sciendo
DOI 10.2478/jazcas-2024-0047

# IN SEARCH OF PATTERNS OF HISTORICAL LANGUAGE VARIATION AND USER INTERACTION (OR: WHO USED WHAT LINGUISTIC FEATURES WITH WHOM, WHEN, WHERE, WHY, AND HOW?)

MARK RICHARD LAUERSDORF
Department of Linguistics, College of Arts and Sciences, University of Kentucky, USA

LAUERSDORF, Mark Richard: In search of patterns of historical language variation and user interaction (or: Who used what linguistic features with whom, when, where, why and how?). Jazykovedný časopis (Journal of Linguistics), 2024, Vol. 75, No. 3, pp. 330 – 347.

**Abstract:** This essay presents a set of interconnected thoughts, considerations, and conceptualizations that address what it is that we are talking about when we discuss language variation and the interaction of language users in historical contexts, and how we research patterns of historical language variation and user interaction.
**Keywords**: data-driven, inductive analysis, language and identity, historical language-documentation fieldwork, historical ethnographic fieldwork, tasks of historical sociolinguistics.

## 1. POSITION STATEMENT 1

The arguments presented in this essay grow out of:

• work in historical situations that involve a high number of (closely-related) language varieties in contact in settings that exhibit a multitude of geophysical, socio-cultural, political, and psychological borders, with ever changing socio-cultural, socio-political, socio-economic contexts, and often no prominent "roofing" standard language or prestige variety, and

• a strong preference for methods that include digital corpora, analytic data visualization, and statistical data analysis, in overarchingly data-driven, inductive investigations (a methodological preference that will figure prominently throughout this discussion).

In other words, the arguments presented here seek to make a case for the necessity of **data-driven corpus-based quantitative investigation of language variation in its complex socio-political and socio-cultural environments, using statistical and visualization methods of data analysis to identify salient patterns.**

One might be tempted to react to this position statement with a dismissive "so what?", wondering whether there is anything new to be said here. After all, wasn't

Barbara Horvath already doing exactly this kind of work in the 1980s in her work in Sydney, Australia, where the rationale behind the methods she employed in that research program was stated as follows (with the parallels to my above-stated position inserted in *[square-bracketed italics]*)?

> Quantitative analyses of large data sets make use of both linguistic and sociological categories in sociolinguistic studies. *[= quantitative; language in its social environment]* [...] The familiar problem of grouping speakers by such sociological characteristics prior to quantitative analysis is addressed *[= data-driven, not category-driven]* and an alternative solution – principal components analysis – is suggested. Principal components analysis is used here as a heuristic for grouping speakers solely on the basis of linguistic behaviour... *[= data-driven; statistical and visualization data analysis]* In addition, by naming the principal components, the major linguistic and social dimensions of the variation in the data can be identified. *[= identify salient linguistic and social patterns]* (Horvath – Sankoff 1987, p. 179)

There are, for our discussion here, two important observations in this regard:
**Observation 1:** Despite the earlier pioneering work of scholars like Horvath, the data-driven, inductive investigation of clusters or constellations of linguistic and social features and their patterning has only recently been more widely taken up (often under the label of "coherence"). See, for example, Meyerhoff and Klaere (2017), the collection of studies in Hinskens and Guy (2016), or the more recent studies in Beaman and Guy (2022), where Cerruti and Vietti (2022) note that:

> there is growing interest in the empirical characterization of aggregates of linguistic variables (Guy and Hinskens 2016; Ghyselen and De Vogelaer 2018; Vietti 2019). [In note 3 here Cerruti and Vietti mention that: "Seminal studies include Guy (1980), Horvath and Sankoff (1987), and Trumper and Maddalon (1990)." – MRL] The co-occurrence of linguistic features in relation to social factors has been framed within the concept of coherence. Coherence at the level of empirical observation has been mainly understood as "external" or as co-occurrence between linguistic facts in relation to socio-communicative contexts (Cerruti – Vietti 2022, p. 263).

Thus, the position that I am promoting here, though gaining in application, is not yet a dominant research paradigm in the study of language variation[1], and this

[1] There is, however, steady refinement of the concepts and the methods, including in the direction of studying much larger datasets (see for example Hua et al. 2021), which is an important methodological step especially as regards the further arguments in this essay concerning the need to "use all the data".

essay is intended to lend another voice to the call for the data-driven, inductive investigation of clusters or constellations of linguistic and social features and their patterning.

**Observation 2:** Such work has, for a variety of reasons, been undertaken nearly exclusively in modern synchronic situations of language variation, whereas the discussion here focuses on language variation and the interaction of language users *in historical contexts*.

Thus, this essay is expressly intended to be a call for *extending to the field of historical sociolinguistics* the growing application (in present-day synchronic contexts) of data-driven, inductive investigation of clusters or constellations of linguistic and social features and their patterning.

## 2. POSITION STATEMENT 2

In promoting this type of data-driven, inductive investigation of clusters or constellations of linguistic and social features and their patterning specifically *in historical contexts*, another position needs to be propounded regarding the focus of research into historical language variation and user interaction. The traditional research paradigm in the investigation of language history was for a long time:

language history = standard language history = history of national identity

However, with the increasing growth and development of socio-historical emphases in historical linguistic work (i.e., the steady rise of the field of *historical sociolinguistics*), there has been strong movement away from this more narrow focus on the history of the standard language and national (nation-state) identity, and an incorporation of some basic tenets from social history, a field that arose in the mid 20th century bringing new perspectives to the study of the past that were no longer about writing "biographies of great men", but rather about relating the experiences of ordinary people. Indeed, Tilly (1985) identifies the "tasks of social history" as:

1) "documenting large structural changes";
2) "reconstructing the experiences of ordinary people in the course of those changes";
3) "connecting the two" (Tilly 1985, p. 31).

