ORIGINAL PAPER

RelChronVis: an interactive web application for visualizing the relative chronology of language changes

Florian Wandl[1] · **Thilo H. K. Thelitz[2]**

Abstract

Models of the relative chronology of language changes resemble the structure of a graph. They consist of a set of changes that are in many cases related to each other. It follows that they can be visualized in a diagram, which opens up new possibilities for publishing historical linguistics data. In this paper, we introduce the RelChronVis application, a web application designed for visualizing language history. Based on detailed models of the relative chronology of Russian and Croatian sound changes, we show how interactive graph representations can help visualize language changes and the relationships between them. In addition to graphs, the application provides textual information about the included changes and relations as well as reconstructions of example lexemes. The RelChronVis application is primarily designed for doing research and for teaching. It allows the researcher to transparently compare different models and to access relevant data. At the same time, the RelChronVis application can be used to teach the history of the included languages as well as methods of historical linguistics. It provides the option of visualizing custom data and can easily be accommodated to novel discoveries. We believe that these features and uses make the RelChronVis application an immensely useful tool for publishing models of the relative chronology of language changes.

Keywords Data visualization · Historical linguistics · Language change · Relative dating · Relative chronology

✉ Florian Wandl
  florian.wandl@uzh.ch

  Thilo H. K. Thelitz
  thilo.thelitz@helsinki.fi

1   University of Zurich, Zurich, Switzerland

2   University of Helsinki, Helsinki, Finland

⚉ Springer

1 Introduction

Recent years have seen a significant increase in the number of digital resources and tools available to historical linguists. Today, it can be considered common standard to publish linguistic data online and to visualize results with different types of diagrams whenever practical. However, such visualizations are often limited to particular datasets within one publication, and the overall mode of presenting and discussing the research has basically remained the same. For almost two centuries, historical linguists have been publishing their research in articles and books, be it in print or, in more recent years, in different formats online. This paper introduces the RelChronVis web application,[1] which provides an alternative way to represent language history based on graph representations. It is based on detailed models of the order according to which the changes of the languages under scrutiny occurred. Among other things, this makes it possible to provide easy access to information that is otherwise distributed across one or more publications, and to research linguistic events more closely with regard to the place they occupy on the timeline. While the models underlying this paper include only sound changes, it should be noted that the digital model can be amended with data from other subfields of linguistics (morphology, syntax etc.). Moreover, the inclusion of further data does not have to be limited to linguistics. The digital model introduced in this paper provides a basis for linking data from different scientific disciplines.

Our paper is structured as follows. In Section 2, we discuss some of the methodology by which the order of linguistic changes can be determined (Section 2.2), and we present how models developed by applying the discussed methods can serve as a basis for creating digital representations of language history (Section 2.4). Section 3 contains the technical part of our paper, explaining how we digitized the linguistic models and how the web application functions. Section 4 is dedicated to the potential of our digital model for doing research (Section 4.1) and for teaching historical linguistics (Section 4.2). The outlook in Section 5 contains our ideas for further developing the resource. Section 6 concludes the paper.

2 Linguistic background

The RelChronVis web application is based on two publications that provide detailed models of the relative chronology of Croatian and Russian sound changes, respectively (Holzer, 2007; Wandl, 2011). To give the reader a better understanding of our data, in this section, we introduce the concept of relative chronology (Section 2.1), present the methods that have been applied in developing these two models (Section 2.2) and discuss how they are presented in traditional formats (Section 2.4).

2.1 The term 'relative chronology'

The term 'relative chronology' seems to have its origin in geology from where it spread to archaeology and, possibly via the latter, to historical linguistics (Anttila, 1989, pp.

---

[1] Accessible at https://relchron.eu.pythonanywhere.com

294-297; Andersen, 2003). In geology and archaeology it is related to stratigraphy, the study of rock layers. For example, archaeologists can establish a relative chronology when two objects are found in two chronologically different strata of an excavation site (O'Brien & Lyman, 2002). Although it may not be possible to provide absolute dates for the origin of the two objects at hand, one of them must be older than the other. Hence, they are dated relatively to each other. We speak of absolute dating, on the other hand, if the objects are dated with methods like radiocarbon dating or dendrochronology.

In historical linguistics, the term relative chronology is typically used with regard to the relative order of changes, and this is also how it is understood in the works serving as the basis for the RelChronVis application (for the use of the term with reference to the order of specific language states see, i.e., Anttila (1989, pp. 294-297); Andersen (2003)). Language changes can be dated relatively to each other when a relationship between them can be inferred. For instance, one change might have created inputs for another change, and must consequently have occurred before it (see Section 2.2 for more details).

Historical linguistics, like archaeology, also makes use of absolute dating. Usually, written records form the basis of this. For example, if a document that for some reason can be dated to a certain year or period in the past contains the reflexes of a certain change, we can establish that this change must have occurred prior to that year or period. Hence, we can establish the *terminus ante quem* ('limit before which') for the change. For innovations that are not yet reflected in a dated manuscript, on the other hand, we can determine the *terminus post quem* ('limit after which'), although one must consider that orthographic systems are often conservative, and thus reflect earlier stages. Strictly speaking, absolute dating, therefore, means dating a certain innovation relative to a certain time or time period.

Due to the above-mentioned issues with regard to the interpretation of written attestations of earlier periods, relative dating can be more reliable than absolute dating. In our opinion, the most promising procedure to develop a model of the history of a language is, therefore, to first establish a model of the relative chronology of the investigated changes, and to adjust the model to a timeline afterwards.

As Section 2.2 will demonstrate, relative dating of language changes, unlike absolute dating, is independent of written attestation. Regardless of whether the investigated changes occurred during the last century or the last millennium, if it is possible to infer a chronological relationship between them, they can be dated relatively to each other.

2.2 Establishing the relative chronology of sound changes

The main method for determining relative chronology is logical inference (e.g., Fortson, 2013).[2] In the following, we demonstrate how this method is applied to actual language data by means of examples from the two studies that provide the basis for our digital tool. Sound changes are the language changes which show the highest degree of regularity, and therefore these methods work best with sound changes. At this stage, it would not be possible to develop models such as those in Holzer (2007, 2011); Wandl

---

[2] For more comprehensive descriptions of reconstruction methods see, e.g., Anttila (1989); Hock (1991).

(2011) and Wandl (2020) based on morphological or syntactic innovations with the methods presented here because it is much more difficult to determine chronological relationships between them. Therefore, it seems most reasonable to start developing a model of the relative order of language changes with sound changes and then to proceed enriching it with data from other domains such as morphology, morphosyntax, syntax, etc.

