Check for updates

ESSAY

# REVISED Open Problems in Computational Historical Linguistics

[version 2; peer review: 5 approved]

Johann-Mattis List [1,2]

[1]Chair of Multilingual Computational Linguistics, University of Passau, Passau, Bavaria, 94032, Germany
[2]Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, 04103, Germany

**v2** **First published:** 20 Nov 2023, **3**:201
https://doi.org/10.12688/openreseurope.16804.1
**Latest published:** 29 May 2024, **3**:201
https://doi.org/10.12688/openreseurope.16804.2

## Abstract

Problems constitute the starting point of all scientific research. The essay reflects on the different kinds of problems that scientists address in their research and discusses a list of 10 problems for the field of computational historical linguistics, that was proposed throughout 2019 in a series of blog posts (see http://phylonetworks.blogspot.com/). In contrast to problems identified in different contexts, these problems were considered to be solvable, but no solution could be proposed back then. By discussing the problems in the light of developments that have been made in the field during the past five years, a modified list is proposed that takes new insights into account but also finds that the majority of the problems has not yet been solved.

## Keywords

historical linguistics, computational linguistics, open problems, scientific problem solving

This article is included in the European Research Council (ERC) gateway.

This article is included in the Languages and Literature gateway.

**Open Peer Review**

**Approval Status** ✓ ✓ ✓ ✓ ✓

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **version 2** (revision) 29 May 2024 | ✓ view | | | | ✓ view |
| **version 1** 20 Nov 2023 | ? view | ✓ view | ✓ view | ✓ view | ? view |

1. **Jeff Good**, University of Buffalo, Buffalo, USA
2. **Thomas Brochhagen**, Universitat Pompeu Fabra, Barcelona, Spain
3. **Siva Kalyan**, The University of Queensland, Saint Lucia, Australia
Australian National University, Canberra, Australia
4. **John Nerbonne**, University of Groningen, Groningen, The Netherlands
Albert-Ludwigs Universität, Freiburg, Germany
5. **Michael Pleyer**, Nicolaus Copernicus University in Toruń, Toruń, Poland

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the Horizon Europe gateway.



This article is included in the Linguistic Diversity collection.

**Corresponding author:** Johann-Mattis List (mattis.list@uni-passau.de)

**Author roles: List JM**: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Project Administration, Resources, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This project has received funding from the European Research Council (ERC) under the [European Unions Horizon Europe research and innovation programme] (Grant agreement No. 101044282).
*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2024 List JM. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** List JM. **Open Problems in Computational Historical Linguistics [version 2; peer review: 5 approved]** Open Research Europe 2024, **3**:201 https://doi.org/10.12688/openreseurope.16804.2

**First published:** 20 Nov 2023, **3**:201 https://doi.org/10.12688/openreseurope.16804.1

**REVISED** **Amendments from Version 1**

The new version provided here does not alter the main idea of the essay substantially. However, following several points raised by the reviewers, I have tried to make sure that these are addressed in the new version. The most important modifications can be found in the framing with respect to the question of how language evolved for the first time, which was criticized by many reviewers, and also in the correction of multiple typos. In addition, the figure caption for the main figure was modified to make sure that the symbols in the figure make more immediate sense (as was also brought up by the reviewers).

**Any further responses from the reviewers can be found at the end of the article**

## Introduction

The driving force of many scientific inquiries are problems – unsolved or *open problems*. We observe phenomena that we cannot explain, we do not understand how phenomena interact and to which degree they influence each other, or we want to know how we can enhance methods that allow us to study phenomena of interest. While questions that ask for the solution of big problems often trigger peoples' initial interest in science, scientists themselves typically work with much smaller problems. While the solution of small problems may seem boring to the public, they are crucial for the field to advance and they can lay the foundation for major breakthroughs at later stages.

Working on the typically small problems that individual subfields offer, all scientists run the risk of loosing track of their discipline's broader challenges. Working fortunately in our ivory tower that shields us from the outside world, it is often through confrontation with laypeople or scientists from other fields that we are made aware of the greater challenges of our own discipline.

In the field of linguistics, specifically in the field of historical linguistics, one of the questions that linguists have stopped asking is how language – that is, the specific communicative faculty of which many think is unique to humans – originated for the first time. Non-linguists are often very surprised that asking for the origin of language has been a taboo question in the field of historical linguistics for a very long time. Already in the 19th century, the question had been officially banned from the agenda of most linguistic endeavors. The *statuts* of the *Société de Linguistique de Paris* from 1866, for example, state that '[the] Société does not accept any contributions either on the origin of the language faculty, or the creation of a universal language.' (Société de Linguistique de Paris, 1871, III).[1] In my impression, this situation has not changed significantly since then. While — as reviewers of the first version of this essay have pointed out — it is true that the interest of scientists who discuss how language originated the first time has increased (which is also reflected in journals dedicated to the topic and in conferences devoted to it), the discipline of historical linguistics in particular and mainstream linguistics in general still largely avoid the question, reflecting the Société's ban until today. One of the first things I learned as a student of Indo-European and Comparative Linguistics at the Freie Universität Berlin in 2003 was that serious historical linguists would never ask about the origin or evolution of language. Since then, I have often experienced dismissive reactions, specifically in the German university context, when talking about *language evolution* instead of *language history*. Major textbooks in historical linguistics exclude the question of how language originated completely, and some even mention this explicitly, as reviewer Michael Pleyer has pointed out (Campbell, 1999[1998]: 1f)

Of course, there are good reasons to avoid asking how language originated. When sifting through the literature that has been published on this matter, one finds a diverse collection ranging from serious attempts to summarize what we can know and what we can't know, up to the weirdest speculations. There are not many similar fields in science where it is so difficult to draw a line between genius and madness. On the one hand, one finds researchers who meticulously assemble tiniest pieces of evidence in the search for a clear picture in the dark glasses. On the other hand, one finds scholars so obsessed with a single idea that they have become blind to counterarguments. Specifically for outsiders from the field, for those not trained in historical linguistics and evolutionary anthropology, it is often very difficult to tell if a theory on the origin of language should be treated as a serious or a senseless idea. Next to Johann Gottfried Herder (1744–1802) who imagined that human beings would have run through the woods imitating the sounds of the objects and phenomena surrounding them (Herder, 1778, see also Table 1), scholars discuss whether Neanderthals could speak or not (Dediu & Levinson, 2013), they have debated for some time whether a single gene was responsible for our language faculty (Atkinson *et al.*, 2018), they propose that research on aphasia might shed light on the structure of early language (Code, 2021), and they speculate about the nature of the 'language of proto-sapiens' in the context of *yin* and *yang* (Papakitsos & Kenanidis, 2018, see also Table 1). There is no doubt: the question invites bold speculation, and when bold speculation becomes careless, it can easily damage the reputation of entire scientific fields. For outsiders who do not know the scientific community that tries to investigate the question of the origin of language carefully, it is often not easy to distinguish between serious and highly speculative attempts. At the same time, however, it feels strange that linguists would deliberately decide to ignore the investigation of a problem that some might consider the most fascinating the field has to offer. While following inconclusive debates and trying to not get angry or amused by weird speculations regarding the "big questions" of one's discipline, one may easily forget that there *are* valid and important problems which one may have forgotten to think about. An example for such a problem that is

[1] My translation, original text: La Société n'adment aucune communication concernant, soit l'origine du langage, soit la création d'une langue universelle.

**Table 1. Collection of quotes on the origin of language.**

| |
|---|
| Sound imitation at the origin of language. |
| *Take for example the sheep. As an image [...] – how much, how difficult to discern! All characteristics are intertwined, next to each other, all still unspeakable! Who can talk shape? Who can sound colors? [...] Who can say what he feels with his hands? But listen, the sheep bleats! [...] "Ha!" says the apprentice [...], "now I will recognize you – you bleat!" The turtledove coos! The dog barks! There are three words, because he was looking for three clear ideas, the latter go into his logic, the former into his dictionary! [...] The soul grasped [for it] – there it has a sounding word!* (Herder, 1778, 76).[2] |
| Yin, yang, and the language of proto-sapiens. |
| *The Proto-Sapiens grammar was so simple that the sporadic references in previous paragraphs have essentially described it. The prime importance of sound symbolism for the people of nature should be noted again before we further detail that the vowel E was felt as indicating the "yin" element, passivity, femininity etc., while "O" indicated the yang element, activeness, masculinity etc.; "A" was neutral or spiritual, indicating things conceived by the mind and emotions rather than with the physical senses.* (Papakitsos & Kenanidis, 2018, 8) |

routinely ignored in historical linguistics is the problem of the "size" of a language. While historical linguists hardly ever ask how many words a language has, or how many sentences one can speak until one would repeat oneself, non-linguists have often asked me these questions. I still remember quite vividly how annoyed I always felt – and at times even reacted – when colleagues from biology asked me if English was bigger than German or French. I tried to explain that languages are open systems that cannot be measured by counting the words in a corpus and that if at all, one would have to count the words in the head of individuals, but that this would not be possible from a practical viewpoint, and that as a result of these complications, linguists preferred to ignore the question completely and look into other problems instead. But with time, I learned to put my linguistic pride aside and began to understand that a solution to the question would have several important implications to other questions that are of vital interest to my research.

In evolutionary biology, for example, scholars have argued that the number of genes that were horizontally transferred among species largely exceeds the number of genes that were vertically inherited (Dagan & Martin, 2006). Horizontal transfer is quite abundant in language history as well, and we usually base our phylogenetic studies on very small collections of basic words often not even exceeding 200 items per language (Greenhill *et al.*, 2023; Sagart *et al.*, 2019). It would therefore be interesting to have a rough estimate of how many words of a language survive over time, but this would require (at the very least) a rough estimate of the words that constitute a language.

In historical linguistics, Starostin (1989) has proposed that every language has about 1000 lexical roots from which most of the words in the language are formed. Up to today, no attempt has been made to count the number of lexical roots, even for well-documented languages like German or Chinese. It would be very interesting to see if Starostin's estimate holds cross-linguistically, and how much variation we should expect when comparing the languages of the world.

Given that languages may differ quite substantially regarding the way in which they build new words from existing ones, it would also be very interesting to see to which degree languages differ regarding the "productivity" of their words to form families (List, 2023), and to which degree the production of new vocabulary is triggered by external events, such as, for example, pandemics or wars.

Lastly, measuring the number of words that different speakers of the same language know might even help us to investigate to which degree individual language faculties vary among humans. The question to which degree speakers of the same language differ regarding their competence is another question that is – unfortunately – rarely asked, although the notion of competence plays a crucial role in some linguistic fields.[3]

While the question of the "size of a language" has been mostly ignored in the context of historical linguistics (Deutscher's assessment that literacy plays a major role in vocabulary size is a rare exception that does, however, not attempt to demonstrate the claim empirically, see Deutscher, 2010: 111), vocabulary size or vocubulary breadth[4] have been routinely

[2] My translation, original text: Da ist z. E. das Schaf. Als Bild [schwebet es dem Auge mit allen Gegenständen, Bildern und Farben auf einer großen Naturtafel vor] wie viel, wie mühsam zu unterscheiden! Alle Merkmale sind fein verflochten, nebeneinander alle noch unaussprechlich! Wer kann Gestalt reden? Wer kann Farben tönen? [Er nimmt das Schaf unter seine tastende Hand das Gefühl ist sicherer und voller, aber so voll, so dunkel ineinander.] Wer kann, was er fühlt, sagen? Aber horch! das Schaf blöket! [Da reißt sich ein Merkmal von der Leinwand des Farbenbildes, worin so wenig zu unterscheiden war, von selbst los, ist tief und deutlich in die Seele gedrungen.] »Ha!« sagt der lernende ]Unmündige wie jener Blindgewesene Cheseldens,] »nun werde ich dich wiederkennen du blökst!« Die Turteltaube girrt! Der Hund bellet! Da sind drei Worte, weil er drei deutliche Ideen versuchte, diese in seine Logik, jene in sein Wörterbuch! [Vernunft und Sprache taten gemeinschaftlich einen furchtsamen Schritt, und die Natur kam ihnen auf halbem Wege entgegen durchs Gehör. Sie tönte das Merkmal nicht bloß vor, sondern tief in die Seele hinein! Es klang!] Die Seele haschte da hat sie ein tönendes Wort!]