A historical sociolinguist's paraphrase of Tilly's tasks, tailoring them as tasks of the social history *of language* (i.e., "tasks of historical sociolinguistics"), might look something like this:

1) documenting structural language variation and change;

2) reconstructing the experiences and interactions of ordinary people in connection with that variation and change; and
3) connecting the two (a key element for the discussion in this essay).

And this, in turn, suggests a recasting of the traditional research paradigm in the investigation of language history that was listed above, leading to a rendering more like this:

| | | | |
|---|---|---|---|
| language history = | history of variation and change in all varieties of language in play in user interactions in a given socio-cultural context | = | a multifaceted approach to questions of history of language and identity |

This view of the investigation of language history places the focus squarely on describing the complexities of the variable and interactional historical contexts, and the individual actors in those contexts, that produce present-day states (that are themselves filled with variation and interaction among individual actors), and it strengthens the case for the research program proposed at the outset of this essay:

> **the necessity of data-driven corpus-based quantitative *sociolinguistic* investigation of *historical* language variation *and user interaction* in *their* complex socio-political and socio-cultural environments, using statistical and visualization methods of data analysis to identify *and correlate* salient patterns *in the linguistic <u>and</u> socio-historical data***

with the additional elements in italics now sharpening the program's focus on historical investigation not just of language but also user interaction, and on the interconnection between the linguistic and social aspects and the patterning of both. Paraphrasing Joshua Fishman's (1965) classic article title, this research program could be described as the investigation of: Who used what linguistic features with whom, when, where, why, and how?

## 3. WHO USED WHAT LINGUISTIC FEATURES WITH WHOM, WHEN, WHERE, WHY, AND HOW?

Seeking, in his 1965 article, to formalize the description of language choice in "within-group (or intragroup) multilingualism", Fishman states: "habitual language choice is far from being a random matter of momentary inclination, even under those circumstances when it could very well function as such from a purely probabilistic point of view" (Fishman 1965, p. 67). In examining Fishman's statement beyond the context in which he was working, Fishman's "multilingualism" can be re-stated as the existence of multiple linguistic codes in a single setting, where “multiple linguistic codes” is understood to include multiple language varieties of any sort (“multi-varietalism”); and the existence of multiple linguistic codes / multi-varietalism in a single setting means that there are multiple linguistic variants that can be accessed by a language user in that setting to fill a given slot in language structure. If, in that setting, the choice of linguistic code by a language user is (in Fishman’s terms) “far from being a random matter of momentary inclination”, then the selection of any given linguistic variant is (again in Fishman’s terms) also not “a random matter of momentary inclination”.

This chain of argumentation is meant to draw attention to three important notions: (1) that Fishman’s conclusions are as applicable to settings involving closely-related language varieties as they are to settings involving non-closely related language varieties[2], (2) that “language choice” in any of these settings ultimately refers to the language users’ selection of features from those available in the language varieties that co-exist in these multi-varietal settings, and (3) that the users’ selection of features is “far from being a random matter of momentary inclination”. Thus, the discussion that follows can be considered to be broadly applicable to many different contexts where different types of language varieties co-exist in a single societal setting; and the investigation of the use of features in that setting can be considered an exercise in discovering meaningful (“non-random”) patterns/clusters of language features as the language users interact with one another in that setting.

This reformulation of Fishman could, once again, be considered a partial description of the type of work that is currently gaining ground among those sociolinguists working on questions of “coherence” in present-day, synchronic contexts, and it is this type of work that I am promoting here for use in historical and diachronic contexts as well.

## 4. THEORETICAL AND METHODOLOGICAL CONSIDERATIONS

Taking a cue from those colleagues working on coherence, what if, in historical linguistics work, instead of investigating a pre-determined linguistic feature (or several pre-determined features) and pre-determined social characteristics or categories of speakers, we address the question of “Who used what linguistic features with whom, when, where, why, and how?” using a data-driven, inductive approach? Instead of *a priori* deciding specifically which linguistic feature(s) and social characteristics to investigate, what if we simply analyze the entire dataset to discover the linguistic patterns present therein? In other words, what if we approach the study of language history as an exercise in *historical language-documentation fieldwork*,

[2] Encompassing the entire spectrum of what is sometimes referred to as “internal multilingualism” and “external multilingualism” (“innere Mehrsprachigkeit” and “äußere Mehrsprachkigkeit”) (see Wandruszka 1975, 1979).

seeking to provide a full description of the patterns in the texts in front of us, treating each text/writer as an individual fieldwork participant who is providing language data and socio-cultural notes?

This approach of studying the full set of linguistic patterns available to us in a given historical context, in a historical language-documentation fieldwork sort of way, finally engages the theoretical stance that Hermann Paul espoused almost 140 years ago:

> The true object of philological study is rather the entire sum of the products of the linguistic activity of the entire sum of individuals in their reciprocal relations. All the groups of sound ever spoken, heard, or represented, with the associated ideas, whose symbols they were; all the numerous relations entered into by the elements of speech in the minds of individuals – all these belong to the history of language, and must, properly speaking, all be thoroughly apprehended to render a full apprehension of its development a possibility[3] (Paul 1891, pp. 2–3).