There are four relationships that can be determined between sound changes: *Feeding*, *Bleeding*, *Counterfeeding* and *Counterbleeding*. In linguistics, these terms are better known with regard to rule ordering (e.g., Halle, 1962), but they are also applied diachronically (e.g., Chen, 1976; Hock, 1991, pp. 42-43; and especially Holzer, 2001, 2007). With regard to language changes, they can be defined as follows:

- *Feeding*: Change A creates inputs for change B, therefore A has to be dated before B.
- *Bleeding*: Change A withdraws inputs from change B, therefore A has to be dated before B.
- *Counterfeeding*: Change A would have created inputs for change B if it had occurred before B, but since it has not, B has to be dated before A.
- *Counterbleeding*: Change A would have withdrawn inputs from change B if it had occurred before B, but since it has not, B has to be dated before A.

In what follows, we present an example for a feeding and for a bleeding relationship from Holzer (2001) or Wandl (2011) (see there for information about the discussed changes and datings as well as references to the secondary literature) thereby making clear the process of inference.

Feeding Two changes are related to each other through a feeding-relationship if one of them creates inputs for the other. Consider the cognates Old Church Slavic *cěna* and Lithuanian *kainà* 'both: price, value'. The comparison of these words with further cognates in other Indo-European languages shows that Lithuanian both with regard to the word-initial consonant and the word-medial vocalism reflects a more archaic state than the Slavic word. Hence, we can posit a change A **k > *c* and a change B **ai > *ě* for a pre-stage of Slavic. However, the change of **k > *c* did not occur in any position (e.g., Old Church Slavic $kak\mathfrak{z}$ ~ Lithuanian *kõks* 'what (kind of)'), but is limited to the position before the diphthong *ai*. Now, considering that the change of **k* to **c* presents a typical palatalization process that is usually induced by an immediately following front vowel (Bhat, 1978; Bateman, 2011), we can infer the relative chronology of the two changes. The fact that the word-initial consonant **k* was followed by a front vowel only after change B had taken place – **ě* is a front vowel while *a* in *ai* is not – shows that change B must have occurred before change A. Change B created inputs for change A, hence the former change "fed" the latter. As Table 1 shows, only reconstruction a. gives the correct result while the order in b. does not account for the palatalization in Old Church Slavic *cěna*.

Bleeding Two changes are related to each other through a bleeding-relationship if one of them withdraws inputs from the other. This is the case, for example, with the rise of secondary palatalization (marked with superscript *j*) before front vowels and the change of **e* to **o* before hard consonants in Russian. Examples such as Russian *den'*

Table 1 Relative chronology of the Common Slavic palatalization *k >*c (A) and the monophthongization *aj >*ě (B)

a.	Proto-Slavic	B		A				Old Church Slavic
	* kajnā	>	*kěnā	>	*cěnā	>		cěna
b.	Proto-Slavic	A		B				Old Church Slavic
	*kajnā	=	*kajnā	>	*kěnā	>		**kěna

[dʲenʲ] 'day' show that the rise of secondary palatalization withdrew inputs from the change of *e to *o. The word final nasal *n was originally followed by the front vowel *ь which softened (palatalized) preceding consonants. For this reason, the change *e > *o did not apply in *den'*. It was "bled" by the rise of secondary palatalization. In Russian *lën* [lʲon] 'linen', on the other hand, *n was originally followed by the back vowel *ъ and hence, was not palatalized which is why *e changed to *o in the root syllable.

Apart from inherited word forms, relative chronology can also be derived from loanwords. Analyzing loanword strata in many cases reveals information about the linguistic situation at the time of the borrowing. This may include the presence or absence of certain innovations which, as a consequence, may allow for conclusions about their chronology. Moreover, two changes can be dated relatively to each other if one of their two possible sequences can be proven to be unlikely because it goes against our knowledge about language change. Datings of the latter kind may be considered less reliable since they depend on current knowledge. The fact that a certain change has not yet been observed in the languages of the world does not render it impossible.

In general, while the methods just described may be considered theory-neutral in the sense that they do not require the assumption of specific theoretical framework apart from the assumption that some languages are related to each other,[3] the interpretation of certain facts remains subjective. This concerns cases which depend on the researcher's intuition (see above) or cases where data are inconclusive or insecure. Different interpretations are the reason why different reconstructions exist. In this regard, it is important to be aware that any proposed reconstruction constitutes a model. Even if we consider all available data, we can only approximate reality. For example, at the current stage, it is not possible to determine relationships between all the sound changes of a language. As a consequence, in many places changes can be arranged in several different orders which may all be correct.

2.3 Verifying models of relative chronology

The models of Croatian and Russian sound changes used for developing the RelChron-Vis application incidentally both include 71 changes (segmental and suprasegmental). In the case of the Croatian model, it was possible to determine 225 relationships be-

---

[3] Cf., for instance, Chen (1976, p. 220): "It should be emphasized that the method of inferring the temporal relationships among linguistic processes is not tied to any particular theoretical model (…)"

tween individual changes, and in the case of Russian, 193. Although some of these relative datings concern identical changes – two changes may, for example, be dated relative to each other both because of a feeding-relationship and loanword data –, the number of involved changes and datings makes it difficult to verify the models. To do so, Holzer (2007) applied the sound changes in the order of his model to reconstruct the entire history of a significant number of word forms starting from Proto-Slavic (forward reconstruction) (cf. also Holzer, 2011; Wandl, 2011; 2020).

An example reconstruction is given in Eq. 1. The numbers following the greater-than signs refer to the corresponding sound changes as discussed in Wandl (2011).

$$\text{Proto-Slavic } *wa\underline{r}\dot{n}\bar{a}\dot{} > 12 \ *wa\underline{r}\dot{n}\bar{a} > 22 \ *wa\underline{r}a\underline{n}\bar{a} > 23 \ *wo\underline{r}o\underline{n}\bar{a} > 28$$
$$*wo\underline{r}o\underline{n}a > 44 \ *wo\underline{r}\hat{o}na > 55 \ *vo\underline{r}\hat{o}na > 67 \ *v\Lambda\underline{r}\hat{o}n\vartheta > 71 \text{ Russian}$$
$$v\Lambda\underline{r}o\underline{n}\vartheta < \text{воро́на} > \text{'crow'}$$
$$(1)$$

Although painstaking, this procedure has a number of advantages. First, it helps the researcher detect mistakes and contradictions in the model, and therefore improve it. Second, the reconstructed examples serve as proof that the proposed model yields viable results in the sense that it allows to derive attested forms from their ancestors. Third, any attempt to reconstruct the entire history of a word form can be considered a test for the model. The more tests of this kind the model 'survives', then, the 'stronger' it is (cf. Popper, 2002). For these reasons, Holzer (2007, 2011) and following his example Wandl (2011, 2020) gather these reconstructed forms in an appendix.

