Steven Moran*, Shelece Easterday* and Eitan Grossman

Current research in phonological typology

https://doi.org/10.1515/lingty-2022-0069
Received October 3, 2022; accepted December 11, 2022; published online February 1, 2023

1 Aims and background

Despite the fact that phonology has historically been integral to the development of modern approaches to linguistic theory and typology (Greenberg et al. 1978; Hockett 1955; Jakobson 1941; Martinet 1955; Trubetzkoy 1939), its study is still underrepresented in many important venues (Hyman 2007; Hyman and Plank 2018). Therefore, we (Moran and Grossman) organized two conference workshops in 2019 aimed specifically at highlighting current research in phonological typology. The first, entitled 'Phonological (in)stability and language evolution,' took place at the annual meeting of the *Societas Linguistica Europaea (SLE)*, from August 21–24, in Leipzig, Germany. The second, which shares the title with this special volume, was held at the meeting of the *Association for Linguistic Typology (ALT)*, from September 4–6, in Pavia, Italy. The present special issue emerged from these two workshops. Its aims are twofold.

First, like those before us (e.g., Hyman and Plank 2018), we aim to raise the profile of phonological typology within typology and beyond. Phonological typology is particularly important for a number of reasons. First, phonology is one of the few domains of language in which we have substantial cross-linguistic data. Second, phonology has traditionally been a testing ground for linguistic theories, including a range of structuralist, generativist, and functionalist or usage-based theories; many concepts first developed for phonology were later expanded to other domains of language, and phonological data have been the basis for diverse perspectives on the relationship between synchrony and diachrony. Third, there is broad agreement among linguists of every persuasion that the sound patterns of languages are ultimately phonetically-based, which means that typological studies can make reference

***Corresponding authors: Steven Moran ['stivɛn 'mɔ˞ræn]**, Department of Biology, University of Neuchatel, Neuchatel, Switzerland; and Department of Anthropology, University of Miami, Coral Gables, FL, USA, E-mail: steven.moran@unine.ch. https://orcid.org/0000-0002-3969-6549; and **Shelece Easterday [ʃə'lis 'ɪstə˞deɪ]**, Department of Linguistics, University of Hawai'i Mānoa, Honolulu, HI, USA, E-mail: shelece@hawaii.edu. https://orcid.org/0000-0003-1186-1044
Eitan Grossman ['eɪtɐn 'gʁosmɐn], Hebrew University of Jerusalem, Jerusalem, Israel, E-mail: eitan.grossman@mail.huji.ac.il. https://orcid.org/0000-0001-7479-7831

to the observable and measurable facts of production and perception, which are themselves grounded in human physiology and cognition. Fourth, phonological typology provides materials for hypotheses linking conventional aspects of language with extra-linguistic variables, such as geography and topography, demographic aspects of speech communities, genes, and more. Despite these advantages, some truly foundational methodological and theoretical issues in phonological typology are still unresolved. For example, the fundamental units of cross-linguistic comparison are still debated, and the reliability of existing cross-linguistic databases has been questioned. Moreover, phonological typology is often marginalized within typology, where morphosyntax and to a lesser extent lexical semantics dominate the discussion. In light of the disparity between the empirical, methodological, and theoretical importance (and achievements) of phonological typology, on the one hand, and its prominence within typology on the other, renewed attention to its study seems warranted.

Second, we are interested in contributing to the ongoing search for the factors that shape the phonological systems of the world's languages, specifically by exploring the question of 'what's where why?' (Bickel 2007; Hyman 2014, 2018). This question represents a recent paradigm shift in linguistic typology, which moves away from the goals of classical typology, such as the classification of languages into types and the identification of gaps between logically possible and attested patterns of language (Comrie 1989; Dryer 1989, 1992; Greenberg et al. 1978). Instead, due to factors inherent to the nature of the population of languages accessible to linguists today, typologists have increasingly come to formulate explicit causal theories about cross-linguistic diversity and its sources (Bickel 2014, 2015, 2017; Nichols 1992; Nettle 1999).

After we issued calls for papers, we received submissions on a wide range of topics within phonological typology – reflecting its diverse and engaged practitioners. The authors of the papers in this special volume demonstrate the challenges involved in doing phonological typology today – both in theory and in practice. Complementing these discussions are papers that explicitly model phonological patterns and processes. These contributions bring together some of the state-of-the-art research underway. They also highlight the many ongoing and lively debates regarding both phonological theory and linguistic typology.