Of course, Paul himself considered such a task an impossibility[4], but then immediately stated that the impossible nature of the task doesn't mean that we shouldn't aspire to it:

> It is good to state the ideal aim of a science in all its bareness of outline. By so doing we become aware of the gulf between our powers and our possibilities. We learn that we must in many questions content ourselves with an avowal of ignorance; and that super-acuteness, which imagines that it can explain the most complicated historical developments by a few ingenious aperçus, is humbled. But it is for us an inevitable necessity to get a general idea of the play of the forces at work in this huge complex – forces which we must always keep before our eyes, if we would endeavour

[3] In a more recent translation: "The real object of investigation for the linguist consists of the entire body of speech events in all individuals and their influence on one another. Indeed, the history of a language includes all the sound combinations ever spoken, heard, or imagined by an individual and the associated representations of which they were symbols, as well as all the manifold relationships that the elements of a language entered into in the minds of individuals. In theory all these facts would have to be known to us in order to reach a complete understanding of language change" (Auer – Murray 2015, p. 48). The original German text: "Das wahre object für den sprachforscher sind vielmehr sämmtliche äusserungen der sprechtätigkeit an sämmtlichen individuen in ihrer wechselwirkung auf einander. Alle lautcomplexe, die irgend ein einzelner je gesprochen, gehört oder vorgestellt hat mit den damit associierten vorstellungen, deren symbole sie gewesen sind, alle die mannigfachen beziehungen, welche die sprachelemente in den seelen der einzelnen eingegangen sind, fallen in die sprachgeschichte, müssten eigentlich alle bekannt sein, um ein vollständiges verständniss der entwicklung zu ermöglichen" (Paul 1886, p. 22).

[4] "It need hardly be said that to solve such a problem is an impossibility" (Paul 1891, p. 3). "No one need object that there is no point in setting up a task that is so obviously impossible to fulfill" (Auer – Murray 2015, p. 48). "Man halte mir nicht entgegen, dass es unnütz sei eine aufgabe hinzustellen, deren unlösbarkeit auf er hand liegt" (Paul 1886, p. 22).

> to arrange correctly the few scanty fragments which we can really claim to possess out of it.[5] (Paul 1891, p. 3)

Hermann Paul is, of course, not the only one who has, in the past, proposed this idea that, in order to even hope to understand language variation and change in historical periods, we must strive to consider the full set of language evidence and language use by individuals. Izidor Kotulič, about 65 years ago, stated the same "ideal aim" as Paul, specifically regarding the social history of Slovak, which he argued:

> ...can only be more comprehensively answered after a thorough and broadly organized investigation of linguistic monuments with the initial goal of a perfect knowledge of the language and documents of the period. The theoretical investigation of the language of the Slovak people must be based not only on a reliable methodological foundation, but also on the relatively complete historical linguistic material that is lying scattered in domestic and foreign archives (Budapest and others). For this it will be necessary to undertake a thorough archival investigation and at the same time publish in greater measure and more systematically than previously these materials on the history of the Slovak language...[6] (Kotulič 1961, pp. 122 – 123; my translation MRL)

[5] "There is a value in imagining the pure, ideal form of a science because: (i) it makes us aware of how remote from the ideal our actual capabilities are, (ii) we learn humility in the face of so many unanswered questions, and (iii) it humbles the know-it-alls who believe they have grasped the most complex historical developments simply by making some witty remarks. It is absolutely necessary to have a general idea of the forces at play in this massive complex, and we need to keep them in mind at all times in order to correctly categorize the scarce fragments that we do have access to" (Auer – Murray 2015, p. 48). "Es ist schon deshalb von wert sich das idealbild einer wissenschaft in seiner ganzen reinheit zu vergegenwärtigen, weil wir uns dadurch des abstandes bewusst werden, in welchem unser können dazu steht, weil wir daraus lernen, dass und warum wir uns in so·vielen fragen bescheiden müssen, weil dadurch die superklugkeit gedemütigt wird, die mit einigen geistreichen gesichtspunkten die compliciertesten historischen entwickelungen begriffen zu haben meint. Eine unvermeidliche notwendigkeit aber ist es für uns, uns eine allgemeine vorstellung von dem spiel der kräfte in diesem ganzen massenhaften getriebe zu machen, die wir beständig vor augen haben müssen, wenn wir die wenigen dürftigen fragmente, die uns daraus wirklich gegeben sind, richtig einzuordnen versuchen wollen" (Paul 1886, pp. 22 – 23).

[6] "...bude možno podrobnejšie riešiť iba po dôkladnom a široko organizovanom výskume jazykových pamiatok, ktorého cieľom v prvej etape bude dokonalé poznanie jazyka a písomností z tohto obdobia. Teoretický výskum jazyka slovenskej národnosti musí sa opierať nielen o spoľahlivú metodologickú základňu, ale aj o relatívne úplný historický jazykový materiál, ktorý leží roztratený v domácich i zahraničných archívoch (Budapešť a i.). Preto bude potrebné vykonať dôkladný archívny výskum a zároveň vo väčšej miere a systematickejšie ako doteraz vydávať materiály k dejinám slovenského jazyka..." (Kotulič 1961, pp. 122 – 123).

though Kotulič's statement is tempered with his recognition that we only have the data that history leaves us. Where Paul says we should strive to examine "...the entire sum of the products of the linguistic activity of the entire sum of individuals in their reciprocal relations...", Kotulič more cautiously encourages us to use "the entire sum of the linguistic products of the entire sum of individuals" that are *available* to us.