Holzer (2007) refers to his model of Croatian sound changes as a computing machine ('Rechenapparat'). Indeed, models of this kind can be converted into a computer program by simply transforming the formulation of the sound changes into computer-readable code (cf., e.g., Sims-Williams, 2018; Marr & Mortensen, 2020; Weinberger, 2021). Automating these reconstructions could potentially accelerate the verification of a model of the relative chronology. However, it does not replace the reconstruction process, at this stage.

2.4 Representing models of relative chronology

By determining all relative datings that can be inferred by means of the above-described methods, a model of the relative chronology of the changes of a language can be established. Surprisingly, models of this kind are still rare. To our knowledge, the only such contribution published as a book is Holzer (2007).[4] In this study, the author discusses the sound changes which occurred between Proto-Slavic and Croatian strictly according to the order resulting from the relationships determined between them. Unfortunately, unlike in less-extensive works Holzer (2005, 2009, 2007) only occasionally provides the reasons for the proposed order of changes, but refers to

---

[4] This is not to say that there are no other works which discuss sound changes in a specific order. (cf., e.g., for Slavic Carlton, 1991; Shevelov, 1964). However, theses publications draw from relative-chronology data to a much lesser extent.

previous studies. Understanding the reasons behind the ordering, therefore, requires consultation of several additional works.

Following the example of Holzer (2005, 2009); Wandl (2011, 2020) explicitly states all relative datings that he was able to determine. While this renders the placement of each change in the model transparent, it also reveals limits in the representation of complex models of the relative chronology in traditional formats such as books or articles. As stated above, Wandl (2011) contains 71 changes which are dated relatively to each other in 193 cases. Naturally, the datings are spread out across the entire publication which makes it difficult to keep track of all relationships between the included changes. Therefore, it is often not clear which consequences different orderings have for definitions or datings of other changes. Moreover, in traditional formats, the model cannot be presented in its entirety which makes it difficult to compare competing models on a larger scale. Due to its structure, the model can, however, be digitized, which provides a remedy for many of the above-stated difficulties.

3 Digitizing models of relative chronology

Bondy and Murty (2008) begin their introduction to graph theory with the following statement:

> "Many real-world situations can conveniently be described by means of a diagram consisting of a set of points together with lines joining certain pairs of these points. For example, [...] the points might be communication centres, with lines representing communication links."

If we replace "communication centres" with "language changes" and "communication links" by "chronological relationships", we can visualize the relative chronology of language changes with a graph. This may at first not seem like a big advantage, considering that the models we intend to represent are quite complex. They contain numerous sound changes and relationships joining them which makes it difficult to comprehend the resulting diagram. However, diagrams can easily be represented digitally, and as digital interactive graphs, complex models can be made accessible and equipped with additional features that are useful for research and teaching (cf. Section 4).

Bubenhofer identifies five main types of visualization that are typically used in linguistics: lists, scores (as in sheet music), maps, vectors and graphs (Bubenhofer, 2020, 134). Even though the visualizations that Bubenhofer reviews effectively communicate a variety of linguistic information, the vast majority of them are static and not interactive. Moreover, while there are some applications that present interactive plots of historical language data (e.g., Schlüter & Vetter, 2020), there are not many such interactive visualizations in historical linguistics in general. As of the writing of this article, the authors have not identified any interactive visualizations for the history of sound changes like the RelChronVis tool offers.

In this section, we describe how we developed an interactive web application to represent the relative chronology of Russian and Croatian sound changes by means

of interactive arc and chord diagrams. We begin by describing our input data in Section 3.1. Section 3.2 discusses how the input data are used to create a digital representation, and Section 3.3 provides a description of our application. Section 3.4 explains how users can create a digital model from their own data.

3.1 Input data

The first step to creating our web application was to transform the models from Holzer (2007) and Wandl (2011) into structured data that can be further processed computationally. We chose to represent each model with three CSV files, which will be described in detail in the following subsections. Note that text data have been translated from the German originals into English.

3.1.1 Sound changes

The first file we created presents a list of all sound changes of a model. The file contains one row per sound change and three columns. The first column assigns an ID to each sound change. The graphs in the web application will order the sound changes based on these IDs, and the IDs must start at 1 and increase by steps of 1. Because our web application is designed as a visualization, the order of the sound changes must be fixed at this stage. This approach has the advantage of eliminating the possibility that contradictory data (e.g., "A before B" and "B before A") are uploaded into the application. However, it also means that some decisions must be made as to the ordering of sound changes which cannot be dated relatively according to existing scholarship.

The second column of the file contains the name of the sound change, and the third column a short description. For the first version of the RelChronVis app, we have kept these descriptions as short as possible, providing only information that is necessary to understand the changes (input, output, context). Figure 1 gives the first six lines of our input files (they are identical in the Russian and Croatian model). Section 3.3 will provide more information about how these data appear in the application.

	A	B	C
1		name	description
2	1	Third or Progressive palatalization	The velar obstruents *k, *g, *x changed to *t′, *d′, *ś if they were immediately preceded by *i, *ī, or *j, or by sequences consisting of *i and a nasal *n.
3	2	Umlaut	Back vowels were fronted after the palatal consonants *j, *t′,*d′, and *ś: *u > *i, *ū > *ī, *a > *e, *ā > *ē.
4	3	Monophthongization	Tautosyllabic sequences of vowel and *w or *j were monophthongized: *aw, *āw, *ew, *ēw > *ō; *aj, *āj > *ē; *ej, *ēj > *ē > *ī.
5	4	Depalatalization	*ē was backened to *ā after the palatal consonants *j, *č, *dž, *š, *t′, *d′, *ś.
6	5	w-prothesis before *u-, *ū-	Word-initial *u-, *ū- received a prothetic glide *w: *u-, *ū- > *wu-, *wū-.
7	6	The change of *ē̦ to *ī	*ē̦, which had been introduced as a result of the monophthongization of *ej, changed to *ī.