[3] See Dąbrowska, 2020 for a rare example where scholars try to deal with differences in competence, investigating genitive marking in Polish, in the context of a research program that looks into individual differences in competence.

[4] The term *vocabulary breadth* is usually understood as "the number of words a learner knows in a foreign language" (Milton, 2010, 211), but it can probably also be used as a term denoting the number of words a native spaker knows.

investigated by scholars focusing on foreign language acquisition (Milton, 2009) and psycholinguistics (Brysbaert *et al.*, 2016; see specifically also Nation & Coxhead, 2021 on the problems of estimating vocabulary sizes). Unfortunately, the majority of these studies has concentrated exclusively on English. We know that languages may differ quite substantially regarding the structure of their word families and the techniques they use to create new words from existing ones (Milton, 2010, 226, see also List *et al.*, 2016b, 9f). As a result, the estimates on vocabulary size are of limited use to address the above-mentioned problems in historical linguistics, and it remains an open problem to measure and compare the size of the vocabularies of the worlds' languages.

As scientists, we cannot ask enough questions. Being used to address small problems as part of our scientific routines, however, we may forget to ask the "big questions" that are at the heart of our particular disciplines. By ignoring certain questions deliberately and limiting the scope of questions we allow ourselves to ask in our work we may easily lose the chance to enrich our studies and open new horizons. Specifically for young and rapidly growing subdisciplines such as the field of computational historical linguistics, it can be very helpful to identify particular problems and challenges that should be addressed in future work. When talking about computational historical linguistics in this context, I refer to those attempts that try to formalize and automate the classical approaches for historical language comparison that have been developed in the traditional historical linguistics (often referred to as the "comparative method"). In the following, I will reflect on challenges that have been discussed in the context of comparative linguistics an contrast them with those challenges that I have identified for my own work. In doing so, I hope to show that an active discussion about open problems can be a useful guiding principle not only for an entire research field, but also for individual researchers.

## Rational, general and historical problems

Major problems and challenges for the field of comparative linguistics have been discussed in the past on different occasions. Weinreich *et al.* (1968, 183–187) identify five general problems with respect to the phenomenon of language change, which they call

(1) the *constraints problem*, dealing with the question of which changes are possible in language change and which conditions could constrain them,

(2) the *transition problem*, dealing with the question of how and where changes are instantiated in concrete during language change,

(3) the *embedding problem*, dealing with the systemic aspects conditioning language change,

(4) the *evaluation problem*, dealing with the question to which degree communities of language users are aware of the changes that are happening, and

(5) the *actuation problem*, dealing with the question of how changes are triggered.

Roberts and Sneller (2020) discuss these in the context of the "four questions" for evolutionary sciences proposed by Tinbergen (1963), pointing to potential empirical implications and programs for future research. Tinbergen's questions themselves have been originally stated for the field of biological evolution (Bateson & Laland, 2013), although they were later adopted by researchers studying cultural evolution (Roberts & Sneller, 2020, 194f). They are nowadays usually presented in a more systematic fashion than the problems by Weinreich *et al.* (1968), distinguishing two major perspectives, *dynamic* (diachronic) and *static* (synchronic), and two major kinds of questions, *how-* and *why*-questions, the former referring to individuals and the latter to species (Tinbergen, 1963). This allows to look at particular problems (e.g., the evolution of a specific trait) from four perspectives, namely

(1) the perspective of the *ontogeny*, focusing on how the trait evolves in individuals,

(2) the perspective of the *mechanism*, focusing on how the trait is structured synchronically in an individual,

(3) the perspective of the *phylogeny*, focusing on how the trait evolves inside a given species, and

(4) the perspective of the *function*, focusing on the role adaptive role the trait plays for a given species.

Similar to Roberts and Sneller (2020), I do not find it very helpful to try to classify the five problems by Weinreich *et al.* (1968) according to the schema proposed by Tinbergen (1963). Despite the apparent systematicity of the latter, I find the schema hard to apply to concrete problems.

As yet another example for an attempt to systematize linguistic endeavor by stating problems, Eugenio Coseriu (1921–2002, see Coseriu, 1973, 65f) suggested distinguishing three basic problems of language change, namely

(a) the rational problem of change ("problema racional del cambio"),

(b) the general problem of change events ("problema general de los cambios"), and

(c) the historical problem of a given change ("problema histórico de tal cambio determinado").

These problems can again be represented by certain questions, as indicated by Coseriu himself. The rational problem asks why languages change after all ("¿por que cambian las lenguas?"). This question does not find a counterpart in the list of problems proposed by Weinreich *et al.* (1968), where language change has been taken for granted, and the goal is to investigate and describe the phenomenon. As Coseriu emphasizes himself, the problem is of a chiefly *theoretical* nature and cannot be resolved by identifying all causes for particular changes that can be observed for particular languages (ibid. 66f), but rather addresses the deeper question of why mutability is one of the fundamental characteristics of language (68f). Coseriu himself sees the reason for the mutable character of language as a result of the fact that languages is

constantly recreated, not only when being learned by speakers, but also when being applied by them (69f).

The second problem of Coseriu is similar to the *actuation problem* by Weinreich *et al.* (1968), addressing the question in which conditions certain changes occur ("¿en qué condiciones suelen occurrir cambios en las lenguas?"). In Coseriu's view, this problem is a problem of *general linguistics* in the sense that general linguistics deals with linguistic phenomena independently of particular languages. Particular changes in particular languages, finally, are addressed by the third problem, which Coseriu calls *historical*, emphasizing the individual character of investigating particular changes in particular languages.

Coseriu's strict distinction between *general problems* in linguistics on the one hand and *historical problems* on the other hand finds a very close counterpart in the distinction between *p(articular language)-linguistics* and *g(eneral) linguistics* by Haspelmath (2020, 5), where a distinction between the investigation of language as a general communication system and the investigation of individual languages is made (see also Haspelmath, 2019).

Judging from numerous discussions with colleagues, distinguishing questions applying to individual languages and language families from questions applying to language in general (as a system of human communication) constitutes a much more important systematization of problems than the attempts by Weinreich *et al.* and Tinbergen. The failure to distinguish questions pertaining to particular languages and questions pertaining to language in general has led to many misunderstandings in the field of comparative linguistics. In my own research, it has happened quite a few times that scholars who reviewed my work were asking me to test new methods which I had designed to account for general problems against previously proposed methods that could only solve problems for particular languages (at times even relying on particular orthographies, while my methods would use phonetic transcriptions). It also happens a lot — specifically when presenting new methods to computational linguists who do not know the particular problems of historical linguistics very well — that newly proposed general methods that work in an unsupervised fashion (i.e., without requiring training data) are rejected with the justification that supervised methods that solve the problem for particular languages (using a substantial amount of training examples) exists.

## Hilbert problems

At the end of 2018, students from the Universidad de Buenos Aires asked me about the biggest challenges for computational historical linguistics. Inspired by this discussion, I decided to make a short list of tasks that I consider challenging, but of which I still think could be solved some time in the nearer or farther future.

The idea to make such a list of questions is not new. Mathematicians, for example, have their well-known *Hilbert Problems*, proposed by David Hilbert in 1900 (published in Hilbert, 1902). In linguistics, I first heard about them from Russell D. Gray, who himself was introduced to this by a talk of the linguist Martin Hilpert, who gave a talk on challenging questions for linguistics in 2014, called "Challenges for 21st century linguistics". Russell D. Gray since then has emphasized the importance to propose "Hilbert" questions for the fields of comparative linguistics and cultural evolution, and has also presented his own challenges in the past.

Due to my methodological background, the problems I identified and assembled were by no means big and in some sense also not necessarily extremely challenging (at least not at first sight). Instead, the problems I selected were problems I wanted to see solved at that time. While the solution of the problems would not directly advance our knowledge about language evolution and linguistic typology, I had the hope that it would help us to do so indirectly, by giving us the possibility to assemble more data and to carry out new analyses that would ultimately help us to search for answers on deeper questions in historical linguistics in specific and in comparative linguistics in general. Due to this goal of providing solutions for historical linguistics in general, I was interested in solutions that could applied to any language, as long as it is represented in some uniform way (in phonetic transcription rather than orthography). Open problems also exist in the context of particular languages, but my interest was in problems in general historical linguistics rather than historical linguistics of particular language families.

One further aspect of the problems that I selected was that I was convinced that they could *all* be solved by algorithms or workflows. Characterizing them as "small" refers to their very specific application range. I did not want to express that the problems I selected were not challenging. It also did not mean that I expected that they all could be solved in the nearer future, although, given that the work in the field of computational and computer-assisted language comparison, is steadily progressing, I had some confidence that at least some of these problems that I assembled by then would indeed be solvable within the next five years.

When writing down my ten open problems for computational historical linguistics, I announced them in a blog post for the blog *The Genealogical World of Phylogenetic Networks* (https://phylonetworks.blogspot.com/), edited by David Morrison, in January 2019 (List, 2019a), with the plan of discussing each of the problems in detail in monthly blog posts throughout the year. I managed to stick with this schedule and concluded the year with a final blog post in December 2019, in which I looked back at one year of discussing problems in my own research for which no solution could have been found by then (List, 2019d).

The 10 problems I came up with are listed in Table 2. I divided the problems into three different groups, which roughly correspond to three different categories I identified as being important for research in general, namely *modeling* (m), *inference*

**Table 2. 10 problems of computational diversity linguistics discussed in a series of blog posts in 2019.** Problems labeled as (i) are inference problems, problems labeled (m) refer to modeling problems, and problems labeled (a) are analysis problems.

| No. | Problem |
|---|---|
| 1 | automated morpheme segmentation (i) |
| 2 | automated borrowing detection (i) |
| 3 | automated sound law induction (i) |
| 4 | automated phonological reconstruction (i) |
| 5 | simulating lexical change (m) |
| 6 | simulating sound change (m) |
| 7 | statistical proof of language relatedness (m) |
| 8 | typology of semantic change (a) |
| 9 | typology of sound change (a) |
| 10 | typology of semantic promiscuity (a) |

(i), and *analysis* (a). This triad, inspired by Dehmer *et al.* (2011, XVII), follows the general idea that scientific research in the historical disciplines usually starts from some kind of idea we have about our research object (the *model* stage), and based on which we then apply methods to infer examples in our data which confirm our ideas (the *inference* stage). Having inferred enough examples, we can then *analyze* them qualitatively or quantitatively (the *analysis* stage) and use this information to update our model, as indicated in the schema in Figure 1.

As an example for this procedure, consider the problem of *cognate detection*, the detection of historically related – or homologous – words across languages. Here, initial surveys of words across different languages have long since confirmed that we can easily find words which are to some extent similar to each other with respect to their form and their meaning, and that the amount of similar words varies quite significantly from one pair of languages to another. As an example, consider word pairs, such as German *Zeh* "toe" and English *toe*, German *zwei* "two" and English *two*, or German Zeichen "sign" and English *token*. Starting from a model of *lexical change* that states that the lexicon of all spoken languages changes slowly over time, be it through the change of individual sounds or through the change of the meaning which a word expresses, we can derive a model of *language split* which assumes that the same language may split into two or more varieties when its speakers separate from each other and their language keeps modifying independently. Based on this model, we can then conclude that similar words observed across different language varieties have been inherited from the common, formerly unified, ancestral variety. In order to detect these *cognates*, one could now design methods that help to find more

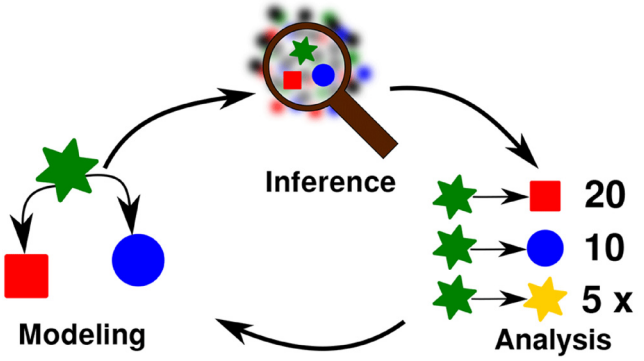


**Figure 1. Research workflow for the investigation of problems in computational historical linguistics.** The stage of modeling assumes certain relations between certain scientific objects. The stage of inference tries to find more examples for the relations proposed by a given model. The stage of analysis would then — for example — compare the frequency by which different processes can be accounted for and use this information to inform the model.

than the so far observed similar words in order to increase the data basis. In the case of German and English, one could design a method by which dictionaries in English and German are searched for words that start with *z-* in German and with *t-* in English and compare their semantics. Once this has been done, and more material has been identified, one can *analyze* the data and try to see if the analysis provides some hints on specific questions, such as, for example, the more detailed branching history of a given language family, or to which degree language contact has masked further evidence. More examples of German words containing a *z* with counterparts containing a *t* in English would, for example, show that there are more nuances to the correspondences (compare cases like German *heiß* "hot" vs. English *hot*, or German *hassen "hate" vs. English "hate"*), and that semantics can vary considerably (compare German *Zaun* "fence" vs. English *town*).