Stepping back to our "tasks for historical sociolinguistics" (derived from Tilly's (1985) "tasks for social history"), we have, up to this point in the discussion, considered the theoretical and methodological frame of the first task: "documenting structural language variation and change", bringing us now to a theoretical/methodological consideration of the second task: "reconstructing the experiences and interactions of ordinary people in connection with that variation and change". And in the same way that it can be helpful to think about the first task as "historical language-documentation fieldwork", it might be of assistance to frame the second task as *historical ethnographic fieldwork*. Ethnographic sociolinguists do fieldwork that involves becoming intimately acquainted with the socio-cultural context of the community that they are working with, attempting to identify the many different lines of affiliation/affinity between the members of the community, and the patterns of interaction of the community members, in order to better understand their patterns of language use. What if we attempt to become intimately acquainted with the writers of our historical texts and the personal interactions among them in order to answer the question of "Who *interacted* with whom, when, where, why, and how?" ? Instead of *a priori* deciding which parts of the historical socio-cultural context to focus on as significant social variables, what if we attempt to draw all possible lines of social affiliation/affinity that we can derive from the available socio-historical data and then inductively look for patterns of social interaction?

This approach of studying patterns of interaction among historical language users allows us to more fully engage with theoretical notions and models that have until now perhaps been under-utilized, or under-emphasized, in exploring specifically *historical* language variation and change:

- Mundane mobility – the "mundane movements we engage in in everyday life" are "small-scale, less dramatic in distance, and perhaps in life impact at the level of the individual, [but] their scale, intensity and pervasiveness at the level of the community as a whole mean they cannot be ignored as a source of rather striking dialect contact" (Britain 2013, p. 165, 168).
- Social networks – the contact patterns formed by instances of mundane mobility can be modeled as social networks that examine the interactions of individuals in their communities and in their constellations of contacts as potential determinants in historical language variation and change. (cf. Bergs 2005; Conde-Silvestre 2012)
- Communities of practice – "[t]he community of practice takes us away from the community defined by a location or by a population. Instead, it focuses on

a community defined by social engagement [the social networks] – after all, it is this engagement that language serves, not the place and not the people as a bunch of individuals" (Eckert – McConnell-Ginet 1992, p. 95).

With these theoretical and methodological considerations, we are now working within a framework that encompasses the first two "tasks of historical sociolinguistics" – "documenting structural language variation and change" and "reconstructing the experiences and interactions of ordinary people..." – but, as mentioned earlier, the third task – "connecting the two" – may be considered the key element in the whole theoretical and methodological conceptualization presented here. And this key element is embedded in a principle that I have espoused in other theoretical and methodological contexts and connections, summed up in the phrase: "Use all the data!".

## 5. USE ALL THE DATA!

This principle of "Use all the data!" has antecedents in the kinds of frames that both Paul (1886) and Kotulič (1961) invoked as "ideal aims" for maximizing the quality of work in historical linguistics, as well as in Janda and Joseph's (2003) "informational maximalism"[7], but the challenge to "Use all the data!" goes beyond the calls for gathering and considering all possible sources of language data (task 1 – "historical language-documentation fieldwork") and maintains that a fuller picture of historical language variation and user interaction will only be visible when we also gather all possible information about the language users' contexts (task 2 – "historical ethnographic fieldwork") and then, critically, bring the two together (task 3) in an examination of "Who used what linguistic features with whom, when, where, why, and how?".

The "Use all the data!" principle states (Lauersdorf 2018a, p. 112; 2018b, pp. 211 – 212; 2021, pp. 215 – 216):

1) Identify all possible sources of language data – data may be "hiding" where you don't expect it, in unexplored physical locations and in unexplored textual locations.
2) Consult the entirety of the language data available to you – avoid selective sampling (inclusion or exclusion) of language data on the basis of *a priori*

[7] "To a great extent, then, what we should really strive for, in diachronic pursuits such as historical linguistics, is what could be called *"informational maximalism"* – that is, the utilization of all reasonable means to extend our knowledge of what might have been going on in the past, even though it is not directly observable. Normally, this will involve a heavy concentration on the immediate present, but it is in fact more realistic just to say that we wish to gain a maximum of information from a maximum of potential sources: different times and different places – and, in the case of language, also different regional and social dialects, different contexts, different styles, different topics, and so on and so forth" (Janda – Joseph 2003, p. 37).

notions of what kind of data you need, how much data you need, where it should come from, etc.

3) Language data isn't the only data – use all the socio-historical data!
   a) Identify and use all possible sources of socio-historical data (again being on the lookout for socio-historical data "hidden" in unexpected places and using the entirety of the socio-historical data available to you).
   b) We only have the language data that history leaves us (what has "survived" through time), so wrap the language data in all possible socio-historical datasets to help complete the picture.

Corollary 1: If you use all the data, view all the data!
a) If you view all the data, view all the combinations.
b) If you view all the data, view all the angles.
c) If you view all the data, use all the techniques.