We begin with a brief historical overview of phonological typology and its key questions, achievements, and challenges. While the amount of new high-profile research due to increased access to comprehensive reference materials, detailed datasets, and computational tools, methods and models, is impressive – it may be even more impressive that some key issues regarding cross-linguistic comparability in phonological typology remain largely unresolved today. We discuss these issues in

the hope that researchers will find constructive ways to move forward, turning critical aporias into theoretical, methodological, and empirical opportunities and challenges.

We then summarize the contributions in this volume, highlighting how the authors tackle some of the big questions in phonological typology today.

2 Phonological typology: key issues

Phonology has always been typological in nature because its aim is to determine the possible range of phonological systems through cross-linguistic comparison (Hyman 2014). Its inseparability from typology is also reflected in their shared historical development. Typology, at least in its structuralist phase, began with the cross-linguistic study of phonological systems (Hockett 1955; Jakobson 1941; Trubetzkoy 1939) and later the classification of languages into types in search of language universals (Greenberg et al. 1978).[1] Recent approaches shift away from language classification and towards explanations for the distribution of linguistic properties and processes across space and time (Nichols 1992).

Maddieson (this volume) writes, "*Why* languages are the way they are today is the most intriguing, but the most difficult part of the question, because the typologist must first correctly answer 'what?' and 'where?'". At first glance, getting the answers to 'what and where?' is now easy because of increased access to data, together with tools for quantitative analysis and data visualization. For example, with a few lines of code, a researcher nowadays can download and combine numerous freely available datasets (with linguistic and non-linguistic information) and then interactively explore patterns of interest. And due to language reference catalogs such as Glottolog (Hammarström et al. 2022), results about these patterns can also easily be plotted as points or polygons onto world maps.

This increased ease of access to data, quantitative approaches, and visualization techniques has given rise to a powerful paradigm for exploratory analysis in linguistic typology. However, this new data science approach has not solved, but rather exacerbated the interpretability of research results in light of fundamental and decades-old questions in phonology and typology. Two of the most fundamental questions both deal with cross-linguistic comparability: (i) what are the units of comparison (e.g., speech sounds, phonemes, syllables, morae, phonological processes)? And (ii) how can they be compared?

1 See also van der Hulst (2017) for an overview.

Today, as in all data-driven science, typologists need the qualitative and quantitative skills to undertake research in typology and also to confidently call bullshit (Bergstrom and West 2021) on research presented in quantitative form – whether through statistics or data visualizations (cf. Bickel and Nichols 2020; McNew et al. 2018). How did we reach our present situation, in which larger quantities of data, increasingly sophisticated statistical tools, and more powerful computational resources, permit large-scale analyses that present challenges in terms of reliability of data, appropriateness of methods, and interpretability and evaluability of results?

The trend towards "big" and open data in typology started nearly 20 years ago, most prominently because of the publication of the *World Atlas of Linguistic Structures* (WALS; Dryer and Haspelmath 2013). First published as a coffee table book and CD-ROM application, it has become an interactive web application with downloadable raw data. The success of open typological databases – first WALS and now many others published by MPI-EVA, MPI-SHH, and by other institutions – has profoundly changed the kinds of questions and the amount of research being published in typology. Nevertheless, the comparability problem remains, and is more prevalent now that researchers combine typological databases curated at different levels of linguistic analysis.

The design of the chapters on phonology in WALS and many other typological databases builds on the landmark work in synchronic phonological typology undertaken by Maddieson (1984) in his publication of the UCLA Phonological Segment Inventory Database (UPSID). By selecting a genealogically balanced sample of the world's languages with carefully analyzed segment inventories and explicit phonological features, Maddieson aimed to address the fundamental questions of comparability and inference of language universals that had been raised in the decades before during the US National Science Foundation funded *Typology and Universals* project at Stanford from 1969 to 1976 (PIs Joseph Greenberg and Charles Ferguson).[2]

Greenberg and his colleagues are best known for their research on the classification of languages into types and the search for language universals.[3] Much of the resulting research during these years was published in the narrowly distributed and now difficult to access 20-volume series titled, *Working Papers on Language Universals*.[4] Some of the highlights from that series were published in 1978 by Stanford University Press in a four-volume collection titled *Universals of Human*

2 https://dlc.hypotheses.org/618.

3 The comparative approach stems from anthropology, itself inspired by typology as a methodology of comparison of codable traits that are part of the early development and heyday of any biological science (Heath 2016).

