61

Davies, M. (2012). The Corpus of Historical American English (COHA): 400 million words, 1810-2009. ICAME Journal, *36*(1), 9-12. https://doi.org/10.1515/icame-20120002

Hilpert, M. (2013). Constructional change in English: Developments in allomorphy, word formation, and syntax. Cambridge University Press. https://doi.org/10.1017/CBO9781139004206

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, ... & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books.

Science, *331*(6014), 176-182. https://doi.org/10.1126/science.1199644

Nevalainen, T., & Traugott, E. C. (Eds.). (2012). The Oxford handbook of the history of English. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199922765.001.0001

Perek, F. (2015). Argument structure in usage-based construction grammar: Experimental and corpus-based perspectives. John Benjamins. https://doi.org/10.1075/cal.17

Tagliamonte, S. A. (2012). Variationist sociolinguistics: Change, observation, interpretation. Wiley-Blackwell. https://doi.org/10.1002/9781444300441

Traugott, E. C., & Dasher, R. B. (2001). Regularity in semantic change. Cambridge University Press. https://doi.org/10.1017/CBO9780511486500