In addition to being a statement that focuses on the necessity of completing all three "tasks for historical sociolinguistics", this call to "Use all the data!" in historical linguistic investigation also derives, in part, from the fact that historical data is inherently "finite", and therefore it is potentially incomplete, limited, fragmentary, unbalanced, for our investigations in ways that the researcher has no control over (unlike the modern contexts that Fishman or Horvath were working in where, in theory, one can always gather more data). It is therefore imperative to identify and gather as much data as possible for a given historical investigation, from all interrelated sources, both linguistic and socio-historical, and to consider especially non-traditional data sources, if one hopes to be able to assemble a sufficient dataset for data-driven, inductive analysis. If one follows the principle of gathering "all the data" in data collection, the extant historical record can often produce richly layered datasets containing linguistic features of language users in their socio-cultural interactional contexts.

This call to "Use all the data!" and "View all the data!" further derives from "the conviction that using only a selective sample and/or selective methods of analysis of the available data (based on *a priori* assumptions about the features and categories that one should expect to find in the data) limits what one is actually able to find, given that portions of the data are not being considered and that only certain analytical viewpoints are being entertained" (Lauersdorf 2021, p. 216).[8] You will

[8] Feagin (2013) makes a similar statement regarding inadequacies of working non-inductively with preconceived categories: "One danger with selecting informants by pre-selected categories is that results can be self-fulfilling or circular. For a more general community study, Horvath (1985) gathered speech data from a stratified judgment sample in Sydney, Australia, and analyzed it using principal components analysis, a statistical technique which grouped speakers into clusters according to their linguistic similarities, and in that way revealed what the sociolinguistic groupings of Sydney were, based

(almost) always find evidence of things that you are specifically looking for; but you will (almost) never find evidence of things that you are not looking for. And what if the truly salient things are the things that you aren't looking for, or the truly significant patterns and correlations are the ones that you are not considering? Thus, it is imperative to assemble and interrogate the data in a way that facilitates data-driven, inductive examination of all possible combinations of all linguistic and socio-cultural information contained in the rich data layers.[9]

## 6. USER INTERACTION AND QUESTIONS OF LANGUAGE AND IDENTITY

In placing the emphasis now on the third "task for historical sociolinguistics" – the task of *connecting together* "structural language variation and change" with "the experiences and interactions of ordinary people in connection with that variation and change" – we open up opportunities for consideration of questions of language and identity. Indeed, the question "Who used what linguistic features with whom, when, where, why, and how?", that we have used throughout our discussion to this point, can also be interpreted as an inquiry into language and identity.

In part, the traditional research paradigm in the investigation of language history, with its focus on the development of the standard language variety, was an attempt to explore *historical identity through language*, deriving from the 19th-century rise of the nation-state and the use of culture (including language) to construct and define nationhood and national identity, with a desire to trace the national lineage, through its culture and language, as far back in time as possible.[10] In other words, in the modern nation-state, national identity is, in part, defined by the (national) standard language, so studying the history of the standard language provides information about the history of national identity, hence the emphasis on *standard* language history in the traditional research paradigm. To represent this implementation of the traditional "language history" research paradigm as an attempt to explore "historical identity through language", we can simply reverse the language history equation provided earlier:

---

entirely on speech, rather than on preconceived notions about class membership, sex, or other social groupings" (Feagin 2013, pp. 27 – 28).

[9] It is tempting here to further drive home this point by paraphrasing a quote usually attributed to Canadian hockey star Wayne Gretzky: "You miss 100% of the shots you don't take." – for our purposes we might state this as: "You miss 100% of the data / correlations you don't consider."

[10] As stated by Milroy (2001) in regard to English, "These histories have until quite recently almost always been designed as histories of the internal structure of one variety – the standard language… They are largely *codifications* of the history of the standard language. […] When the language is given an authoritative (almost 'official') history in this way, this assures us that it has not merely sprung up overnight like a mushroom, and it becomes important to trace it as far back as possible" (Milroy 2001, p. 548).

Traditional research paradigm for the investigation of language history:
*language history = standard language history = history of national identity*

Traditional research paradigm for the exploration of historical identity through language:
*history of national identity = standard language history = language history*

This traditional interpretation is, in some ways, an example of the aphorism "history is written by the victors", in the sense that the interpretation is driven from a present-day position and view on the past, and as noted by Roach, "Whether we like it or not, we view the past from a modern standpoint, privileging (consciously or otherwise) the interests and ideals of the world we know. As a result, we tend to treat developments towards modernity as natural – and disparage the apparent dead ends that stood in its path" (Roach 2023, p. 8). Applying this to our discussion, this means that we have present-day notions of language and identity based on "how things have turned out", i.e., based on the state of how things are right now – so we take the language and identity frameworks that we have now, and the names that we give them, and we go looking for them in the past to build a lineage for what we have now. But, even if we (probably erroneously) assume that people in the past were always developing their social context in a direction toward the frameworks and names that we have now, they themselves likely didn't frame things in the way we do now and with the names that we use now. Additionally, given the notion that was just mentioned that "we tend to treat developments towards modernity as natural – *and disparage the apparent dead ends that stood in its path* [emphasis added, MRL]", there have likely been many different language and identity frameworks and names in existence historically that have not survived cycles of "disparagement", having been continually disregarded in our ongoing views and perceptions of the "path to modernity". With just those two arguments pointing to the likelihood of many different language and identity frameworks and naming conventions having existed in the past, the question arises: should the starting point of our historical investigations of language and identity really be our present-day frameworks and names?