Fig. 1 Extract of a CSV file listing sound changes

3.1.2 Relations

The second file we created contains what we here call "relations". These represent the relationships detailed in Section 2.2 (Feeding, Bleeding, etc.). The file lists one relation per row and has five columns: source, target, type, confidence and description (Fig. 2). The "source" and "target" fields contain the IDs of the sound changes which have a relationship. The field "type" contains an abbreviation for the type of relationship. Currently, there are the following types: Feeding (F), Bleeding (B), Counterfeeding (CF), Counterbleeding (CB), Attestation (A), Loanword (LW), Naturalness (N), Plausibility (P) and Simplicity (S). The first four terms have been discussed in Section 2.2. "Attestation (A)" refers to written attestation and "Loanword (LW)" to a relative dating based on loanword data. The term "Naturalness (N)" is used if a certain relative chronology can be ruled out or at least be considered highly unlikely because it can for some reason be considered unnatural. Similarly, "Plausibility (P)" refers to the plausibility of a certain ordering. In case of a dating marked with the label "Simplicity (S)", one rule ordering has been given preference over another since it can account for the attested facts in a simpler way than the reversed ordering. The web application will display information about these types in text form, and will color certain graph elements in a color scheme according to the relation type they represent.

The "confidence" field can be set to either TRUE or FALSE, indicating whether the relationship can be confidently established. If the user is insecure about a dating, they can indicate this by setting the field to FALSE. This choice will then influence the representation of the relationship in the application (see Section 3.3). Finally, the "description" field contains a string that the visualization will display when a user has selected a relation (see Section 3.3).

In the case of the data from Wandl (2011), there are 193 relations and thus 193 rows in this file (without the header row). Multiple relations are allowed between two sound changes, since relative datings can sometimes be established in more than one way (cf. Section 2.2).

With these two files, the web application can already build an interactive graph that visualizes the entire model. However, another CSV file is needed to model the

	A	B	C	D	E
1	source	target	type	confident	description
2	1	2	F	TRUE	Feeding: Palatal consonants resulting from the Third Palatalization (§ 1) triggered umlaut (§ 2), e.g. PSl. *suĺᵊnik
3	1	2	CF	TRUE	Counterfeeding: High front vowels resulting from umlaut (§ 2) did not trigger the Third Palatalization (§ 1), e.g.
4	1	9	F	TRUE	Feeding: The Third Palatalization (§ 1) created inputs for the change of *ś to *s' (§ 9), cf. PSl. *wįxa > *wįśa > *v
5	1	13	F	TRUE	Feeding: The Third Palatalization (§ 1) created inputs for the Assibilation of *t' and *d' to *c' and *dz', respecti
6	1	23	LW	TRUE	Loanword: The loanwords Fi. arnitsa 'thick linen' (cf. ORu or(ь)nica), Fi. katitsa 'fishing weir' (cf. gen. sg. ursl. *k
7	2	3	F	TRUE	Feeding: Umlaut (§ 2) of *aj to *ej created inputs for the monophthongization of *ej to *ī (§ 3), cf. PSl *jąjxu >
8	2	3	B	TRUE	Bleeding: Umlaut (§ 2) of *aj to *ej withdrew inputs from the monophthongization of *aj to *ē (§ 3), PSl *jąj
9	2	4	F	TRUE	Feeding: Umlaut (§ 2) created inputs for the depalatalization of *ē to *ā (§ 4), cf. PSl. *brā˙tijā˙ > *brā˙tijē˙ > *br
10	2	9	CB	FALSE	Counterbleeding: Assuming that Umlaut (§ 2) occurred only after palatal consonants, the change of *ś to *s' (§
11	2	17	B	FALSE	Bleeding: Umlaut (§ 2) of *ū to *ī withdrew inputs from the change of *ū to *ȳ (§ 17), cf. PSl. *sjų˙jā˙ > *sjį˙jē˙ >
12	2	24	CF	FALSE	Counterfeeding: Had j-Prothesis (§ 24) occurred before Umlaut (§ 2), the former change would have created inʲ
13	3	6	F	TRUE	Feeding: The monophthongization of *ej to *ē (§ 3) created inputs for the change of *ē to *ī (§ 6), cf. PSl. *zejⲙ
14	3	7	F	TRUE	Feeding: The monophthongization (§ 3) created inputs for the Second Palatalization (§ 7), cf. PSl. *kajnā˙ > *kēʳ
15	3	8	B	TRUE	Bleeding: The monophthongization of *aw, *āw, *ew, *ēw to *ō (§ 3) withdrew inputs from l-epenthesis (§ 8), ⱸ
16	3	26	F	TRUE	Feeding: The monophthongization of *aj, *āj to *ē (§ 3) created inputs for *ē to *ě (§ 26), cf. PSl, *sḛ˙weru > *s
17	3	27	LW	FALSE	Loanword: The toponym Lat. Maximus > rom.*Mājsęmu > Sl. *Mājsьmъ > Čak. Majsan was borrowed after the

Fig. 2 Extract of a CSV file listing relations

	A	B	C	D	E	F	G	H	I	J	K
				1	2	3	4	5	6	7	8
1	Ru.	Phonetic	PSl.								
2	бдеть	bd'ɛt'	*budeˇteʃ			*budeˇteˉ			*budeˇtiˉ		
3	беден	b'ɛd'ɪn	*bajdinu			*beđinu					
4	бела	b'ɪla	*beɾaˉ								
5	бело	b'ɪlo	*beɾa								
6	берёза	b'ɪr'ozə	*berˈzaˉ								
7	беремя	b'ɪr'em'ɪ	*berˈmeñ								
8	блоха	bləxa	*bluxaˉ								
9	блюдо	bl'udə	*bjawˈda		*bjewˈda	bjoˇda					*bljoˇda
10	блюсти	bl'us't'i	*bjawsteʃ		*bjewsteʃ	*bjošteˉ			*bjoštiˉ		*bljoštiˉ
11	боб	bop	*babu								
12	боба	bʌba	*babaˉ								
13	бобёр	bʌb'or	*babru								
14	бог	bok	*bagu								
15	боец	bʌjec	*bajiku	*bajitˊu	*bajitˊi						
16	бойцы	bʌjcy	*bajikuj	*bajitˊuj	*bajitˊij	*bajitˊiˉ					
17	болото	bʌlotə	*baɭˈta								

Fig. 3 Extract of a CSV file listing examples

information about the reconstructed examples (cf. Section 2.3) that are present in Holzer (2007); Wandl (2011), so that this information can be viewed in the application.

3.1.3 Examples

The last data file models reconstructions of example lexemes, one lexeme per row. The header row contains the names of the investigated language and its proto-language. In the case of the model from Wandl (2011), these usually are Russian ("Ru.") and Proto-Slavic ("PSl."), respectively. Column B contains a phonetic transcription. The web application will extract these fields and use them to display reconstructions (see Section 3.3). Each column after these has a header with a sound change ID, starting at 1 and counting up to 71, the final change of the model. Each row first contains the most recent and oldest form, and then contains a new form for each relevant step in the example reconstruction. The forms in the column with ID 3, for example, contain the forms that were reconstructed to exist after change 3 has taken place. In the example below (Fig. 3), this sound change is the first one to affect the lexemes in rows 2 and 3.