As one can see from the example, the term *modeling* is used in a rather loose sense, unlike its usage in phylogenetic approaches, where *modeling* is often treated as synonymous with *stochastic modeling*, referring to a transition matrix which shows how certain characters can turn into other characters during an evolutionary process (Nunn, 2011).

With respect to my ten open problems from 2019, the first four problems belong to the family of inference problems, since they all deal with tasks where something has to be inferred from the data, be it morphemes from words (Problem 1, List, 2019e), lexical borrowings from word lists (Problem 2, List, 2019g), sound laws from data on ancestral languages and their descendants (Problem 3, List, 2019f), or proto-forms from cognate sets (Problem 4, List, 2019h). Note that the inference of morphemes in wordlists is listed as a problem pertaining to the field of historical linguistics here, because the identification of the morphemes in a language is an important step for *internal reconstruction*.

Although the handling of morphemes and the identification of the principles underlying word formation in particular languages is often seen as a classical problem in synchronic linguistics, such analyses are typically diachronic in nature, or — to put it in other terms — it is not always easy to distinguish diachrony from synchrony when dealing with questions of morphology and word formation. Note also that not all linguists agree on the notion of the morpheme. From the perspective of historical linguistics, where one must try to control for allomorphic alternation, the notion of morpheme boundaries and morpheme segmentation is crucial, even if it is rarely discussed in classical handbooks introducing the traditional methodology of historical language comparison.

The next three problems in my list belong to the family of modeling problems, since they all require to understand the processes by which certain aspects of languages, such as their lexicon (Problem 5, List, 2019i) or their sound systems (Problem 6, List, 2019j) change over time. Proving language relatedness (i.e. that two or more languages have descended from the same ancestral language) statistically (Problem 7, List, 2019k) does not directly model any aspect of language evolution, but it requires a model of language relatedness that can then be tested against a random model in which languages are thought to be unrelated.[5]

The last three problems in my list all had "typology" in their title. They belong to the family of analysis problems, aiming to gain insights into phenomena of language change by comparing major processes, such as semantic change (Problem 8, List, 2019l) and sound change (Problem 9, List, 2019b). What was meant by "typology" in this context was a data-driven estimate of the overall cross-linguistic dynamics of these phenomena. Lacking consistent accounts on the general tendencies of these processes and phenomena when excluding areal and genetic factors, the task I thought of ways to come up with a consistent estimate on each of them. While semantic change and sound change are probably self-explaining in this context, the last problem – dealing with the question of what I called "semantic promiscuity" by then – deserves some explanation (Problem 10, List, 2019c). What I meant with this term was the degree to which certain words, due to their original meanings, are re-used or re-cycled in the human lexicon. While the term *promiscuity* has been used before in other contexts in linguistics (Schweikhard, 2018), the specific usage of promiscuity to denote what one could also call *semantic productivity* or *concept productivity*, was first proposed in List *et al.* (2016a), where biological and linguistic processes were consistently compared with each other, and semantic promiscuity was identified as a phenomenon similar to *domain promiscuity* in protein evolution in biology (Basu *et al.*, 2008), with an explicit analogy being identified between the processes of *word formation* in linguistics and *protein assembly* (Ahnert *et al.*, 2015) in biology (List *et al.*, 2016a, 5). As I'll discuss in more detail below, I now tend to call the phenomenon *lexical root productivity*, a term supposed to denote the propensity of certain words and morphemes to be used to form new words due to the meaning they express.

## New and old open problems

The series of blog posts was generally well received, and some posts triggered interesting discussions. However, my hope that at least a few of the problems might be considered as "solved" after half a decade turned out to have been a false one. When looking back at the list of ten problems now, almost five years after I first proposed them, I do not have the feeling that we are any closer in solving any of them. For this reason, the list of open problems that I consider important for my own field and my own work has remained almost unchanged, although more than five years have passed by now.

## Inference problems

As far as inference problems are concerned, no significant progress was made with respect to the tasks of automated morpheme segmentation (Problem 1) and automated sound law induction (Problem 3). Research on natural language processing has profited from new segmentation approaches that avoid to take the word as the basic unit of texts, proposing to segment texts in large corpora into units beyond the word. These methods, however, cannot be applied to the specific problem of morpheme segmentation outlined in my list of problems, where much fewer words in phonetic transcription constitute the basic unit of analysis.

Quite a few new methods have been proposed to address the detection of lexical borrowings (Problem 2). Among these are supervised approaches (which I had deliberately excluded, since I consider unsupervised approaches as more useful when it comes to inference problems) that made use of recurrent neural networks (Miller *et al.*, 2020), there are tree-based approaches that even try to identify the direction of borrowings (Neureiter *et al.*, 2022),[6] and there are numerous attempts to handle very specific cases of lexical borrowing, such as contact across language families (Hantgan *et al.*, 2022; List & Forkel, 2022) or contact induced by dominant languages (Kaiping & Klamer, 2022; Miller & List, 2023).

While all of these methods may contribute to the detection of borrowings in particular cases, all of them suffer from

[5] While there are numerous attempts in the literature to come up with a convincing statistical model to prove genetic relationship (Baxter & Manaster Ramer, 2000; Blevins & Sproat, 2021; Ceolin, 2019; Kassian *et al.*, 2015; Kessler, 2001; Mortarino, 2009; Ringe, 1992), none of the attempts which have been proposed so far deals with lexical comparisons in all their complexity. Either, scholars only compare initial consonants with each other (Kessler, 2001; Ringe, 1992), or they resort to sound classes (Baxter & Manaster Ramer, 2000; Kassian *et al.*, 2015), and even if scholars compute random models for whole alignments of potentially related words (List, 2014), they have the problem of not accounting for the factor of closeness due to borrowing.

[6] I write "try", since the study does not provide a detailed error analysis and many directions seem to be problematic).

the problem that they need very specific conditions to work. For supervised approaches, we need labeled training data, for tree-based approaches to borrowing detection, we need phylogenetic trees, for borrowing detection across language families, we need languages to belong to different families, and for the detection of borrowing from dominant languages, we need to deal with languages that are spoken in a region where dominant languages occur.

It is well possible, that the current pocket knife solution that employs various forms of evidence to find borrowings in very specific contexts is the only feasible way to handle the problem of language contact in computational historical linguistics. Even classical approaches to historical language comparison do not use one unified approach to identify borrowed traits, but rather hope to accumulate enough evidence until borrowing is the only convincing solution to explain the data at hand. But even if we accept that we need to embrace arguments based on "cumulative evidence" (Berg, 1998) or "consilience" (Whewell, 1847; Wilson, 1998) in order to solve the problem of borrowing detection (see also List, 2019n, 11), we are still quite far away from being able to handle all the evidence with automated approaches which goes into the argumentation of classical qualitative approaches to borrowing detection. From today's perspective, I would adjust the problem of borrowing detection in order to make it more specific. Here, a problem I would really love to be solved would be the detection of the *layers of contact* across a group of languages (Lee & Sagart, 2008). Contact layers have been discussed for a long time in the literature on language contact. The idea behind contact layers is that the individual traits of a language can be *stratified* and assigned to different groups that would point to different phases in which these traits were borrowed through specific contact events. Developing a method that would be able to group borrowings into different strata that could then be identified with specific contact events in time would be extremely beneficial for the discipline of historical linguistics. The problem is, however, also very challenging, since it is not clear whether contact layers *can* be identified in all cases (evidence might just have been lost), and on what kind of evidence one should base the detection of contact layers. This makes contact layer detection a truly hard problem, although I would not consider it impossible to solve.

Rather huge progress – at least when looking back specifically at the last couple of years – has been made with respect to supervised phonological reconstruction. Here the task is different from the problem I had originally stated. Instead of inferring the proto-language from a sufficiently large number of aligned cognate sets, the method is given aligned cognate sets (or simply cognate sets) along with a certain number of already reconstructed proto-forms that can be used to train a machine learning model in a first instance. In a second instance the model can then be used to infer proto-forms for data that has not been seen previously.

While scholars had been working on this problem before, a first impressive demonstration of the capability of machine learning methods was done by Meloni *et al.* (2021) for a test set in which Latin words needed to be reconstructed from words in several Romance languages. The authors used recurrent neural networks to learn to predict Latin word forms from word forms in the Romance languages. What was remarkable about their approach was the high accuracy reported. Using the edit distance (Levenshtein, 1966) to compare the proposed Latin word with the attested Latin word, their best parameter settings reported scores in which the proposed word would on average diverge less than one character from the attested word form. Kim *et al.* (2023) repeated the experiments by Meloni *et al.* (2021) but used Transformers (Vaswani *et al.*, 2017), the famous architecture for neural networks, that has revolutionized artificial intelligence applications in many areas, but is best known in the context of large stochastic language models that are used to run chat bots. Not unexpectedly, the Transformer models further outperform the recurrent neural network architecture employed by Meloni *et al.* (2021).

The problem of supervised phonological reconstruction can be stated in the broader context of *reflex prediction*. Reflex prediction refers to the task of predicting how a word sounds when knowing the pronunciation of historically related words in related languages. For example, observing words like German *Zoll* and Swedish *tull* "customs", I might predict that the English should have a corresponding words *toll* with a meaning similar (but not necessarily identical) to the meaning of the words in Swedish and German. At times, this may even work with language pairs, and as learners of languages closely related to languages we know intimately, we may even intuitively predict how certain words in the foreign language might sound, based on our knowledge of similar words in the languages we know.

We introduced the task in a preregistered study in which we first predicted word forms in languages that had so far been insufficiently studied and then checked the predictions against word forms verified in field work carried out after the predictions had been registered (Bodt & List, 2022). For the prediction, we used a supervised approach that goes back to a new method for the detection of regular sound correspondence patterns which I had introduced in 2019m. Having refined this approach in a later study, testing it on a larger collection of datasets from different language families (List *et al.*, 2022b), we used it as a baseline for a *shared task on reflex prediction* where we invited scholars with a background in machine learning and historical linguistics to design their own approaches for the task of reflex prediction (List *et al.*, 2022c). For this shared task, we used an even larger number of datasets from the Lexibank repository (List *et al.*, 2022a) and introduced a computer-assisted pipeline to make sure that all participants would have access to the data in the same form. The best systems were presented by Kirov *et al.* (2022) who used sophisticated neural network architectures and data processing pipelines to augment the very sparse input data. Interestingly, their best model was not based on transformers, but on convolutional network architectures originally designed for the task of restoring images (Liu *et al.*, 2018). While the recent success of neural network approaches in the task of phonological reconstruction and reflex prediction is definitely impressive, the

approaches have the drawback of not telling us how they arrive at their decisions. The consequence of their blackbox character is that we cannot use them to learn about the tasks they solve.

With respect to the originally proposed problem of *unsupervised phonological reconstruction*, not much has happened in the meantime. In 2013, Bouchard-Côté *et al.* (2013) showed that unsupervised automated phonological reconstruction is possible for Austronesian languages, using a complex framework in which stochastic transducers were applied to model the evolution of individual words across known reference trees. Since then, only Jäger (2019) has presented an alternative approach to the problem, in which methods for ancestral state reconstruction (Nunn, 2011, 63–89) were applied to a test set of Romance languages. The results were apparently disappointing, but closer inspection easily shows that the problem is less the method itself (although a closer analysis would be needed), but even more the quality and the nature of the original data, which treats Latin as the ancestor of the Romance languages, although it has been known for a long time that a reconstruction of Romance languages cannot reveal Latin entirely, since many distinctions have been lost across all Romance languages (Hall, 1950).