The issue of "named languages" has become a topic in linguistic research (see, for example, Horner – Weber 2018; Horner – Bradley 2019; Saraceni – Jacob 2019), driven by studies of present-day multilingualism and linguistic "superdiversity", and focusing on the role of the inherited "framing and naming" conventions of the "the victors" who brought us the modern nation-state paradigm: one nation ~ one culture ~ one language. Horner and Bradley (2019) make the point that: "The notion that languages exist as clearly identifiable and bounded objects constitutes a widespread and fundamental language ideological belief. To be sure, the construction of named languages functions similarly to that of other categories, such as ethnicities and nations, all of which can be interpreted and instrumentalized in various ways due to their potential malleability and their situatedness in given social and political contexts" (Horner – Bradley 2019, p. 298).

"Naming things" is ultimately an act of categorizing/classifying according to a specific framework, thereby reifying our perceptions within and through that framework, in this case involving the categorization/classification of historical language and identity. As an example of this, consider the investigation of language variation in written documents in the Slovak-speaking areas of Central Europe during the historical period before the official codification of a Slovak standard language. Written language in documents from that time and place is generally considered to have been influenced to varying degrees by Czech, Polish, and Slovak language varieties (both spoken and written), with the additional presence, in this socio-historical context, of Hungarian, German, and Latin; and much of the linguistic research into documents from that time and place has traditionally revolved around attempts to identify features that show the position of any individual text on a continuum of "Czech-ness" or "Polish-ness" or "Slovak-ness".[11] As I have argued elsewhere (Lauersdorf 2018b), this kind of research paradigm operates with a specific set of *a priori* assumptions and categories, as well as specific investigative methods and goals that derive from these assumptions and categories, that can be stated something like this:

1) *assumption*: the language of the texts must show some kind of overall genetic alignment / affiliation with a specific "named language" category (with the names and categories often coming from a present-day context);
   *goal*: identify the specific genetic affiliation of the language of the texts (e.g., are the texts Czech?, Polish?, Slovak?);
2) *assumption*: certain linguistic (phonological, morphological, syntactic, lexical, semantic) features that have developed distinct variants can be considered diagnostic for genetic affiliation with one of the "named language" categories (different variants of a linguistic feature become "genetic markers" of, for example, "Czech-ness", "Polish-ness", or "Slovak-ness");
   *method*: search the texts for instances of these specific pre-determined linguistic features that have developed distinct variants that will mark genetic affiliation with one "named language" category or another;
3) *assumption*: a preponderance of "genetic markers" of one type or another in the language features of a text is indicative of the overall genetic affiliation of the text with one of the "named language" categories;
   *method and goal*: track the number of occurrences of each diagnostic variant for each linguistic feature (perhaps determining whether the occurren-

[11] For more on the linguistic and socio-historical context of written documents from the Slovak-speaking area in the 15th, and especially 16th, centuries, and a review of the linguistic research into these documents, see Lauersdorf 1996 and 2010 and the literature cited there.

ces show uniformity or variation in the texts) with the hope that there will be a clearly identifiable profile that points to predominantly one "named language" affiliation or another (and hence one language identity or another) (adapted from Lauersdorf 2018b, pp. 208 – 209).

Not only does this kind of research paradigm involve the use of (likely anachronistic) social and linguistic frames and names from modern times as the position from which it investigates language and identity in earlier periods (a position that this discussion directly argues against), but this practice also limits the scope of the investigation to only selective portions of the overall data using only a selective set of methods and viewpoints, which goes in the opposite direction from the "Use all the data!" principle espoused above.[12]

In contrast with all of this, Horner and Weber's (2018) "social approach" to multilingual contexts emphasizes the role of social interaction and the use of language socially, and the importance of investigating language in its actual interactional context – in our case its actual *historical* interactional context. With this in mind, what if, in our investigations of language variation and user interaction in historical contexts, we do not begin with pre-determined names and categories that channel and constrain the analysis of our data and results? In performing task 3 of the "tasks for historical sociolinguistics" (the task of *connecting* the variation in language features with the social interactions of their users), taking an approach of discovering *inductively* "Who used what linguistic features with whom, when, where, why, and how?", and describing the resulting patterns, allows us to present information about the past and how it evolved into the present state without imposing (likely anachronistic) assumptions that are based on present-day framing and naming.

## 7. A FINAL POSITION STATEMENT

In the revised research paradigm for the investigation of language history, that I put forth in section 2 above:

[12] This questioning of the use of present-day frames and names in our research paradigms for historical language settings should also prompt questions regarding what the writers of historical documents themselves thought about their written linguistic practice and their identity – i.e., their framing and naming conventions that they used in their time period. Did they give a *name* to the collection of linguistic features that they were writing in any given document (in this example "Czech" or "Polish" or "Slovak"); and did they equate the linguistic features that they were writing with a *named* identity for themselves or with a *named* identity for the intended recipient(s) of any given document? Consideration of these questions goes beyond the scope of this essay, but see Lauersdorf 2018b for some additional discussion of these and related questions of "the writing of historical identity".

| | | | |
|---|---|---|---|
| language history = | history of variation and change in all varieties of language in play in user interactions in a given socio-cultural context | = | a multifaceted approach to questions of history of language and identity |

it is the third portion – "a multifaceted approach to questions of history of language and identity" – that is realized by pursuing the "tasks of historical sociolinguistics" all the way through to the third, highly important task of *connecting language and users*. In sum:

A data-driven, use-all-the-data connection and correlation of:

patterns of language variation derived by data-driven, inductive methods from full-scale "historical language-documentation fieldwork" that collects and analyzes the full scope of available historical data (i.e., the products of task 1)

AND

patterns of social interaction of language users carefully reconstructed from full-scale "historical ethnographic fieldwork" that reconstructs, for example, mundane mobility, social networks, communities of practice (i.e., the products of task 2)

expressly seeks to investigate:

who used what linguistic features with whom, when, where, why, and how? (i.e., task 3).