3.2 Technical details

The RelChronVis application takes the input files described in Section 3.1 as a basis and constructs the visualization from them. We have built the system in a modular way that allows input files to be changed or modified without needing to touch any of the other code.

The application is kept as simple as possible, using a Python server[5] and the Flask micro-framework,[6] and is deployed via a free hosting service, pythonanywhere.com.[7] Users can also download the source code from our online repository[8] and run the

---

[5] https://www.python.org

[6] https://flask.palletsprojects.com/en/2.3.x/

[7] https://www.pythonanywhere.com

[8] https://github.com/RelChron/relchronvis_1

server locally. The pages are naturally built with HTML and CSS, as well as Sass,[9] Jinja[10] and the Bootstrap[11] framework for layout and styling. The application makes use of the D3.js[12] library to draw an arc diagram and a chord diagram based on the data files. By using D3.js, the web application can easily link certain data (such as sound changes) to corresponding elements in the diagrams (such as circles), but also the rest of the interface (such as information cards about sound changes). In this way, a diagram can for example be drawn with a number of circles corresponding to the number of sound changes listed in the corresponding file (cf. Section 3.1.1). D3.js also offers tools to enable user interaction with the diagram, such as hovering and double clicking to access the associated data, but also zooming and panning.

In his discussion of different "coding cultures", Bubenhofer (2020, p. 131) points out that it is common practice to borrow snippets of code when using JavaScript, and specifically D3.js, because of the wealth of code that is publicly available. He also, however, points to a disadvantage of this borrowing, namely that when re-using code for a visualization, one adopts the assumptions and rules of that visualization. Using D3.js directly allowed us to sidestep this disadvantage. While we certainly took inspiration from existing snippets, the scale of our application meant that we had to largely write our own code to implement the desired features.

Another reason to create a web application for our visualization was that it is easy to distribute. It is deployed on the web, so it can be accessed by anyone with an internet connection, on any operating system, on Firefox, Safari and Chrome, without having to install or run anything. If downloaded and run locally, the application can even be used without an internet connection.

3.3 The RelChronVis web application

When visiting the landing page of the RelChronVis application, users can select between Russian, Croatian or custom data (cf. Section 3.4). After choosing a language, or uploading custom data, they will see an arc diagram, which is the first of two graphs that the application offers. This diagram arranges sound changes as a horizontal line of circles, and relations as arcs which connect the sound changes. We chose this diagram because it can organize changes and their relationships according to a timeline. In addition to the diagram, the application shows a sidebar to the left, where users can find general functions. The buttons at the top of this sidebar allow users to download the graph in its current state, and to hide or show the labels, which may be needed when users want to use the graph in a publication or presentation. If no sound change is selected (see below), all sound changes and relations will be highlighted in the downloaded image. Furthermore, the sidebar's filter button allows users to redraw the graph including only certain sound changes. This filtering function also allows restricting the relations to only incoming or outgoing ones, i.e., the changes that preceded or

---

[9] https://sass-lang.com

[10] https://jinja.palletsprojects.com/en/3.1.x/

[11] https://getbootstrap.com

[12] https://d3js.org

followed a selected change according to the model at hand. Moreover, relations that have been labeled unconfident (see Section 3.1.2) can be excluded from the diagram. While this reduces the data basis for the diagram, it renders the model more reliable (less secure datings are exluded).

Initially, all circles and arcs will be gray, but when users hover their mouse over a circle, that circle will change to a darker color, and so will all connected circles. The connected arcs will change to the color of their associated type (Feeding, Bleeding, etc., see Section 3.1.2), according to the color scheme. Arcs for datings which are not confident are rendered with a dashed line. Double-clicking a circle will freeze this highlighted state for that sound change, and double-clicking again will unfreeze it. When a sound change is frozen, an information card appears in the sidebar on the left part of the view. Additionally, in this state, users can click on colored arcs to show an information card about the type and reason of the relative dating, as well as an information card about the sound change that is connected via that relation. Figure 4 shows the web application in this state.

When a sound change is frozen, users can also view example reconstructions of lexemes which undergo the selected sound change taken from the Examples file (see Section 3.1.3). Clicking the "Show Examples" button on the top right will open a sidebar which contains suitable example lexemes in their present-day variety. Figure 5 demonstrates what this looks like. This sidebar contains a list that was filtered from the whole pool of examples based on the selected sound change. Upon clicking an example, the sidebar closes and the application writes the reconstruction of the selected lexeme into a box at the bottom of the view, like in Fig. 6. This shows the reconstruction by which the lexeme changed from its earliest form to its present-day form, presenting the selected sound change in bold font (cf. Fig. 6).

The RelChronVis application also offers an alternative visualization, the chord diagram. This diagram arranges sound changes in a circle as segments, and relations
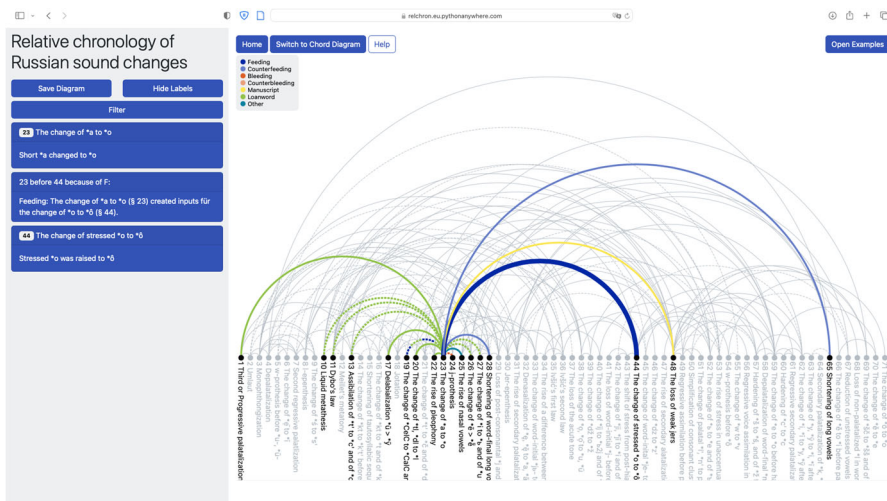
Fig. 4 Screenshot of the application's arc diagram while a sound change and relation is selected