In summary, we can conclude that the problem of automated phonological reconstruction remains a difficult problem for which no satisfying solutions exist so far. Even the very good results reported for the supervised reconstruction on Romance languages should be taken with considerable care, since the original dataset by Ciobanu & Dinu (2014) is far too large to provide a realistic test case that would be applicable to other language families. Not only does it seem impossible to find a comparable number of cognate sets for other language families of the same time depth as that of the Romance language family, I would also suspect that the majority of supposed cognates in the data do not qualify as true cognates (referring to etymologically related words, see Trask, 2000, 63) but rather reflect late borrowings from Latin into the individual Romance languages. Since borrowings show very different rules of transformation, which are – at least in the case of borrowings from Latin into Romance languages – often much simpler than the complex sound change processes that can be observed in the languages of the world, it would be important to test the approach by Meloni *et al.* (2021) and the follow-up approach by Kim *et al.* (2023) on the much sparser dataset that we proposed for our shared task on reflex prediction in order to understand their real potential.

## Modeling problems

With respect to the modeling problems, there was – as far as I can judge – no substantial progress in the simulation of lexical change and sound change, but there were some interesting studies dealing with the problem of finding a statistical proof for language relatedness.

Ceolin (2019) tested the methods by Baxter & Manaster Ramer (2000), Kessler & Lehtonen (2006) and Kessler (2001) on a dataset consisting of six different language varieties, including Turkish, Mongolian, and Manchu, three languages that some scholars assume to be genetically related, showing that no conclusive results could be obtained in favor of the highly disputed Altaic language family (Georg, 2017).

Kassian *et al.* (2021) employed the test reported by Turchin *et al.* (2010), originally inspired by Dolgopolsky (1964) to another dataste of languages from the disputed Altaic family, with the difference that they used reconstructed proto-languages. Their test also failed to provide conclusive evidence for the whole language family, although they argue that rather clear support can be found for a deeper relationship of the families tested by Ceolin (2019). While both Ceolin (2019) and Kassian *et al.* (2021) applied methods proposed before to newly compiled datasets, Blevins & Sproat (2021) designed a new workflow for to test for supposed deep language relations in order to find evidence for the hypothesis that the language isolate Basque is related to Indo-European. In contrast to previous studies, Blevins & Sproat (2021) test their approach on a rather large sample of languages where the language relations are known, showing that their approach is rather conservative, showing a tendency to reject grouping two languages within the same family in case of sparse evidence. Whether this is enough to prove the case of Basque and Indo-European, however, remains to be seen, since the data used in the test and the data used to test the potential relationship between Basque and Indo-European were not identical in design. This makes it more difficult to interpret the results of the test.

Given that Ceolin (2019) and Kassian *et al.* (2021) report contradicting results, both using methods that are supposed to deliver clear mathematical proofs, and given the difficulty to compare the evaluation study reported by Blevins & Sproat (2021) with the tests they conducted, I think it is safe to argue that the last word on the problem of finding a convincing statistical proof for language relatedness has not yet been spoken.

## Analysis problems

No progress that I would be aware of has been made with respect to the establishment of first typologies for semantic change (Problem 8) and sound change (Problem 9). This confirms my original suspicion that both problems are indeed rather tough ones, requiring large amounts of annotated data – along with ideas of *how* the data should be annotated – that we simply do not have at the moment. While individual inquiries into semantic change based on corpus data of large well-documented language such as English have enjoyed a considerable popularity over the last years (e.g., Xu *et al.*, 2017), the problem I identified relates to cross-linguistic tendencies of semantic change, which cannot be observed by corpus studies for a few languages with a long documentation history. As a result, these corpus studies — as interesting as they may be — do not contribute to the solution of the problem I identified. Regarding the last problem on my list, the problem of establishing a typology of what I called *semantic promiscuity* by then,

there has been no progress regarding the methodology of studying the phenomenon, but I think that there was at least some progress in explaining and defining the problem itself more properly. When I initially stated the problem, I was not aware of the rather large body of literature devoted to the topic of *lexical motivation*, referring to the process by which new words are created from existing ones (Koch, 2001; Koch & Marzo, 2007; Urban, 2016). Most of the words in the lexicon of human languages are composed from other words and individual word histories can be very complex, as can be seen from the illustration in Figure 2, where I have described, how the term *Ellenbogengesellschaft* in German ("dog-eat-dog society", lit. "elbow society") derives from individual words and suffixes. For a long time, I have been trying to find a way to investigate to which degree the meaning of individual words contributes to their reuse.

Geisler (2018) makes the rather strong claim that word reuse results from bodily experiences made in early life, such as the act of "falling" or "standing". This could explain why words build from the verbal roots meaning "to fall" and "to stand" are so frequently met in German. An alternative possibility would be that word reuse reflects actual trends that could change over time (Alinei, 2001). This could explain, for example, the recent increase in the use of metaphors from psychology, popularly also known as "therapy speech" (Prendergast, 2022). Most likely, none of the two extreme positions is absolutely true. Instead, it is quite likely that both actual trends and important (potentially bodily) experiences contribute to the reuse of words in the world's languages. But a detailed investigation of the semantics underlying these processes by which new words are formed from existing ones has so far not been carried out yet.

When I first stated my problem of establishing a typology of semantic promiscuity, I did not know that quite some detailed work on the processes of *lexical motivation* – albeit mostly qualitative in nature – had already been carried out and that some authors had proposed concepts quite similar to what I had in mind when proposing the term *semantic promiscuity*. Blank (1997, 21), for example, describes the processes of

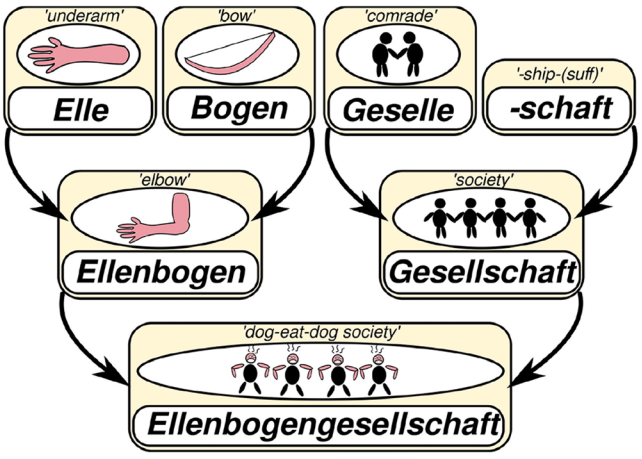


**Figure 2. The lexical motivation of the term "dog-eat-dog society" in German.**

*attraction* and *expansion*. Attraction refers to cases where a given concept "attracts" different words to express it. In theory, we might be able to measure the *attractivity* of concepts, that is, their propensity to be expressed by multiple words. Expansion refers to cases where a word receives new meanings. If one agrees that the *expansivity* of words typically depends on the meaning they express originally, one could take this idea one step further and measure the *expansivity* of concepts and compare it across languages. Taking it one additional step further, one could then ask not only which concepts are good at triggering the extension of a word's meaning, but also which concepts are good at triggering the reuse of a word in word formation processes, which is what I meant to denote with the term "semantic promiscuity".

In a recent study in which I proposed new methods for the automated inference of words that share certain parts resulting from word formation processes (called *partial colexifications* in that study, see List, 2023), I found that there is a tendency for words expressing certain concepts to be reused much more frequently in other words than words expressing different concepts. I also found a tendency for certain concepts to be expressed by words that are composed rather than being expressed by single morphemes. Since my analyses are based on a very rough approach that has not yet been tested any further so far, they should be taken with certain care. Given the confusion that the term "semantic promiscuity" has created in discussions with colleagues as well as in discussions following my original blog post (List, 2019c), I decided – inspired by a comment of Alexandre François – to use the term "lexical root productivity" from now on, in order to refer to the reuse potential of words in the lexicon – resulting from their meaning.

Regarding the problem of establishing a typology of lexical root productivity, I would no longer consider this as the most important problem for the field of lexical typology. Instead, I think, that one could state the problem in broader terms as the problem of establishing a typology of processes of *lexical motivation* that would allow us to investigate both how words are reused across the languages in the world (*form-based perspective*, also called *semasiological perspective*), and how concepts are expressed with the help of reusing lexical material (*concept-based perspective*, also called *onomasiological perspective*). Despite earlier attempts to solve certain aspects of this problem (Urban, 2012), a real typology of lexical motivation has not yet been established, and the problem can therefore be considered as an unsolved one.

## Outlook

In the end of 2018, I identified ten unsolved problems in computational historical linguistics that I considered as important and solvable at the same time. In my opinion, the importance of solving these problems has not changed since then. While quite some progress has been made in the past five years, most of the problems are still not solved, and it is not clear if and when we will find solutions for them. In order to be able to compare where I see the field of computational historical linguistics now, five years later, I have created a revised list of open problems, which is shown in Table 3. In two cases (Problem 3, contact layer detection, and

**Table 3. Updated list of open problems in computational historical linguistics.**

| No. | Problem (2019) | Problem (2023) | Progress |
|---|---|---|---|
| 1 | automated morpheme segmentation | | - |
| 2 | automated sound law induction | | - |
| 3 | automated borrowing detection | contact layer detection | + |
| 4 | automated phonological reconstruction | | + |
| 5 | simulation of lexical change | | - |
| 6 | simulation of sound change | | - |
| 7 | statistical proof of language relatedness | | (+) |
| 8 | typology of semantic change | | - |
| 9 | typology of sound change | | - |
| 10 | typology of semantic promiscuity | typology of lexical motivation | (+) |

Problem 10, typology of lexical motivation), I have decided to shift the focus and therefore modified the title of the problem. In two out of ten problems (Problem 3, automated borrowing detection, and Problem 4, automated phonological reconstruction), I would argue that some substantial progress has been made in the field, although I do not consider any of the problems as "solved" as of today.

Given that my optimism from 2018, when I assumed that most problems could be solved in five to ten years' time, has not turned out to be very reliable, I would now refrain from making any further assumptions on whether the ten problems outlined in Table 3 are solvable or not. If, however, in five years from now, the progress of the field of historical computational linguistics has been at a similar rate as it has been in the past five years, I assume that we will see some substantial progress, even if none of the problems can be solved completely.

## Data and software availability

### Source data

No data are associated with this study.

## Software availability

No software are required for this study.

### Acknowledgments

I would like to thank Nathan W. Hill for ongoing interesting discussions on open problems in historical linguistics, and Russell D. Gray for introducing me to the fascinating world of open problems in science.

## References

Ahnert SE, Marsh JA, Hernández H, *et al.*: **Principles of assembly reveal a periodic table of protein complexes.** *Science.* 2015; **350**(6266): aaa2245.
**PubMed Abstract** | **Publisher Full Text**

Alinei M: **Aspects of a theory of lexical motivation (iconymicity).** *Versus. Quaderni di studi semiotici.* 2001; **88/89**(1): 89–97.