This set of tasks, carried out in this way as a research paradigm, seeks to specifically avoid *a priori* assumptions that lead to results based on pre-determined selective sampling of data and methods and to interpretations based on likely anachronistic present-day frameworks and names, and thereby holds the potential to paint a more detailed and accurate picture of the social history of language varieties and their users, providing greater richness in our overarching understanding of language history and historical linguistic identity.

## References

AUER, Peter – MURRAY, Robert W. (eds.) (2015): *Hermann Paul's* Principles of Language History *Revisited: Translations and Reflections*. Berlin and Boston: Walter de Gruyter.

BEAMAN, Karen V. – GUY, Gregory R. (eds.) (2022): *The Coherence of Linguistic Communities: Orderly Heterogeneity and Social Meaning*. New York: Routledge.

BERGS, Alexander (2005): *Social Networks and Historical Sociolinguistics: Studies in Morphosyntactic Variation in the Paston Letters (1421–1503)*. Berlin and New York: Mouton de Gruyter.

BRITAIN, David (2013): The Role of Mundane Mobility and Contact in Dialect Death and Dialect Birth. In: D. Schreier – M. Hundt (eds.): *English as a Contact Language*. Cambridge: Cambridge University Press, pp. 165–181.

CERRUTI, Massimo – VIETTI, Alessandro (2022): Identifying Language Varieties: Coexisting Standards in Spoken Italian. In: K. V. Beaman – G. R. Guy (eds.): *The Coherence of Linguistic Communities: Orderly Heterogeneity and Social Meaning*. New York: Routledge, pp. 261–280.

CONDE-SILVESTRE, Juan Camilo (2012): The Role of Social Networks and Mobility in Diachronic Sociolinguistics. In: J. M. Hernández-Campoy – J. C. Conde-Silvestre (eds.): *The Handbook of Historical Sociolinguistics*. West Sussex: Wiley-Blackwell, pp. 332–352.

ECKERT, Penelope – MCCONNELL-GINET, Sally (1992): Communities of Practice: Where Language, Gender, and Power All Live. In: K. Hall – M. Bucholtz – B. Moonwomon (eds.): *Locating Power, Proceedings of the 1992 Berkeley Women and Language Conference*. Berkeley: Berkeley Women and Language Group, pp. 89– 99.

FEAGIN, Crawford (2013): Entering the Community: Fieldwork. In: J. K. Chambers – N. Schilling (eds.): *The Handbook of Language Variation and Change*. 2nd ed. West Sussex: Wiley-Blackwell, pp. 19– 37.

FISHMAN, Joshua A. (1965): Who speaks what language to whom and when? In: *La Linguistique,* Vol. 1, No. 2, pp. 67– 88.

GHYSELEN, Anne-Sophie – DE VOGELAER, Gunther (2018): Seeking Systematicity in Variation: Theoretical and Methodological Considerations on the 'Variety' Concept. In: *Frontiers in Psychology*, Vol. 9, Article 385. Available at: https://doi.org/10.3389/fpsyg.2018.00385

GUY, Gregory R. (1980): Variation in the Group and in the Individual: The Case of Final Stop Deletion. In: W. Labov (ed.): *Locating Language in Time and Space*. New York: Academic Press, pp. 1–36.

GUY, Gregory R. – HINSKENS, Frans (2016): Linguistic Coherence: Systems, Repertoires and Speech Communities. In: *Lingua*, Vol. 172–173, pp. 1–9. Available at: https://doi.org/10.1016/j.lingua.2016.01.001

HINSKENS, Frans – GUY, Gregory R. (eds.) (2016): Coherence, Covariation and Bricolage. Various Approaches to the Systematicity of Language Variation. Special issue of *Lingua*, Vol. 172–173. Available at: https://www.sciencedirect.com/journal/lingua/vol/172

HORNER, Kristine – WEBER, Jean-Jacques (2018): *Introducing Multilingualism: A Social Approach*. 2nd ed. London and New York: Routledge.

HORNER, Kristine – BRADLEY, Andrew J. (2019): Language Ideology. In: J. Darquennes – J. C. Salmons – W. Vandenbussche (eds.): *Language Contact: An International Handbook, vol. 1*. (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science / Manuels de linguistique et des sciences de communication [HSK], 45/1). Berlin: De Gruyter Mouton, pp. 296–307.

HORVATH, Barbara (1985): *Variation in Australian English: The Sociolects of Sydney*. Cambridge: Cambridge University Press.

HORVATH, Barbara – SANKOFF, David (1987): Delimiting the Sydney speech community. In: *Language in Society*, Vol. 16, No. 2, pp. 179–204.