Fig. 5 Screenshot of the application while browsing examples

as bands between those segments. Each relation is rendered as a separate band, even if they apply to the same two sound changes. This makes it easy to determine how many types of relationships have been established between two changes of interest. The more relations are connected to a circle segment, the longer it gets, which shows at a glance which sound changes are most "useful" in dating others. Moreover, the length of bands in the chord diagram is similar in all instances while that of the arcs in the arc diagram varies significantly. Therefore, the representation of relationships between two changes which are far away from each other chronologically requires
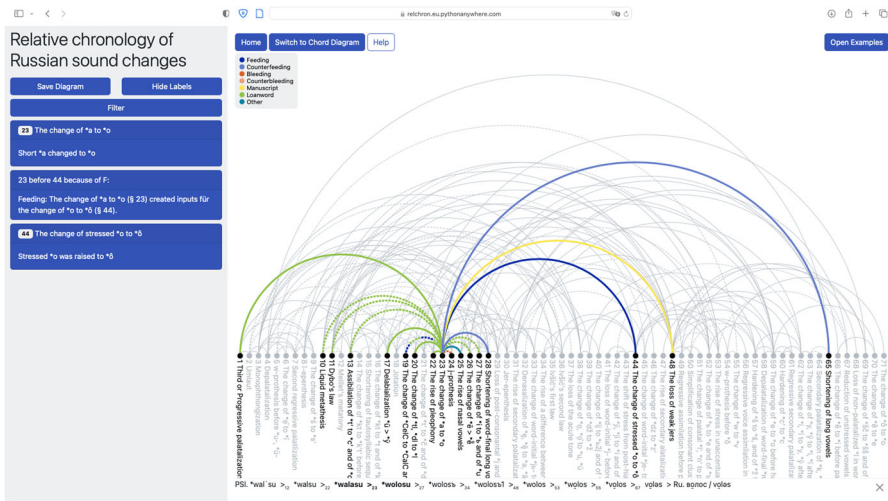
Fig. 6 Screenshot of the application after selecting an example
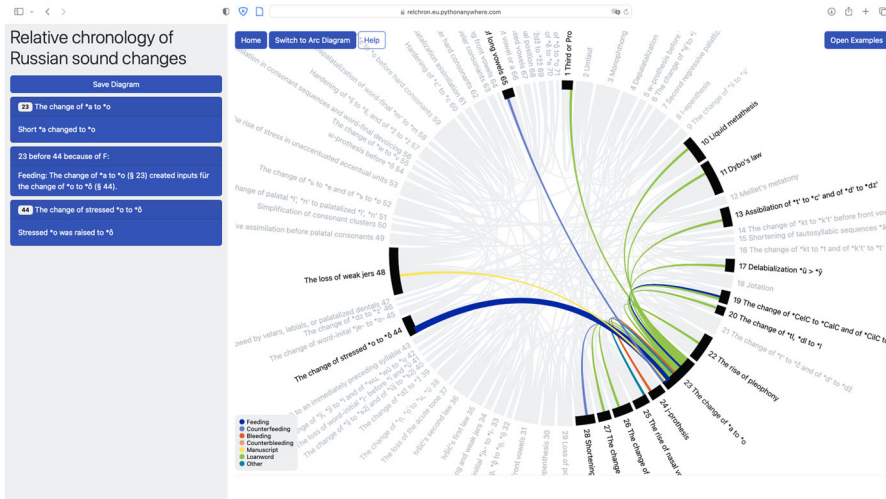
Fig. 7 Screenshot of the application's chord diagram

less space. Figure 7 shows what the chord diagram looks like when selecting the same sound change as in Fig. 4.

3.4 Custom data

The application furthermore allows users to upload custom data on an upload page that can be accessed from the landing page. Users can simply upload data files like those explained in Section 3.1, or they can download, modify and re-upload our data files, since they are publicly available. For example, it would be fairly trivial to add new relations, change their confidence values, or delete entire sound changes, without having any programming ability. Currently, the only requirement for the data is that it use our relation types (Section 3.1.2). This is because the relation types and color scheme are hard-coded. However, because our application is open source, more advanced users could easily adapt relation types and color scheme.

The file containing the example reconstructions is optional, because it is not strictly necessary to view the graph and the reconstructions may not exist yet. In this way, an infinite number of visualizations can be created for any kind of custom datasets, which will always display the entire model on a single screen. This can be beneficial to researchers, students, and laypeople alike, as the next section will explore.

4 Applications of the digital model

We believe that our model is primarily useful in two contexts: research and teaching. However, since the RelChronVis application is easy to interpret, it additionally

provides a resource for interested laypeople. Therefore, it contributes to the dissemination of linguistic findings to the public.

4.1 Research

According to Bubenhofer (2020, p. 6), diagrams are so useful to humans because they not only visualize data, but also emphasize the relationships between those data and thus allow us to derive new knowledge. We believe that this is very well exemplified by the RelChronVis application. It visualizes complex data in an accessible way, thereby allowing for new conclusions. One aspect that is often not considered systematically outside of studies dealing with specific problems of relative chronology are indirect chronological relationships between changes. For instance, if a change A can be dated before a change B and the change B can be dated before a change C, it follows that A must have occurred before C. Extracting the relevant information from historical grammars can be tedious since it is not made explicit. The RelChronVis application provides a possibility to retrieve information about sound changes and their place in the relative chronology by a simple mouse click. After selecting one sound change, all connected sound changes can be hovered with the mouse to view their relations, thus allowing inspection of individual indirect relationships. Including them into related research is, therefore, made much easier which potentially makes them more precise with regard to questions of chronology. The RelChronVis application helps users get a more detailed and thus clearer picture of the history of the investigated language. Figure 8 demonstrates this functionality.

Another advantage of the digital model is that it can be represented in its entirety. Of course, it could also be simply printed as a diagram in a book, but the complexity of the model would make it very difficult or even impossible to derive information without
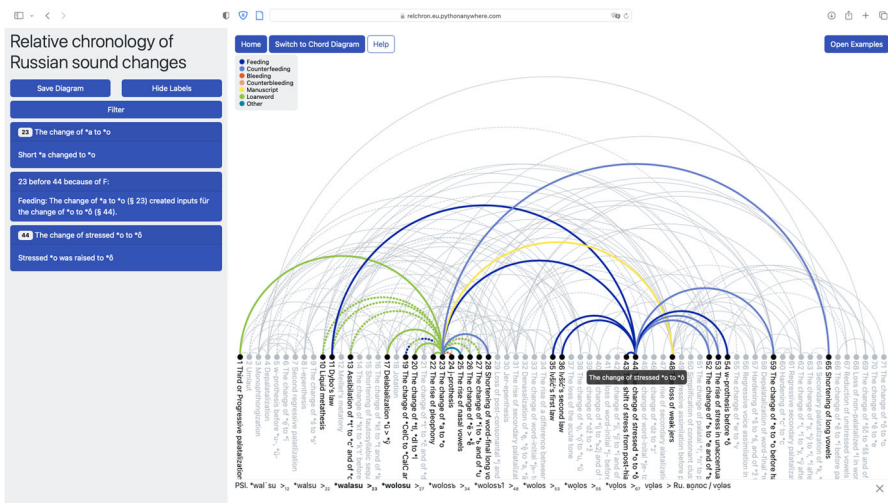