Atkinson EG, Audesse AJ, Palacios JA, *et al.*: **No evidence for recent selection at *FOXP2* among diverse human populations.** *Cell.* 2018; **174**(6): 1424–1435. e15.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Basu MK, Carmel L, Rogozin IB, *et al.*: **Evolution of protein domain promiscuity in eukaryotes.** *Genome Res.* 2008; **18**(3): 449–461.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Bateson P, Laland KN: **Tinbergen's four questions: an appreciation and an update.** *Trends Ecol Evol.* 2013; **28**(12): 712–718.
**PubMed Abstract** | **Publisher Full Text**

Baxter WH, Manaster Ramer A: **Beyond lumping and splitting. Probabilistic issues in historical linguistics.** In: Colin Renfrew, April McMahon, and Larry Trask, editors, *Time Depth in Historical Linguistics.* McDonald Institute for Archaeological Research, Cambridge, 2000; 167–188.
**Reference Source**

Berg T: **Linguistic structure and change: an explanation from language processing.** Clarendon Press, Gloucestershire, 1998.
**Reference Source**

Blank A: **Prinzipien des lexikalischen bedeutungswandels am beispiel der romanischen sprachen.** Niemeyer, Tübingen, 1997.
**Reference Source**

Blevins J, Sproat R: **Statistical evidence for the Proto-Indo-European-Euskarian hypothesis. a word-list approach integrating phonotactics.** *Diachronica.* 2021; **38**(4): 506–564.
**Publisher Full Text**

Bodt TA, List JM: **Reflex prediction. a case study of western kho-bwa.** *Diachronica.* 2022; **39**(1): 1–38.
**Publisher Full Text**

Bouchard-Côté A, Hall D, Griffiths TL, *et al.*: **Automated reconstruction of ancient languages using probabilistic models of sound change.** *Proc Natl Acad Sci U S A.* 2013; **110**(11): 4224–4229.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Brysbaert M, Stevens M, Mandera P, *et al.*: **How many words do we know? practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age.** *Front Psychol.* 2016; **7**: 1116.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Campbell L: **Historical linguistics. An introduction**. MIT Press: Cambridge. 2nd edition, 1999[1998].
**Reference Source**

Ceolin A: **Significance testing of the altaic family.** *Diachronica.* 2019; **36**(3): 299–336.
**Publisher Full Text**

Ciobanu AM, Dinu LP: **Automatic detection of cognates using orthographic alignment.** In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers).* 2014; **2**: 99–105.
**Publisher Full Text**

Code C: **The prehistory of speech and language is revealed in brain damage.** *Philos Trans R Soc Lond B Biol Sci.* 2021; **376**(1824): 20200191.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Coseriu E: **Sincronía, diacronia e historia. El problema del cambio lingüístico.** Biblioteca Románica Hispánica, Madrid, 1973.
**Reference Source**

Dagan T, Martin W: **The tree of one percent.** *Genome Biol.* 2006; **7**(10): 118.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Dąbrowska E: **Language as a phenomenon of the third kind.** *Cogn Linguist.* 2020; **31**(2): 213–229.
**Publisher Full Text**

Dediu D, Levinson SC: **On the antiquity of language: the reinterpretation of neandertal linguistic capacities and its consequences.** *Front Psychol.* 2013; **4**: 397.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Dehmer M, Emmert-Streib F, Graber A, *et al.*: **Applied statistics for network biology: methods in systems biology.** Wiley-Blackwell, Weinheim, 2011.
**Publisher Full Text**

Deutscher G: **Through the language glass. Why the world looks different in other languages**. Metropolitan Books: New York, 2010.
**Reference Source**

Dolgopolsky AB: **Gipoteza drevnejego rodstva jazykovych semej Severnoj Evrazii s verojatnostej točky zrenija.** *Voprosy Jazykoznanija.* 1964; **2**: 53–63.

Geisler H: **Sind unsere wörter von sinnen? Überlegungen zu den sensomotorischen grundlagen der begriffsbildung.** In: Kerstin Kazzazi, Karin Luttermann, Sabine Wahl, and Thomas A. Fritz, editors, *Worte über Wörter. Festschrift zu Ehren von Elke Ronneberger-Sibold.* Stauffenburg, Tübingen, 2018; 131–142.

Georg S: **The role of paradigmatic morphology in historical, areal and genealogical linguistics.** *Journal of Language Contact.* 2017; **10**: 353–381.
**Publisher Full Text**

Greenhill SJ, Haynie HJ, Ross RM, *et al.*: **A recent northern origin for the Uto-Aztecan family.** *Language.* 2023; **99**(1): 81–107.
**Publisher Full Text**

Hall RA: **The reconstruction of proto-romance.** *Language.* 1950; **26**(1): 6–27.
**Publisher Full Text**

Hantgan A, Babiker H, List JM: **First steps towards the detection of contact layers in Bangime: a multi-disciplinary, computer-assisted approach [version 2; peer review: 2 approved].** *Open Res Eur.* 2022; **2**: 10.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Haspelmath M: **Confusing p-linguistics and g-linguistics: philosopher Ludlow on "framework-free theory".** *Diversity Linguistics Comment.* 2019; **7**(6).
**Reference Source**

Haspelmath M: **Human linguisticality and the building blocks of languages.** *Front Psychol.* 2020; **10**: 3056.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Herder JG: **Abhandlung über den Ursprung der Sprache, welche den von der königl. Academie der Wissenschaften für das Jahr 1770 gesetzten Preis erhalten hat.** Christian Friedrich VoSS, Berlin, 1778.

Hilbert D: **Mathematical problems.** *B New York Math So.* 1902; **8**(1): 437–479.
**Publisher Full Text**

Jäger G: **Computational historical linguistics.** *Theor Linguist.* 2019; **45**(3–4): 151–182.
**Publisher Full Text**

Kaiping GA, Klamer M: **The dialect chain of the Timor-Alor-Pantar language family a new analysis using systematic Bayesian phylogenetics.** *Lang Dyn Change.* 2022; **2022**(2): 1–53.
**Publisher Full Text**

Kassian AS, Starostin G, Egorov LM, *et al.*: **Permutation test applied to lexical reconstructions partially supports the altaic linguistic macrofamily.** *Evol Hum Sci.* 2021; **3**: e32.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Kassian A, Zhivlov M, Starostin GS: **Proto-Indo-European-Uralic comparison from the probabilistic point of view.** *J Indo-Eur Stud.* 2015; **43**(3–4): 301–347.
**Reference Source**

Kessler B: **The significance of word lists.** CSLI Publications, Stanford, 2001.
**Reference Source**

Kessler B, Lehtonen A: **Multilateral comparison and significance testing of the Indo-Uralic question.** In: Peter Forster and Colin Renfrew, editors, *Phylogenetic Methods and the Prehistory of Languages.* McDonald Institute monographs,Cambridge UK , Oxford UK , Oakville CT USA,, McDonald Institute for Archaeological Research; Distributed by Orbow Books, 2006; 33–43.
**Reference Source**

Kim YM, Chang K, Cui C, *et al.*: **Transformed protoform reconstruction.** In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers).* Toronto, Canada, Association for Computational Linguistics, 2023; **2**: 24–38.
**Publisher Full Text**

Kirov C, Sproat R, Gutkin A: **Mockingbird at the SIGTYP 2022 shared task: two types of models for the prediction of cognate reflexes.** In: *The Fourth Workshop on Computational Typology and Multilingual NLP.* Online, Association for Computational Linguistics, 2022; 70–79.
**Publisher Full Text**

Koch P: **Lexical typology from a cognitive and linguistic point of view.** In: Martin Haspelmath, Ekkehard König, Wulf Oester- reicher, and Wolfgang Raible, editors, *Language Typology and Language Universals.* number 20.2 in Handbook of Linguistics and Communication Science, de Gruyter, Berlin and New York, 2001; 1142–1178.
**Reference Source**

Koch P, Marzo D: **A two-dimensional approach to the study of motivation in lexical typology and its first application to french high-frequency vocabulary.** *Stud Lang.* 2007; **31**(2): 259–291.
**Publisher Full Text**

Lee YJ, Sagart L: **No limits to borrowing: the case of Bai and Chinese.** *Diachronica.* 2008; **25**(3): 357–385.
**Publisher Full Text**

Levenshtein VI: **Binary codes capable of correcting deletions, insertions, and reversals.** *Soviet Physics Doklady.* 1966; **10**(8): 707– 710.
**Reference Source**

List JM: **Investigating the impact of sample size on cognate detection.** *J Lang Relat.* 2014; **11**(1): 91–101.
**Publisher Full Text**

List JM: **Future challenges for computational diversity linguistics.** *The Genealogical World of Phylogenetic Networks.* 2019a; **6**(1): 1–3.
**Publisher Full Text**

List JM: **Typology of sound change (open problems in computational diversity linguistics 9).** *The Genealogical World of Phylogenetic Networks.* 2019b; **6**(10): 1–8.
**Publisher Full Text**

List JM: **Typology of semantic promiscuity (open problems in computational diversity linguistics 10).** *The Genealogical World of Phylogenetic Networks.* 2019c; **6**(11): 1–5.
**Publisher Full Text**

List JM: **Open problems in computational diversity linguistics: Conclusion and outlook.** *The Genealogical World of Phylogenetic Networks.* 2019d; **6**(12): 1–5.
**Publisher Full Text**

List JM: **Automatic morpheme segmentation (Open problems in computational diversity linguistics 1).** *The Genealogical World of Phylogenetic Networks.* 2019e; **6**(2): 1–5.
**Publisher Full Text**

List JM: **Automatic detection of borrowing (Open problems in computational diversity linguistics 2).** *The Genealogical World of Phylogenetic Networks.* 2019f; **6**(3): 1–6.
**Publisher Full Text**

List JM: **Automatic sound law induction (Open problems in computational diversity linguistics 3).** *The Genealogical World of Phylogenetic Networks.* 2019g; **6**(4): 1–8.
**Publisher Full Text**

List JM: **Automatic phonological reconstruction (Open problems in computational diversity linguistics 4).** *The Genealogical World of Phylogenetic Networks.* 2019h; **6**(5): 1– 7.
**Publisher Full Text**

List JM: **Simulation of lexical change (Open problems in computational diversity linguistics 5).** *The Genealogical World of Phylogenetic Networks.* 2019i; **6**(6): 1–6.
**Publisher Full Text**

List JM: **Simulation of sound change (Open problems in computational diversity linguistics 6).** *The Genealogical World of Phylogenetic Networks.* 2019j; **6**(7): 1–6.
**Publisher Full Text**

List JM: **Statistical proof of language relatedness (Open problems in computational diversity linguistics 7).** *The Genealogical World of Phylogenetic Networks.* 2019k; **6**(8): 1–6.
**Publisher Full Text**

List JM: **Typology of semantic change (Open problems in computational diversity linguistics 8).** *The Genealogical World of Phylogenetic Networks.* 2019l; **6**(9): 1–5.
**Publisher Full Text**

List JM: **Automatic inference of sound correspondence patterns across multiple languages.** *Comput Linguist.* 2019m; **45**(1): 137–161.
**Publisher Full Text**

List JM: **Automated methods for the investigation of language contact, with a focus on lexical borrowing.** *Lang Linguist Compass.* 2019n; **13**(10): e12355.
**Publisher Full Text**

List JM: **Inference of partial colexifications from multilingual wordlists.** *Front Psychol.* 2023; **14**: 1156540.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

List JM, Forkel R: **Automated identification of borrowings in multilingual wordlists [version 3; peer review: 4 approved].** *Open Res Eur.* 2022; **1**: 79.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

List JM, Forkel R, Greenhill SJ, *et al.*: **Lexibank, a public repository of standardized wordlists with computed phonological and lexical features.** *Sci Data.* 2022a; **9**(316): 316.
**Publisher Full Text**

List JM, Hill NW, Forkel R: **A new framework for fast automated phonological reconstruction using trimmed alignments and sound correspondence patterns.** In: *Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change.* Dublin, Association for Computational Linguistics, 2022b; 89–96.
**Reference Source**

List JM, Lopez P, Bapteste E: **Using sequence similarity networks to identify partial cognates in multilingual wordlists.** In: *Proceedings of the Association of Computational Linguistics 2016 (Volume 2: Short Papers).* Berlin, Association of Computational Linguistics, 2016a; **2**: 599–605.
**Publisher Full Text**

List JM, Pathmanathan JS, Lopez P, *et al.*: **Unity and disunity in evolutionary sciences: process-based analogies open common research avenues for biology and linguistics.** *Biol Direct.* 2016b; **11**: 39.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

List JM, Vylomova E, Forkel R, *et al.*: **The SIGTYP shared task on the prediction of cognate reflexes.** In: *Proceedings of the 4th Workshop on Computational Typology and Multilingual NLP.* Seattle, Association for Computational Linguistics, Max Planck Institute for Evolutionary Anthropology, 2022c; 52–62.
**Publisher Full Text**

Liu G, Reda FA, Shih KJ, *et al.*: **Image inpainting for irregular holes using partial convolutions.** In: Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Proceedings of the 15th European Conference on Computer Vision (ECCV 2018).* Munich, Germany, Springer International Publishing, 2018; **11215**: 89–105.
**Publisher Full Text**

Meloni C, Ravfogel S, Goldberg Y: **Ab antiquo: neural proto-language reconstruction.** In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* Association for Computational Linguistics, 2021; 4460–4473.
**Publisher Full Text**

Miller JE, List JM: **Detecting lexical borrowings from dominant languages in multilingual wordlists.** In: *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. Short Papers.* Association of Computational Linguistics, 2023; 2591–2597.
**Reference Source**

Miller JE, Tresoldi T, Zariquiey R, *et al.*: **Using lexical language models to detect borrowings in monolingual wordlists.** *PLoS One.* 2020; **15**(12): e0242709.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Milton J: **Measuring second language vocabulary acquisition.** DeGruyter, Berlin and New York, 2009.
**Publisher Full Text**