4 https://linguistics.stanford.edu/news/truly-archives-language-universals-project.

Language. Volume Two in this series is devoted specifically to phonological typology and includes the chapters:

- Correlations of stops and fricatives (Gamkrelidze)
- Palatalization (Bhat)
- Universals of vowel systems (Crothers)
- Syllabic consonants (Bell)
- Nasal vowels (Ruhlen)
- Consonant clusters (Greenberg)
- Consonant harmony (Vihman)
- Metathesis (Ultan)
- Tone (Maddieson)
- Phonological processes (Ferguson)
- Word demarcation (Hyman)
- Intonation (Bolinger)
- Size-sound symbolism (Ultan)

Most of the research projects in the *Typology and Universals* project were based on convenience samples of languages in grammars available in the personal libraries of Ferguson and Greenberg or in the Stanford libraries.

Their work on phonological typology involved the development and analysis of the Stanford Phonology Archive (SPA), the first computer-based typological database (Crothers et al. 1979; Sherman and Vihman 1972). SPA is a convenience sample of 200 languages that includes information about phonological inventories, including the description of phonemes, allophones and their conditioning environments, and marginal sounds. Work on and with SPA raised two important questions regarding characterizing the distribution of cross-linguistic variation; both are still relevant and unresolved today:

- What constitutes adequate descriptive categories for cross-linguistic comparison (Sherman and Vihman 1972: 163)?
- How do we construct a cross-linguistic sample of languages that captures genealogical and areal diversity and thus allows for proper characterization of the distribution of linguistic phenomena (Sherman 1975)?

At the heart of the question of "what constitutes the basis for cross-linguistic comparison?" is whether "formal" descriptive (i.e., grammatical) categories can be equated and made comparable across languages (cf. Heath 2016). What has ensued in the decades since SPA is the debate about equatable concepts in phonology (e.g., Simpson 1999; Vaux 2009; Anderson 2017; Kiparsky 2018; Maddieson 2018), but also broadly in typology (Haspelmath 2010; Newmeyer 2010). The question about what the "right" descriptive categories (or comparative concepts) are, is highlighted by the

perennial problem of phonemes – and "emic" categories generally. Here again one sees that phonological typology inherits the conceptual problems of both phonology and typology.

The second question is how does one create a cross-linguistic sample of languages that captures genealogical and areal diversity, and therefore provides for proper characterization and inference about the synchronic and diachronic distributions of linguistic phenomena (Sherman 1975)? What confounds must be taken into account when applying quantitative methods on cross-linguistic datasets in which we know there are biases? Languages are not statistically independent data points. Languages may share features because they are genealogically related or because they have been in contact. This statistical sampling problem was raised by Sherman (1975) and Bell (1978) and in the decades since – due particularly to increased data access – has become more pertinent, resulting in many debates in the typological literature (e.g., Rijkhoff and Bakker 1998; Janssen et al. 2006; Bakker 2011; Moran et al. 2012). While balanced samples are still used, we note a recent trend towards denser databases that allow factors such as genealogical affiliation and geographical location to be taken into account.

Work on and with SPA was presumably informed by Hockett's (1955) *A Manual of Phonology*. Much of the manual is devoted to the development of a typology of sound systems based on a sample of 55 languages which were analyzed in terms of their consonant and vowel inventories, syllable structures, tones, and phonotactics.[5] One example is shown in Figure 1 (Hockett 1955: 139). Hockett provides a hand-calculated and hand-drawn plot of the vowel-consonant balance from his language sample. The plot contains several layers of information about the distribution of vowels in the languages in the sample.

Also insightful is Hockett's discussions in *A Manual of Phonology* and elsewhere on information theory, entropy, and functional load (e.g., Hockett 1955, 1967), in part because at the time these measures were conceivable for cross-linguistic analysis, but were impractical to calculate by hand.[6] Furthermore, principles such as functional load require a notion of phonological contrast on which language comparison can be made. Thus, early work in phonological typology by Hockett and others built on research in previous decades on defining the notion of phonological contrast.

Trubetzkoy (1939) proposed some general principles of the organization of sound systems and their limitations. He provided a full exposition of the theory of phonemes as bunches of binary oppositions (adhering to the principle that ideally, when describing a particular language, one should strive to use as few oppositions as

5 One chapter is literally titled: "A Typology of Phonologic Systems".

6 In those days the word for "computer" was still mainly associated with humans – often women – solving mathematical equations by hand.