Fig. 8 Screenshot of the application's arc diagram while a primary and a secondary sound change are selected and both of their relations are highlighted

consulting its description which leads us back to the disadvantages of traditional publication formats. In the digital model, on the other hand, all information can be displayed without losing the overall view. This makes it, for example, easier to detect certain tendencies in the development of a language (cf., e.g., the so-called "law of open syllables in Slavic", see Carlton, 1991, p. 100). While the RelChronVis application does not replace the reconstruction of the relative chronology, it makes it easy to investigate whether certain changes occurred closely to each other on the timeline.

The fact that an entire model of the relative chronology of the changes of a language can be represented digitally, moreover, allows the researcher to compare competing models not only for particular changes, but also on a larger scale. As mentioned above (Section 2), ambiguous or inconclusive data lead researchers to different interpretations of certain changes. From special studies dealing with these problems in the context of a specific time frame, it is often not clear what the consequences of diverging interpretations may be for other changes that can be related to the changes at hand and consequently for the entire model. In the digital model, these differences can be made explicit, which makes the evaluation of advantages and disadvantages of competing reconstructions easier.

Related to the advantage just discussed is the fact that the digital model helps reveal shortcomings or deficiencies in the reconstruction. As noted in Section 2, the model remains incomplete because changes can usually be dated only relatively to several other changes and in some cases no relationship to any other change can be established. Consequently, the place the changes take in the relative chronology is to some degree arbitrary. While this may not be immediately apparent from publications such as Holzer (2007); Wandl (2011, 2020), it is noticeable right away from the digital model.

An asset of perhaps any digital representation is easy access to data. As described in Section 3, the current model contains information about all the included sound changes, relative datings as well as examples for the changes at hand. All this information can be displayed by mouse click whenever needed. This makes access to relevant data much easier and faster. Moreover, digital representations are open for adding additional data, which is an immense advantage in comparison with printed publications. Extension and revision of traditional formats require the publication of new editions which are not only time-consuming in preparation but may depend on the decision of editors or publishers. Digital models, on the other hand, can easily be revised and amended with additional data. This also concerns data that one would rarely find in one and the same publication (cf. Section 5).

4.2 Teaching

The model also provides a handy tool for teaching historical linguistics. When studying language changes from books, the temporal dimension of language evolution is often not immediately apparent. This is especially true for studies that simply juxtapose a proto-language with a daughter-language. One could say that the history of a language is presented somewhat ahistorically in these works. Likewise, the subsequent

descriptions of language changes in formulas, as it is common in historical grammars, may not convey the temporal depth of the development very well.

This is entirely different with graph representations such as the arc diagram. From looking at the diagram, it is immediately clear that the changes proceeded according to a certain timeline, even though it needs to be stressed that the order of changes is to some extent arbitrary. The temporal dimension becomes apparent even more clearly if the relative chronology is adjusted according to absolute datings. Students can, therefore, follow the development of a language, as represented by the included changes, through time. This can be done, on the one hand, by means of the different diagrams, and, on the other hand, by looking at the examples provided for every sound change (Section 3). By investigating the development of individual words, students can acquire a more realistic view of language change.

At the same time, the model can be used to teach individual changes and the relationships observable between them. Since all data are easily accessible, this can be done as part of self-study assignments. The students can be given exercises which they are to solve by means of the digital model. This concerns, for example, questions about why certain changes must have occurred before others, or about what a specific word must have looked like at a certain stage. Furthermore, the custom-built function allows the students to build their own model. They could be given a number of words and sound changes together with the task to find the order in which the changes must have occurred and to reconstruct the history of the given words according to the relative chronology established by them. After converting their results into a computer-readable format, the students can then upload them to the RelChronVis application to build their individual models. In this way, the results can be neatly presented as part of in-class presentations, or they can be downloaded and included into seminar papers or posters.

From the perspective of linguistics, assignments of this kind introduce the students to analytical approaches to language data, and teach them methods that are essential for doing historical linguistics. By converting their data into a computer readable format, they acquire a skill that is valuable far beyond linguistic research. Relevant on a more general level is, moreover, the training in logical thinking that the involved reconstruction methods require.

5 Outlook

As mentioned above (Section 4), one advantage of open-access resources is their extensibility. Therefore, in this section, we intend to present several possibilities for how our tool could be extended in the future. They mainly concern the inclusion of further data, the linkage with other resources, and the representation of the data.

As regards the already existing models, further information about absolute datings and about non-phonological changes can be amended. Information about the absolute datings largely concerns changes reflected in written records (cf. Section 2.1). By including these data, the model of the relative chronology can be at least in parts aligned to an absolute timeline. Another point that is worth exploring concerns the possibility of linking the included data to other online resources such as dictionaries

and databases. To make our tool useful for a larger group of researchers, we intend to add more languages to our inventory.

Apart from the already existing diagrams, we want to implement the option of representing the data as a network diagram. An advantage of this representation is that the data can be organized according to the density of the networks arising from the chronological relationships. Thus, it is not built upon the chronological order resulting from the relationships. Apart from highlighting which changes can be dated more often than others, network diagrams clearly show that the order of changes in the arc and chord diagrams is to some extent arbitrary.

Moreover, we want to develop diagrams that allow to represent models of the relative chronology of several related languages in one visualization. This could help researchers better understand the split of the included languages from their proto-language. The information retained from this diagram is, therefore, relevant for phylogenetic modelling.

As for the limitations in dating language changes, the possibility to capture large-scale models of relative chronology in machine-readable form has the potential to contribute to our overall understanding of how languages change. While it is unlikely that we will ever by able to reconstruct language history in all its details, developing comprehensive data bases about the order of language changes is an important step in this direction since it allows us to conduct research by means of modern computational tools. Applying such tools to a large number of language-history models has the potential to uncover hitherto unnoticed patterns of language change which may then help us to fill in some gaps in the reconstructions of individual languages.