Milton J: **The development of vocabulary breadth across the CEFR levels. A common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, and textbooks across Europe.** In: INge Bartning, Maisa Martin, and Ineke Vedder, editors, *Communicative proficiency and linguistic development: intersections between SLA and language testing research.* Eurosla, York, 2010; 211–232.
**Reference Source**

Mortarino C: **An improved statistical test for historical linguistics.** *Stat Method Appl.* 2009; **18**(2): 193–204.
**Publisher Full Text**

Nation ISP, Coxhead A: **Measuring native-speaker vocabulary size.** John Benjamins, 2021.
**Reference Source**

Neureiter N, Ranacher P, Efrat-Kowalsky N, *et al.*: **Detecting contact in language trees: a Bayesian phylogenetic model with horizontal transfer.** *Humanit Soc Sci Commun.* 2022; **9**(1): 205.
**Publisher Full Text**

Nunn CL: **The comparative approach in evolutionary anthropology and biology.** University of Chicago Press, Chicago and London, 2011.
**Reference Source**

Papakitsos EC, Kenanidis IK: **Going to the root: Paving the way to reconstruct the language of homosapiens.** *International Linguistics Research.* 2018; **1**(2): 1–16.
**Reference Source**

Prendergast C: **Therapy dialect is a thing right now – but could it be toxic?** *Vogue.* 2022; **129**(7).
**Reference Source**

Ringe DA: **On calculating the factor of chance in language comparison.** *T Am Philos Soc.* 1992; **82**(1): 1–110.
**Publisher Full Text**

Roberts G, Sneller B: **Empirical foundations for an integrated study of language evolution.** *Language Dynamics and Change.* 2020; **10**: 188–229.
**Publisher Full Text**

Sagart L, Jacques G, Lai Y, *et al.*: **Dated language phylogenies shed light on the ancestry of Sino-Tibetan.** *Proc Natl Acad Sci U S A.* 2019; **116**(21): 10317–10322.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Schweikhard NE: **Semantic promiscuity as a factor of productivity in word formation.** *Computer-Assisted Language Comparison in Practice.* 2018; **1**(11).
**Reference Source**

Société de Linguistique de Paris: **Statuts.** *Bulletin de la Société de Linguistique de Paris.* 1871; **1**: III–IV.

Starostin SA: **Sravniteln'o-istoričeskoe jazykoznanie i leksikostatistika.** In: S. V. Kullanda, Ju. D. Longinov, A. Ju. Militarev, E. Je. Nosenko, and V. A. Shnirelan, editors, *Materialy k diskussijam na konferencii.* Institut Vostokovedenija, Moscow, 1989; **1**: 3–39.

Tinbergen N: **On aims and methods of ethology.** *Zeitschrift fur Tierpsychologie.* 1963; **20**(4): 410–433.
**Publisher Full Text**

Trask RL, editor.: **The dictionary of historical and comparative linguistics.** Edinburgh University Press, Edinburgh, 2000.
**Publisher Full Text**

Turchin P, Peiros L, Gell-Mann M: **Analyzing genetic connections between languages by matching consonant classes.** *J Lang Relat.* 2010; **3**: 117–126.
**Reference Source**

Urban M: **Analyzability and semantic associations in referring expressions: A study in comparative lexicology.** Phd, Leiden University, Leiden, 2012.
**Reference Source**

Urban M: **Motivation by formally analyzable terms in a typological perspective: An assessment of the variation and steps towards explanation.** In: Päivi Juvonen and Maria Koptjevskaja-Tamm, editors, *The lexical typology of semantic shifts*. De Gruyter Mouton, Berlin and New York, 2016; 555576.
**Publisher Full Text**

Vaswani A, Shazeer N, Parmar N, *et al.*: **Attention is all you need.** In: I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems.* Curran Associates, 2017; **30**: 1–11.
**Reference Source**

Weinreich U, Labov W, Herzog MI: **Empirical foundations for a theory of language change.** In: Winfred Philipp Lehmann and Yakov Malkiel, editors, *Directions for historical linguistics: A symposium.* (-195?) University of Texas Press, Austin, 1968; 95–189.
**Reference Source**

Whewell WDD: **The philosophy of the inductive sciences, founded upon their history, volume 2.** John W. Parker, London, 2 edition, 1847; **2**.
**Reference Source**

Wilson EO: **Consilience. The unity of knowledge.** Vintage Books, New York, 1998.
**Reference Source**

Xu Y, Malt BC, Srinivasan M: **Evolution of word meanings through metaphorical mapping: systematicity over the past millennium.** *Cogn Psychol.* 2017; **96**: 41–53.
**PubMed Abstract** | **Publisher Full Text**

# Open Peer Review

## Current Peer Review Status: ✔ ✔ ✔ ✔ ✔

**Version 2**

Reviewer Report 19 August 2024

https://doi.org/10.21956/openreseurope.19255.r41040

© **2024 Pleyer M.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Michael Pleyer**
Center for Language Evolution Studies,, Nicolaus Copernicus University in Toruń, Toruń, Poland

The author has appropriately responded to my comments. I accept that our views on some of the issues surrounding language evolution and its perception differ, and the author presents a valid perspective even if I might disagree with some aspects. So I give my approval to the indexing of this article.

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Language Evolution, Evolutionary Linguistics, Cognitive Linguistics, Construction Grammar

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 19 Aug 2024

**Johann-Mattis List**

Many thanks for this assessment. I look forward to pursuing the discussion in future work.

***Competing Interests:*** No competing interests were disclosed.

Reviewer Report 07 June 2024

https://doi.org/10.21956/openreseurope.19255.r41037

**© 2024 Good J.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Jeff Good**
University of Buffalo, Buffalo, USA

The revisions based on my feedback are appropriate in my view, and I appreciate the time the author took to respond to them.

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** linguistics, historical linguistics, linguistic typology

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Version 1**

Reviewer Report 22 February 2024

https://doi.org/10.21956/openreseurope.18157.r37446

**© 2024 Pleyer M.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Michael Pleyer**
[1] Center for Language Evolution Studies,, Nicolaus Copernicus University in Toruń, Toruń, Poland
[2] Center for Language Evolution Studies,, Nicolaus Copernicus University in Toruń, Toruń, Poland

In this essay, the author discusses the current state of ten “open problems in computational historical linguistics” that he proposed in a series of blog posts in 2019 and envisioned as solvable within the next five years. Now, five years later he takes stock of the progress that has been made. He comes to the conclusion that although none of the problems have been solved, at least for some of them, progress has been made. This leads the author to slightly modify his list.

This is an interesting essay, written in a personal style (which probably wouldn’t be appropriate for a research article but is fitting for an essay offering a personal perspective) that highlights a number of interesting open problems for computational historical linguistics, and historical linguistics in general. The author does a good job of relating these problems to the question of “big” and “small” problems in science and previous conceptualisations of the problems that historical linguistics should address.

The two major issues I see relate to its discussion of language evolution and the field of language

evolution. The author starts with a discussion of the question of language evolution and how it is seen by (historical) linguists. This discussion is interesting, but I felt that the connection to the next parts of the essay, the discussion of types of problems, and the ten specific open problems in computational linguistics could have been made clearer. What exactly does the discussion of language evolution contribute to the other two issues, “problems” and “problems in computational historical linguistics” and how are they related?

The other major issue concerns the discussion of the state of language evolution research itself. It is certainly true, as the author attests from his own experience, that historical linguistics has mostly steered away from the question of the evolutionary emergence of language and the human language capacity. Here the author reports from his own experience, but the case might be further strengthened by also referring to published works that make this explicit, such as [1] introduction to historical linguistics:

“Another topic not generally considered to be properly part of historical linguistics is the ultimate origin of human language and how it may have evolved from non-human primate call systems, gestures, or whatever, to have the properties we now associate with human languages in general. Many hypotheses abound, but it is very difficult to gain solid footing in this area. Historical linguistic theory and methods are very relevant for research here, and can provide checks and balances in this field where speculation often far exceeds substantive findings, but this is not a primary concern of historical linguistics itself.” (see also discussion in [4]).

The author cites the 1866 “ban” by the Linguistic Society of Paris and states that “The situation has not changed since then.” This might indeed be true for the field of historical linguistics in general, but when taken to refer to linguistics as a whole this seems to me an overstatement. In general in language evolution research, the publication of [6] has been credited with initiating a broad revival of interest in the question, and by now language evolution research and evolutionary linguistics can be seen as strong interdisciplinary fields that also features many contributions by linguistics. There is a biannual “Conference on the Evolution of Language” (Evolang) and a *Journal of Language Evolution*, as well as an Oxford Handbook of Language Evolution [7] and a number of textbooks (e.g. [3]; [5]).
In particular, the view of the impact of the ban stands in stark contrast to some published evaluations of it:

“In the interim [before 1990], the story goes, all that happened was a comical series of silly unscientific hypotheses, nicknamed “bow-wow,” “heave-ho,” and “ding-dong” to expose their basic absurdity. *This view of the field is a myth*. Darwin himself, and subsequent linguists such as Jespersen, made important contributions to this literature after the famous ban, and there was a major, important revival of interest in the 1960s and 1970s when many of the issues under discussion today were already debated insightfully (e.g. Hockett and Ascher, 1964; Hewes, 1973; Harnad et al., 1976).” [3].

So again, I think with regard to historical linguistics, the author might be spot on, but with regard to the field of linguistics as a whole, I believe there have been significant changes regarding whether language evolution is a respectable area of inquiry, so I am not sure how true it is that as a whole “linguists would deliberately decide to ignore the investigation of a problem that some might consider the most fascinating the field has to offer.”

The point that specifically for outsiders " it is often very difficult to tell if a theory on the origin of language should be treated as a serious or a senseless idea" is very well-taken, but the essay's juxtaposition of examples of this seemed a bit weird to me. It seems clear to me that Herder's 1778 essay contribution on the origins of language should not be seen as a modern theory of language evolution, and that a theory invoking the concepts of *yin* and *yang* in a not very well-known journal, by authors who are not really active members of the language evolution community, should be treated with more skepticism than a careful consideration of the evidence for Neanderthal language capacities by highly-cited researchers.

I am also a bit unhappy with the characterisation that there is "debate whether a single gene was responsible for our language faculty" as this is a minority view, and most discussions of FOXP2 revolve around its role in language evolution and the evolution of speech as one (important) contributing factor, and not the "language gene", a view that has already been outdated in the 90s.

Again, the point that there is a danger of bold and careless speculation is well-taken, but given the reality of the field of language evolution research as a whole, I think the framing of this point could be improved.

It also seems to me just slightly too strong of a generalisation that "historical linguists never ask how many words a language has" – although saying that they "hardly ever ask" is probably correct. As one case in point, [2] makes reference to estimates of the vocabulary sizes of different societies. (The problem here of course is that such estimations such as the ones mentioned by [2] historically have been tied to racist and colonialist assumptions, which makes this issue a difficult one to navigate).
But a short reference to this history of size assumptions in comparative linguistics (and its dangers) might be helpful for this essay in addition to the ones that are mentioned.

There are also some smaller issues/minor comments, that can be easily addressed but should be addressed:
- There are number of typos in the manuscript that should be taken care of, e.g.
- p. 4 "loose" à "lose" p.6 "extend", p.8 "was" à ways, p. 9 "workflow for to test for", "dataset"
- the plain language summary doesn't make reference to the "computational" part of historical linguistics.
- Page 4, fn 3, it might be good to mention that Dąbrowska (2020) is part of a bigger research programme of the author to investigate the question of "differences in competence" that she has also investigated for English, for example.
- p.5/6 when discussing Hilbert problems, the author could make it a bit clearer if it is or can be related to the previous distinction of p- vs g-linguistics
- Table 2: it might be useful to add to this table which of the problems are modelling, inference, and analysis problems (as is then discussed in the text).
- p.10 In his discussion of the concept-based vs form-based perspective, the author might want to relate this to the traditional distinction in linguistics between a semasiological and an onomasiological perspective.