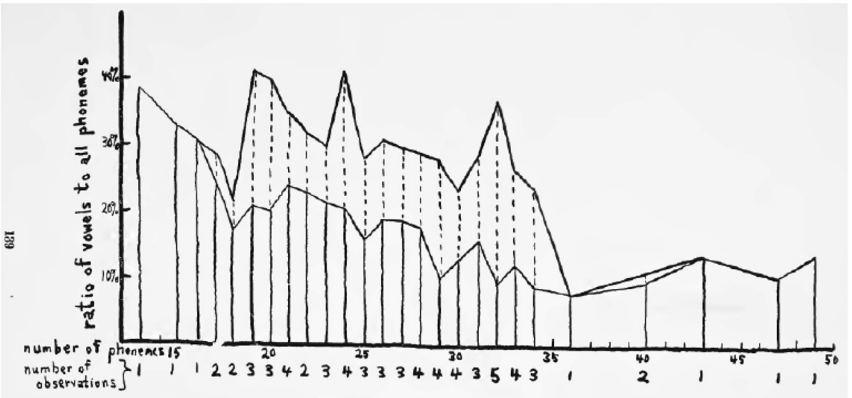

FIG. 12. VOWEL-CONSONANT BALANCE

A vertical line shows the occurrence of at least one system with the indicated number of segmental phonemes (among the systems examined). The lower broken curve joins together the lowest ratios of vowels to all phonemes, and the upper broken curve the highest ratios. Total sample: 68 languages.

Figure 1: Vowel-consonant balance from Hockett (1955: 139).

possible). However, distinctive features as the most basic phonological unit may differ in how they are modeled, e.g., do they describe articulatory or acoustic features and how do they map to the physiological mechanisms involved in production and perception? This key issue in phonological analysis is directly relevant to the general problem of comparability in linguistic typology.

So when did the comparability problem emerge in phonology and typology? The much cited quote by Sapir – nearly 100 years ago now – shows the roots of the relationship between the two:

> … it almost goes without saying that two languages, A and B, may have identical sounds but utterly distinct phone[mic] patterns; or they may have mutually incompatible phonetic systems, from the articulatory and acoustic standpoint, but identical or similar [phonemic] patterns (Sapir 1925: 43).

Despite the fundamental problems inherent to phonological typology, research on the cross-linguistic distribution of sound patterns has continued in the 21st century, with a recent monograph presenting the field (Gordon 2016); studies of particular sound patterns (e.g., Bateman 2007, 2011 on palatalization, Nikolaev and Grossman 2018 on affricates, Easterday 2019 on highly complex syllable structure) and of phonological universals (Hyman 2008; Pericliev 2008); investigations of the structure of phonological segment inventories (Clements 2003, 2009; Coupe et al. 2009; Dunbar and Dupoux 2016; Nikolaev and Grossman 2020); typological reconsiderations of fundamental assumptions in phonological theory (e.g., Mielke 2008; Mackie and Mielke 2011); analyses of areal aspects of phonology (e.g., Stolz and Levkovych 2021;

Nikolaev and Grossman 2018; Blevins 2017), studies of the impact of the physical environment on sound patterns (e.g., Everett 2013, 2017; Everett et al. 2015, 2016; Roberts 2018). The 21st century has also seen a number of new databases with global scope, such as LAPSyD (Maddieson et al. 2013), P-Base (Mielke 2013), PHOIBLE Online (Moran and McCloy 2019), as well as databases of particular areas, families, or phenomena, e.g., SAPHON for South America (Michael et al. 2015), EURPHON for Eurasia (Nikolaev 2018), Australian phonemic inventories (Round 2019), BDPROTO for ancient and reconstructed languages (Marsico 1999; Marsico et al. 2018; Moran et al. 2020), and SEGBO for borrowed phonological segments (Grossman et al. 2020), just to name a few. Data from phonological typology have been used in studies of human history and prehistory, as well as language evolution, e.g., Atkinson (2011; see also responses, including Cysouw et al. 2012; Maddieson et al. 2011; Sproat 2011; Wang et al. 2012), Creanza et al. (2015), Dediu et al. (2017), Blasi et al. (2019). And more recently, computational approaches to phonological typology have flourished (e.g., Cotterell and Eisner 2017; Pimentel et al. 2020; Salesky et al. 2020) and include expansive cross-linguistic corpora for phonetic-acoustic analysis (e.g., Black 2019; Jackson et al. 2020; Cohen Priva et al. 2021; Seifart et al. 2022; Ylonen 2022).