A larger project that we intend to realize in the future concerns the inclusion of data from other scientific disciplines. For instance, the analysis of loanwords takes an important role in the reconstruction of the history of Slavic languages. Among them are toponyms which have been borrowed from Slavic into different languages, and in some areas even survived the Slavic languages. Thus, in some parts of Austria, Germany or Greece that are not inhabited by Slavic speaking people anymore, we find Slavic toponyms as remnants of their earlier presence. These toponyms provide valuable data for the reconstruction of the Slavic language at the time when they were borrowed. In some instances, they may or may not reflect a certain sound change which makes it possible to link the sound change at hand to a certain period in the relative chronology. Vice versa, the sound change at hand can be linked to a certain place or area. By doing so, the model gains a geographical dimension. The model of the relative chronology can, therefore, be amended with maps that provide the user with information about the areal diffusion of the discussed changes.

6 Conclusions

In this paper, we have introduced the RelChronVis web application. It draws information about models of the relative chronology of Croatian and Russian sound changes from Holzer (2007) and Wandl (2011) respectively, and displays each model in a single interactive arc or chord diagram. In addition to the diagrams, the application provides textual information about sound changes, reasons for relative datings and

reconstructions of example lexemes. Furthermore, we publish the source code and data files in a repository, and we allow users to upload and visualize their own data (cf. Section 3.4).

Our novel visualization tool aims to be useful for research and teaching (cf. Section 4). Even though it is relatively simple in technical terms, it provides a tool for making information from many print sources more accessible and easily digestible. For this reason, it presents a way to publish models of relative chronology in the future. However, our application is not intended to replace traditional publication methods. For example, it is not designed to represent knowledge in consecutive chapters that build on each other as is the case, for instance, with many handbooks. Rather, the two formats complement and contribute to each other. For example, models that are newly published in print can be digitized and added to the model in the same way that we have digitized the models from Holzer (2007) and Wandl (2011) (cf. Section 3), and the web application can generate static images to be used in print publications.

In creating the RelChronVis application, we used pre-existing snippets as a basis but have attempted to sidestep assumptions of the "coding cultures" of various programming languages by coding the vast majority of the application ourselves. Moreover, our project is open source, which means that it can continuously be amended, by ourselves or other contributors. In this way, potential limitations can be circumvented, mistakes can be corrected, new data can be added, and new visualization types can be implemented by any researcher, student, or the general public. With a growing number of computational tools, it may also become easier and faster to create new input data that can be visualized with our web application. We hope that our project can be a part of a lively new ecosystem in this way, and we hope that it will inspire others to develop similar models.

Declarations

Competing interests The authors declare no competing interests.

References

Andersen, H. (2003). Introduction. In H. Andersen (Ed.), *Language contacts in prehistory: studies in stratigraphy* (pp. 1–10). John Benjamins.

Anttila, R. (1989). *Historical and comparative linguistics* (Second Revised). John Benjamins.

Bateman, N. (2011). On the typology of palatalization. *Language and Linguistics Compass, 5*(8), 588–602.

Bhat, D. N. S. (1978). A general study of patalization. In J. H. Greenberg (Ed.), *Universals of human language. vol. 2: Phonology* (pp. 47–92). Stanford University Press.

Bondy, J. A., & Murty, U. S. R. (2008). *Graph theory*. Graduate Texts in Mathematics; 244. Springer.

Bubenhofer, N. (2020). *Visuelle Linguistik: Zur Genese, Funktion und Kategorisierung von Diagrammen in der Sprachwissenschaft*. de Gruyter.

Carlton, T. R. (1991). *Introduction to the phonological history of the slavic languages*. Slavica.

Chen, M. Y. (1976). Three methods of reconstruction. *Journal of Linguistics, 12*(2), 209–258.

Fortson, B. (2013). Relative chronology. In G. K. Giannakis, V. Bubenik, E. Crespo, C. Golston, A. Lianeri, S. Luraghi, & S. Matthaios (Eds.), *Encyclopedia of ancient greek language and linguistics*. Brill. https://doi.org/10.1163/2214-448X_eagll_COM_00000306. Accessed 24 July 2023

Halle, M. (1962). Phonology in generative grammar. *Word, 18*(1–3), 54–72.

Hock, H. H. (1991). *Principles of historical linguistics* (Second revised and updated edition). de Gruyter.

Holzer, G. (2001). Zur Lautgeschichte des baltisch-slavischen Areals. *Wiener slavistisches Jahrbuch, 47*, 33–50.

Holzer, G. (2005). Zur relativen Datierung prosodischer Prozesse im Gemeinslavischen und frühen Kroatischen. *Wiener slavistisches Jahrbuch, 51*, 31–71.

Holzer, G. (2007). *Historische Grammatik des Kroatischen. Einleitung und Lautgeschichte der Standardsprache*. Peter Lang.

Holzer, G. (2009). Urslavische Prosodie. *Wiener slavistisches Jahrbuch*, 151–178.

Holzer, G. (2011). *Glasovni razvoj hrvatska jezika*. Institut za hrvatski jezik i jezikoslovlje.

Marr, C., & Mortensen, D. R. (2020). Computerized forward reconstruction for analysis in diachronic phonology, and Latin to French reflex prediction. *Proceedings of LT4HALA 2020 - 1st Workshop on Language Technologies for Historical and Ancient Languages*, 28–36

O'Brien, M. J., & Lyman, R. L. (2002). *Seriation, stratigraphy, and index fossils. The backbone of archaeological dating*. Kluwer Academic Publishers

Popper, K. (2002). *The logic of scientific discovery*. Routledge.

Schlüter, J., & Vetter, F. (2020). An interactive visualization of Google Books Ngrams with R and Shiny: Exploring a(n) historical increase in onset strength in a(n) huge database. *Journal of Data Mining & Digital Humanities, Special issue on Visualisations in Historical Linguistics*. https://doi.org/10.46298/jdmdh.5582

Shevelov, G. Y. (1964). *A prehistory of slavic*. Winter.

Sims-Williams, P. (2018). Mechanising historical phonology. *Transactions of the Philological Society, 116*(3)

Wandl, F. (2011). *Diachrone Lautlehre des Russischen. Ein Modell des Lautwandels und seiner relativen Chronologie* (Master's thesis). University of Vienna.

Wandl, F. (2020). *Vom Urindogermanischen zum Russischen: Eine Studie zum Lautwandel und seiner relativen Chronologie* (Doctoral dissertation). University of Zurich.

Weinberger, H. (2021). Vom Urslavischen zu den slavischen Einzelsprachen. Zur computergestützten Modellierung von Lautgesetzen. In A. Z. Jürgen Fuchsbauer Wolfgang Stadler (Ed.), *Kulturen verbinden. Festband anlässlich des 50-jährigen Bestehens der Slawistik an der Universität Innsbruck* (pp. 55-73). University Press