**References**

1. Lyle Campbell: Historical Linguistics: An Introduction. *Edinburgh University Press*. 1998. Reference Source

2. Guy Deutscher: Through The Language Glass: Why The World Looks Different in Different Languages. *Metropolitan Books Henry Holt and Company, LLC*. 2010. Reference Source
3. Fitch W: The Evolution of Language. 2010. Publisher Full Text
4. Hartmann S: Language change and language evolution: Cousins, siblings, twins?. *Glottotheory*. 2020; **11** (1). Publisher Full Text
5. McMahon A, McMahon R: Bibliography. 2012. 267-301 Publisher Full Text
6. Pinker S, Bloom P: Natural language and natural selection. *Behavioral and Brain Sciences*. 1990; **13** (4): 707-727 Publisher Full Text
7. The Oxford Handbook of Language Evolution. 2011. Publisher Full Text

**Is the topic of the essay discussed accurately in the context of the current literature?**
Yes

**Is the work clearly and cogently presented?**
Yes

**Is the argument persuasive and supported by appropriate evidence?**
Yes

**Does the essay contribute to the cultural, historical, social understanding of the field?**
Yes

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Language Evolution, Evolutionary Linguistics, Cognitive Linguistics, Construction Grammar

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 22 Feb 2024

**Johann-Mattis List**

Many thansk for this very inspiring response. I hope to find time soon to reply to all reports in a revised version of the essay. Here, I hope to be able to answer all points in a satisfying way.

***Competing Interests:*** No competing interests were disclosed.

Author Response 16 May 2024

**Johann-Mattis List**

Many thanks for the critical remarks. I learned a lot from this review and found interesting new literature. The revised version addresses most of the points raised, I hope. Regarding the problem of coherency (why discuss language evolution and language origins at length,

when then discussing concrete problems), I agree that this coherency may be lacking, I would however, feel reluctant to change the draft substantially in this regard, since — as other reviewers have also acknowledged — this is an essay, that should give me the possibility to jump in my lines of thoughts and to also report personal opinons and viewpoints that I would not present in this form in a proper research paper. I have added a reference to Campbell (1998) now, as an example for a text book introduction to historical linguistics that is very explicit on not treating language origin as a topic. I have seen that my statement on language evolution and language origin as scientific inquiries may have been a bit provocative and also not fair to the field. I have tried to modify the statements now. I hope that the demanded restriction to historical linguistics — including mainstream linguistics where historical linguistics is also usually ignored — makes the previous claims appear less general. The qualititative difference between the studies that I mention as examples is of course very obvious. But there are nuances one should not forget. While the yin-yang example is a strong outlier, one should place Herder into the historical context, and here, he was quite revolutionary, since he claimed that humans came up with the language faculty without a god. As silly as the theory may have sounded, in its pre-evolutionary context, the study was still quite important and freed the way for bolder demands. I have modified the passage now in such a way that I emphasize the difficulty of distinguishing serious from less serious attempts. The fact that some of the highly speculative proposals are published in big journals like Nature (think of the out of Africa discussion based on phoneme inventories) makes it in my opinion clear that a distinction is not that easy to draw between serious and non-serious research here. I have modified the wording on Fox-P and placed it into past tense, emphasizing that the debate is not recent. The fact that this research makes it into the big magazines like Cell shows again the suggestive force the topic bears, which is the main reason why it is discussed in this context. I hope that these modifications are sufficient to modify and adjust the potentially negative framing of the part on language evolution / origin in the previous version. Regarding the size of languages, I have modified the wording as suggested. Deutscher's book is an interesting reference for which I thank the reviewer. I have added a note mentioning it to the new version now. I have also addressed the problem of typos, and the minor points mentioned by the reviewer.

***Competing Interests:*** No competing interests were disclosed.

Reviewer Report 15 February 2024

https://doi.org/10.21956/openreseurope.18157.r36348

**© 2024 Nerbonne J.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**John Nerbonne**
[1] Computational Linguistics and chair of Humanities Computing, University of Groningen,

Groningen, The Netherlands
[2] Germantistische Linguistik, Albert-Ludwigs Universität, Freiburg, Germany
[3] Computational Linguistics and chair of Humanities Computing, University of Groningen, Groningen, The Netherlands
[4] Germantistische Linguistik, Albert-Ludwigs Universität, Freiburg, Germany

Summary: Johannes-Mattis List (JML) provides an update on a paper from 2019 on ten outstanding problems in computational historical linguistics. It will serve as an excellent survey for researchers in this field, and may help focus work there.

Recommendation: Definitely index, drawing the author's attention to (some of) the comments below, which he may care to take as reason to modify the paper a bit.

General points: JLM refers several times to questions that will will/could/should be solved in the nearer (or farther) future. It would be interesting to hear him make the time references more concrete, e.g. in the next five vs. the next fifteen years.

I liked the discussion of Weinreich et al.'s vs. Coseriu's crucial questions, and they serve the sensible purpose of anticipating objections from readers.

Issues:
p.4, col.2 the size of a language. George Miller's 'The Science of Words" provides an estimate of the vocabulary size of Americans leaving high school, and he put it at 50K, many of which were derivationally related, using it to estimate how many new words were learned daily, and thence to why structure sensitivity was need to explain this.

Comments & questions:

1. Abstract. Would it be sensible to identify the blog where these ideas were discussed? This would be an indirect acknowledgment of contributions from blog participants.
2. pp. 3-4. I applaud the author's attempt to identify significant problems in computational historical linguistics and thereby focus work on them. This is great.

But I cannot resonate to an introduction that justifies this by claiming that the "driving force of all scientific inquiries are [sic] problems". I suspect that one could defend this against many objections by noting that attempts to find generality, for example, attack a "problem of too little generality", but I also think, first– from an amateur's view of the history of science-- that lots of advances have been stimulated by searches for generality or simplicity (a lot of physics in the last half century, chemistry in the Mendeleyev period), and second, that the focus on problem solving doesn't characterize science alone. The Pragmatists Peirce and Dewey identified the solving of problems as the goal of all human thought.
So, please keep the focus on the outstanding problems, but build up to it in a less highfalutin introduction.
At the risk of repetition, this is a sensible and worthwhile paper.

1. p.7 "Inference problems. Maybe clarify that you're aware that many morphologists reject the notion of morpheme, e.g., paradigm function morphology, and insist on the primacy of processes within paradigms. In the alternative view Umlaut (but also Ablaut) is a morphological process that isn't comfortably reduced to the effect of a morpheme.
2. p.9 "Analysis problems" I was surprised that there was no mention of all the work in

computational linguistics on semantic change. A collection on this, with many references: Tahmasebi *et al.* (2021[1]).

Form: The paper is written clearly in nearly flawless English, which I appreciated. Nonetheless, here are some details which might be improved.

- p.3,col.1 “working luckily” => working fortunately.
- p.3, col. 1 I’d prefer a translation of the Société’s ban on papers on the origin of language, “La Société n’adment aucune communication concernant, soit l’origine du langage, soit la création d’une langue universelle,” as “the Société does not accept any contributions concerning either the origin.
- of the language faculty, or the creation of a universal”, i.e. instead of neither … nor
- p.4, col.1 “the amount of genes” => the number of genes In general, ‘amount’ specifies quantity (of mass noun), ‘number’ quantifies plural count nouns. But this is coming from an older, conservative writer.
- p.5, col ‘suggested to distinguish’ => suggested distinguishing.
- p.6, col.2, Table 2: Add to caption what I, m, and a indicate (Sci. Am.’s rules for captions: make graphics and table independently understandable, if possible.)
- p.6, Fig. 1 Explain significance of stars vs. squares vs circles, ditto blue vs. yellow stars.
- p.7, col.1, l.^3 closing parenthesis.
- p.7, col.2, Section “ New and …” last line: verb should be in present perfect.

**References**

1. Tahmasebi N, Borin L, Jatowt A, Xu Y, et al.: Computational approaches to semantic change. *Berlin: Language Science Press.*2021. Reference Source

**Is the topic of the essay discussed accurately in the context of the current literature?**

Yes

**Is the work clearly and cogently presented?**

Yes

**Is the argument persuasive and supported by appropriate evidence?**

Yes

**Does the essay contribute to the cultural, historical, social understanding of the field?**

Yes

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Computational Linguistics, Dialectology

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 15 Feb 2024

**Johann-Mattis List**

Thanks a lot for this detailed and helpful review, I will try to make sure to answer to all points raised, once I sit down to work on the revision.

***Competing Interests:*** No competing interests were disclosed.

Author Response 16 May 2024

**Johann-Mattis List**

This review has been very helpful to me and I hope that my modified version does justice to all remarks and points raise. I have tried to make clearer now when I think that problems may be solved (although I am hesitant to give concrete time frames). I was trying to identify the work by Miller but could not receive it in the original form for this revision. So I did not cite it in the current version, hoping also, that the reference to the more recent study by Nation (2021) is providing a suitable starting point for readers interested in the problem. The blog is linked now in the abstract. I have modified the characterisation of scientific inquiries as suggested by the reviewer. Problem solving is now one, but not the only scientific task. Inference problems have been clarified now. I have added another reference to a study by Xu et al (2017), that also illustrates how semantic change can be studied with corpora, emphasizing that these studies are particular language studies and not typological surveys. So one can do this research with English, but solving semantic change in English does not solve questions on semantic change in general. I have incorporated all hints on spelling errors and grammatical problems for which I am very grateful.

***Competing Interests:*** No competing interests were disclosed.

Reviewer Report 13 February 2024

https://doi.org/10.21956/openreseurope.18157.r37453

**© 2024 Kalyan S.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Siva Kalyan**
[1] The University of Queensland, Saint Lucia, Queensland, Australia
[2] Australian National University, Canberra, Australian Capital Territory, Australia

[3] The University of Queensland, Saint Lucia, Queensland, Australia
[4] Australian National University, Canberra, Australian Capital Territory, Australia

This essay starts by reflecting on the importance of open problems in setting research priorities in a scientific discipline, and points out the large difference between what non-linguists consider to be important questions about language (e.g. how “big” a language is) and what linguists consider to be important questions about language.

The author then reviews preceding attempts to list the open problems in historical linguistics, and finds them all to be unsatisfying, as they do not translate into concrete research questions.

The author then introduces his own list of ten open problems in computational historical linguistics, which he announced in January 2019. He reviews the progress that has been made on each of the ten problems since then, and concludes while there has been substantial progress towards solving two of the problems (automated borrowing detection and automated phonological reconstruction), the problems are far from being solved. He also revises the formulation of two of the problems, to better reflect his understanding of the field.

Overall, I find the first three sections a bit lengthy, and would prefer the author’s list of open problems to be introduced sooner. However, if the audience for this essay includes non-linguists (or non-historical linguists), then the first three sections contain valuable background information.

My only suggestion for improvement is that the replacement of the term “semantic promiscuity” with “lexical root productivity” should be introduced earlier, since the former term is very confusing, and especially so for non-linguists.

**Is the topic of the essay discussed accurately in the context of the current literature?**
Yes

**Is the work clearly and cogently presented?**
Yes

**Is the argument persuasive and supported by appropriate evidence?**
Yes

**Does the essay contribute to the cultural, historical, social understanding of the field?**
Yes

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Computational historical linguistics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 15 Feb 2024

**Johann-Mattis List**

Many thanks for these insightful comments, which I'll try to work into a revision of the essay.

***Competing Interests:*** No competing interests were disclosed.

Author Response 16 May 2024

**Johann-Mattis List**

Again, many thanks for this very positive review. In the revised version, I have tried to address the point of being more explicit -- earlier in the text -- about lexical root productivity.

***Competing Interests:*** No competing interests were disclosed.

Reviewer Report 31 January 2024

https://doi.org/10.21956/openreseurope.18157.r37450

**© 2024 Brochhagen T.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Thomas Brochhagen**
[1] Universitat Pompeu Fabra, Barcelona, Catalonia, Spain
[2] Universitat Pompeu Fabra, Barcelona, Catalonia, Spain

This work discusses motivations and progress made toward addressing ten problems in historical computational linguistics, originally proposed by the author five years ago. I see the contribution as twofold. On the one hand, as argued within, there is use in taking a step back and evaluating where the field stands; what challenges it faces; and how they may be solved. On the other, after establishing a set of challenges, there is worth in iteratively looking back and evaluating to which extent these challenges have been solved; adjusting the problems and their framing as necessary.