HUA, Xia – MEAKINS Felicity – ALGY, Cassandra – BROMHAM, Lindell (2022): Language Change in Multidimensional Space: New Methods for Modelling Linguistic Coherence. In: *Language Dynamics and Change*, Vol. 12, No. 1, pp. 78 – 123. Available at: https://doi.org/10.1163/22105832-bja10015

JANDA, Richard D. – JOSEPH, Brian D. (2003): On Language, Change, and Language Change – Or, Of History, Linguistics, and Historical Linguistics. In: B. D. Joseph – R. D. Janda (eds.): *The Handbook of Historical Linguistics*. Oxford: Blackwell Publishing, pp. 3–180.

KOTULIČ, Izidor (1961): Z mestských kníh Partizánskej Ľupče zo 16. storočia. In: E. Pauliny (ed.): *Štúdie a pramene k dejinám jazyka slovenskej národnosti. Jazykovedné štúdie VI*, Bratislava: Vydavateľstvo SAV, pp. 119–177.

LAUERSDORF, Mark Richard (1996): *The question of 'cultural language' and interdialectal norm in 16th century Slovakia: A phonological analysis of 16th century Slovak administrative-legal texts*. München: Otto Sagner. [Slavistische Beiträge, 335.]

LAUERSDORF, Mark Richard (2010): *The morphology of 16th-century Slovak administrative-legal texts and the question of diglossia in pre-codification Slovakia*. München: Otto Sagner. [Slavistische Beiträge, 473.]

LAUERSDORF, Mark Richard (2018a): Linguistic Visualizations as *objets d'art*? In: N. Bubenhofer – M. Kupietz (eds.): *Visualisierung sprachlicher Daten*. Heidelberg: Heidelberg University Publishing, pp. 91–122. Available at: https://doi.org/10.17885/heiup.345.c4397

LAUERSDORF, Mark Richard (2018b): Historical (Standard) Language Development and the Writing of Historical Identities: A Plaidoyer for a Data-Driven Approach to the Investigation of the Sociolinguistic History of (Not Only) Slovak. In: S. M. Dickey – M. R. Lauersdorf (eds.): *V zeleni drželi zeleni breg: Studies in Honor of Marc L. Greenberg*. Bloomington, Indiana: Slavica Publishers, pp. 199–218.

LAUERSDORF, Mark Richard (2021): Historical Sociolinguistics and the Necessity of Interdisciplinary Collaboration. In: A. Burkette – T. Warhol (eds.): *Crossing Borders, Making Connections: Interdisciplinarity in Linguistics*. Berlin: De Gruyter Mouton, pp. 207–230. Available at: https://doi.org/10.1515/9781501514371-015.

MEYERHOFF, Miriam – KLAERE, Steffen (2017): A case for clustering speakers and linguistic variables: Big issues with smaller samples in language variation. In: I. Buchstaller – B. Siebenhaar (eds.): *Language Variation – European Perspectives VI: Selected papers from the Eighth International Conference on Language Variation in Europe (ICLaVE 8), Leipzig, May 2015*. Amsterdam and Philadelphia: John Benjamins, pp. 23–46. Available at: https://doi.org/10.1075/silv.19.02mey

MILROY, James (2001): Language Ideologies and the Consequences of Standardization. In: *Journal of Sociolinguistics*, Vol. 5, No. 4, pp. 530–55.

PAUL, Hermann (1886): *Principien der Sprachgeschichte*. Zweite Auflage. Halle: Max Niemeyer.

PAUL, Hermann (1891): *Principles of the History of Language*. Translated from the second edition of the original by H. A. Strong. New and revised edition (1890). London: Longmans, Green, and Co.

ROACH, Levi (2023): The powerful of one era are not the same as those of the preceding one. Contribution to: Is History Written by the Winners? In: *History Today* 73.6 (June): 8. Available at: https://www.historytoday.com/archive/head-head/history-written-winners

SARACENI, Mario – JACOB, Camille (2019): Revisiting borders: Named languages and de-colonization. In: *Language Sciences*, Vol. 76. Available at: https://doi.org/10.1016/j.langsci.2018.05.007

TILLY, Charles (1985): Retrieving European Lives. In: O. Zunz (ed.): *Reliving the Past: The Worlds of Social History*. Chapel Hill: University of North Carolina Press, pp. 11–52.

TRUMPER, John – MADDALON, Marta (1990): Il problema delle varietà: L'italiano parlato nel Veneto. In: M. A. Cortelazzo – A. M. Mioni (eds.): *L'italiano regionale. Atti del XVIII Congresso Internazionale di Studi. Padova – Vicenza, 14–15 settembre 1984*. Roma: Bulzoni, pp. 159–191.

VIETTI, Alessandro (2019): La varietà di lingua come insieme di tratti coerenti: Verso una caratterizzazione empirica. In: *Rivista Italiana di Dialettologia. Lingue dialetti società*, Vol. 43, pp. 11–32.

WANDRUSZKA, Mario (1975): Mehrsprachigkeit. In: H. Moser (ed.): *Sprachwissenschaft und Sprachdidaktik: Jahrbuch 1974 des Instituts für deutsche Sprache*. *Sprache der Gegenwart, 36*. Düsseldorf: Schwann, pp. 321–350.

WANDRUSZKA, Mario (1979): *Die Mehrsprachigkeit des Menschen*. München: Piper.