Overall, I believe the manuscript in its present form already convincingly achieves the above so I'll just raise a few minor issues for the consideration:

1. I believe the Introduction is too pessimistic. First, while the ban from the Societe de Linguistique de Paris may have had its impact and is striking in its own right, it was abandoned within a decade and is best judged in the historical context it happened in (see, e.g., https://doi.org/10.1142/9789814401500_0133 for pointers). Second, and more importantly, much interesting and productive research on the evolution of language (both in terms of its emerge and in terms of the dynamics of change involved) has been conducted in the last decades. It may be

that many proposals/models will turn out to be wrong in one way or another but, as I see it, that's the currency of science. The tone of the manuscript (particularly around the 4th paragraph) struck me as dismissive and as a bit of caricature. We can agree that there is outrageous speculation in the field, but I don't think this area of research is unique in this respect. I'd encourage the author to elaborate more on their stance, if it is indeed as pessimistic as it sounds in the current version, or to consider attenuating this part a little.

2. There is little to no discussion on "automated morpheme segmentation". I found this curious considering that (i) subword tokenization is an important component of the NLP-pipeline of all successful Large Language Models, which are also increasingly becoming multilingual (see, e.g., mT5 [https://github.com/google-research/multilingual-t5] or BigScience's Bloom [ https://bigscience.huggingface.co/blog/bloom]); and that (ii) since 2019 morpheme segmentation has, for instance, figured as a shared task in the Special Interest Group on Computational Morphology and Phonology (https://sigmorphon.github.io/sharedtasks/). It may be that these NLP-oriented approaches are not what the author had in mind since they are often not linguistically motivated. But they do fit the manuscript's preference for unsupervised solutions. I was left wondering what they lack to not even be considered solutions toward the challenge.

3. I concur with the Jeff Good (https://doi.org/10.21956/openreseurope.18157.r36345) that (i) the manuscript would benefit from careful proofreading to eliminate mistakes (e.g., "[...] we may easily loose the chance us to enrich our [...]" on p. 4) and that, in particular, (ii) Figure 2 would benefit from some further context to be interpreted as intended, either in the caption or in the text.

**Is the topic of the essay discussed accurately in the context of the current literature?**
Yes

**Is the work clearly and cogently presented?**
Yes

**Is the argument persuasive and supported by appropriate evidence?**
Yes

**Does the essay contribute to the cultural, historical, social understanding of the field?**
Yes

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** computational cognitive modelling; Bayesian methods; lexical typology

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 15 Feb 2024

**Johann-Mattis List**

Thanks a lot for these very helpful comments which I'll try to make sure to include in a

revision of the essay that I am planning to submit soon.

***Competing Interests:*** No competing interests were disclosed.

Author Response 16 May 2024

**Johann-Mattis List**

Thanks again for this very useful review. I have now finished my revision and tried to address the points raised. Specifically, regarding the point that I was too pessimistic about language evolution, I have tried to modify the wording now. I agree that my introduction was pessimistic, and I have modified this statement now, by also emphasizing that it is my personal opinion and that it reflects the impression I have from the research situation in Germany, but I think that given the essay character, it is justified to share this impression. I think my harsh judgment on the field is also particularly justified since linguistics has not managed to bridge the gap of several hundred thousand years since language evolved the first time and where out classical methods for historical language comparison date back. While we have a lot of convergence regarding the theories of particular language families, we do not find a communis opinio on how language originated. We do not even find a concrete scenario in the current research devoted to the topic. Regarding the topic if subword tokenization and morpheme segmentation, I hope the new version addresses this problem appropriately. The problem of subword tokenization is that it does not solve the problem at hand. We are talking about a very specific problem here, where the goal is to find the morphemes in a list of about 1000 words of a given language. In the original blog post I reference, it is illustrated (and this has not changed until now) that tools like Morfessor fail to identify morphemes when given only 1000 words of a German text. Subword tokenization also does not address this task, since the goal consists in a concrete solution of the problem, not in finding the most frequent forms in a text. I have clarified this in the revised text. Errors in the form of typos have hopefully also all been accounted for now.

***Competing Interests:*** No competing interests were disclosed.

Reviewer Report 08 January 2024

https://doi.org/10.21956/openreseurope.18157.r36345

© **2024 Good J.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

**Jeff Good**
[1] University of Buffalo, Buffalo, USA
[2] University of Buffalo, Buffalo, USA

The core of this paper provides an update on progress towards a solution to ten problems in computational historical linguistics that were proposed by the author in 2018/2019. It additionally provides more general discussion of the role of problems in advancing scientific research following the tradition of Hilbert's problems in mathematics, and much of the framing of the contribution of the paper is in terms of "big picture" questions about problems in science even if the problems outlined for computational historical linguistics in the paper are somewhat narrower in scope. The paper is somewhat unusual in the context of linguistics due to personal nature of the rhetorical framing. However, this seems reasonable since it is being submitted under the category of "essay".

Because the author has already made substantial contributions to computational historical linguistics, I believe that it would be good to see this essay accepted in some form since his ideas about major problems in this area should be of general interest, though I would recommend certain changes. Most of these represent relatively minor issues centering around clarification of the manuscript, but two are more significant. I'll discuss those first.

More significant issues:

1. As discussed above, the focus of the paper is the ten problems in computational historical linguistics that were proposed by the author in 2018/2019. While a revised set of problems is presented here, this set is still more or less based on the original set. I think, in a paper like this, it would be beneficial for the author to also consider whether he would still propose the same basic ten problems proposed in 2018/2019 as being the most significant problems today. Perhaps, for example, some of them seem less important now than they did before, and there are only (say) eight "key" problems in his view. Or, perhaps, there is an eleventh problem he thinks should be added.

2. I think that, for the arguments of the paper to be clearly interpretable, it would be important for there to be a clear definition of what the author believes "computational historical linguistics" is, what its key goals are, and what model it assumes for "language" and "language change". In practice, for example, my sense is that most work in computational historical linguistics has, as its key goal, linguistic reconstruction, and this is suggested by work such as Jaeger (2019). Achieving this goal also means arriving at a better understanding of language change, but that does not usually seem to be the main goal of such work but, rather, a means of achieving the goal of reconstruction. Put differently, computational historical linguistics seems to emphasize the reconstruction of prehistory more than, for example, developing universal theories of sound change. I have no particular idea regarding what "computational historical linguistics" should include (or not include), but I do think it would be good for a paper like this one to take a clear position on this.

Minor issues:

- I noted various minor typos in the manuscript. Before it is accepted, it should be proofread carefully.

- "Non-linguists are often very surprised that asking for the origin of language is a taboo question in the field of historical linguistics." I don't have the impression that this is a taboo anymore.

Rather, my sense is that it is understood to be outside of the domain of much of historical linguistics which is focused on how languages changed after language (as we understand it) had developed. This is, in large part because that's what the methods of historical linguistics are designed to investigate. The evolution of language is now studied, of course, but it tends to be done outside of historical linguistics in a narrow sense.

-The summary given of Weinreich et al.'s (1968, 183–187) problems does not fully match my reading of that work. For example, I don't read the "evaluation problem" as involving the degree to which change happens "consciously" but, rather, as whether language communities are aware that a change is taking place (even if it is originating in an unconscious way). I think it would be good to double check this summary. Similarly, I have always interpreted the "actuation problem" as being a version of "why do languages change". Maybe the author is correct in saying that this is not the case, but double checking the interpretation would still be a good idea in my view.

-"The failure to distinguish questions pertaining to particular languages and questions pertaining to language in general has led to many misunderstandings in the field of comparative linguistics." Can some examples be cited here?

-Table 2: I don't understand how automated morpheme segmentation is a historical computational problem rather than a general linguistic one. Can that be clarified?

-Figure 1: Can the interpretation of 20 x, 10 x, and 5x be explicitly explained in the text?

-"As an example for this procedure, consider the problem of cognate detection, the detection of historically related – or homologous – words across languages." This paper seems to be aimed at a general audience, not linguistic specialists. If so, I think the description of the concept of cognate needs to be expanded since it is discipline specific.

-"Based on this evolutionary model, we can then conclude that similar words observed across different language varieties have been inherited from the common, formerly unified, ancestral variety." Given that the paper opens with a discussion of the question of the evolution of language (from non-language), it might be best to avoid the term "evolutionary model" here since "evolution" has an ambiguous meaning in historical linguistics.

-"...but it requires a model of language relatedness that can then be tested against a random model in which languages are thought to be unrelated.": What is meant by "language relatedness"? It seems like what is meant is "genealogical relatedness under so-called normal transmission" (as opposed to, say, sharing features through contact). I think it would be good to make that explicit.

-"While there are numerous attempts in the literature to come up with a convincing statistical model to prove genetic relationship..." (footnote 5): I would add Nichols (1996) to this list.

-"Rather huge progress – at least when looking back specifically at the last couple of years – has been made with respect to supervised phonological reconstruction.": I think it would be useful to say something about the supervision involved in these approaches.

-"While scholars had been working on this problem before, a first impressive demonstration of the

capability of machine learning methods...": I think that it is worth mentioning that these are of a different character than other approaches since there is the so-called "black box" effect with them. That is, they may work well, but we don't really know why. This is opposed to computational models that require an amount of computation that would be practically impossible for humans to conduct by hand, but where each step in the process is known and understood.

-"I identified ten unsolved problems in computational historical linguistics that I considered as important and solvable at the same time." Following on from my general comment above, would the author still consider all of these important (or as important) today?

**References**

1. Jäger G: Computational historical linguistics. *Theoretical Linguistics*. 2019; **45** (3-4): 151-182
Publisher Full Text
2. Durie M, Ross M: The Comparative Method Reviewed : Regularity and Irregularity in Language Change. *oxford university press*. 1996.

**Is the topic of the essay discussed accurately in the context of the current literature?**
Yes

**Is the work clearly and cogently presented?**
Yes

**Is the argument persuasive and supported by appropriate evidence?**
Yes

**Does the essay contribute to the cultural, historical, social understanding of the field?**
Yes

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** linguistics, historical linguistics, linguistic typology

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 15 Feb 2024

**Johann-Mattis List**

Thanks a lot, these comments are very helpful, and I will try to make sure to reply to all of them when doing my revision.

***Competing Interests:*** No competing interests were disclosed.

Author Response 16 May 2024

**Johann-Mattis List**

Thanks a lot for this very useful review of the essay. In my revision, I have tried to address all points mentioned. I summarize major modifications below.

1. I have added one sentence clarifying that due to the fact that the problems have not been solved, I would not revise the list by now.
2. I have added a sentence explaining what I understand when using the term "computational historical linguistics". My definition in this context is much more narrow than trying to find a computational solution to questions on sound change, but rather aims to cover those approaches that try to formalize with a computer what people have been doing manually so far.
3. I have tried to account for all errors and hope I have identified most of the ones that were noted by the reviewer.
4. Regarding the origin of language, I have added a statement now, also since other reviewers were not happy with my wording, where I emphasize that it is my personal experience that the question is avoided, and also emphasized that it may be restricted to the context of Germany, where we have a tradition of distinguishing evolution and language origins from language history and language development in terms of the terminology we use.
5. Regarding Weinreich et al.'s article, I am a afraid that I do not know how to read the "why" in "why languages change". There are two answers here, following the notion by Coseriu, namely "for what goal" (why is it useful to change) or "what specific aspect of the nature of languages triggers their change, of which we observe that it always happens?". In this context, I think that my reading of the "actuation" problem as referring to the triggers of change seems to hold to me, but if it can be further specified where my reading fails, I would be very glad and change it accordingly. Regarding the evaluation problem, my intention was to specifically point to the reading the reviewer suggests, so I changed that part in the text.
6. I have tried to address the problem of distinguishing p- and g-linguistics. I have made an attempt to specify by pointing to methods that would only work on one orthography. Examples with references are not mentioned here, since the majority of encounters is in review situations. One could extend the passage by referring to the problem of equating English with Language, but I hope it is clearer already in the current form.
7. I have clarified why morpheme segmentation is important in historical linguistics.
8. I have added a concrete example to the example on cognates (German and English).
9. I avoid the term "evolutionary model".
10. I specify what I mean by language relatedness ("common ancestry").
11. Regarding Nichols (1996): The study could be added, but since the work is usually not cited in this context, since the theory proposed in this paper cannot be quantified (it is an example, but nobody has made a computer model out of it, unlike the work by Baxter and others mentioned in this context), I'd prefer to not quote the article in this context.
12. Supervised vs. unsupervised approaches are now explained (I hope it is in sufficient detail).
13. A short statement on the black box character of some appraoches has been added.

14. I try to be more explicit regarding the importance of the problems in my work and if I see new ones by now.

***Competing Interests:*** No competing interests were disclosed.