3 Contributions to this special collection

This special issue's title conveys a broad theme and of course we cannot cover all of the vibrant current research going on in phonological typology. We solicited contributions from participants at both conference workshops and we also circulated widely a general call for papers. We asked for submissions related to these questions, but we were also open to other questions related to phonological typology:

- What should be the big questions in phonological typology? Are any new ones on the horizon?
- What is the current theoretical and empirical status of phonological universals?
- To what extent can we do typology at the macro versus micro levels?
- What phonological properties have been under-studied typologically?
- Why are phonemic categories (and descriptive categories in general) inadequate for typology? What are the alternatives?
- How do different approaches to typology, e.g., Canonical Typology or Distributional Typology, apply to phonology?
- What can sound changes and synchronic alternations in typological databases tell us about diachrony?
- How can we model the impact of language contact on phonological distributions?

– What can we discern from the landscape of particular geographic areas from a typological perspective?
– How does language acquisition affect worldwide phonological diversity?

We received a wide range of submissions from leading scholars in the field – and to our delight – from both the new and old guards. The views of these researchers provide not only breadth, but also depth, on current issues in phonology and typology.

We submit these three categories, based on themes in the articles, as our typology of the papers (however, we admit many other typologies are possible):
– Diachrony & synchrony
– Challenges
– Models

3.1 Diachrony & synchrony

The contributions in this volume that deal with the relationship between diachrony and synchrony include:
– The 'what,' 'where' and 'why' of global phonological patterns (Ian **Maddieson**)
– Diachronic phonological typology: Understanding inventory structure through sound change dynamics (Shelece **Easterday** & Joan **Bybee**)
– The typological frequency of consonants is significantly predictive of their order of acquisition, while language-internal frequency and functional load are not (Caleb **Everett** & Sophie **Schwartz**)

3.1.1 Maddieson

Maddieson's contribution to this volume looks at the big picture of 'what's where why?' in global phonological patterns. To get to 'why?', a typologist must first answer the 'what' and 'where' questions. In Maddieson (this volume), phonological patterns are either phonological inventories or phonological processes – inventories being the well-known elements (vowels, consonants, syllable structures, tones, stress patterns) and their distributions and relative frequencies both across and within languages. Processes are the dynamics of the systems under study. Segment databases have long existed for large diverse samples because data on phonological inventories is easy to collect; however, databases of processes less so (but see Mielke's P-base).[7]

7 https://pbase.phon.chass.ncsu.edu.

Maddieson also addresses the "where" question by using point data (geo-coordinates for latitude and longitude), allowing for the creation of global maps of the distributions of linguistic phenomena, which have been very successful in projects like WALS (Dryer and Haspelmath 2013). With these two tools in hand – large-scale inventory databases and quantitative tools for data visualization – one can begin to analyze worldwide linguistic patterns, taking into account factors for explaining phonological typological patterns: physiological, perceptual, and processing constraints involved in speech production; confounds of genealogical descent and areal contact; and sociolinguistic and potentially environmental factors on language change and evolution of phonological systems. As a case study, Maddieson discusses recent research on potential climatic and environmental factors and how they have an impact on the distribution of phonological features. This is an area of recent heated debate, but one that we agree is a potentially interesting way forward: environment does modify the transmission of acoustic signals, but does that have an observable impact on spoken languages?

3.1.2 Easterday & Bybee

The contribution by **Easterday & Bybee** advocates for a diachronic approach to the study of phonological inventories. By focusing on how specific sound change pathways emerge (in this article, ejectives and palatalized consonants), they argue for an emergentist view that takes into account the dynamic processes that create the structures reported in phonological inventories cross-linguistically. This approach produces more accurate predictions about the phonological patterns reported cross-linguistically and leads to a fuller picture of structural generalizations regarding phonological inventories. This work incorporates both synchronic and diachronic aspects of languages and reiterates that phonological inventories are dynamic systems. As such, the authors provide a method for diachronic typology that can be applied to all aspects of phonological structures to investigate paths of sound change and language evolution.

3.1.3 Everett & Schwartz

The contribution by **Everett & Schwartz** examines phoneme acquisition in English to test whether speech sounds that are easier to produce and perceive are acquired earlier due to biological factors that hold across the human species. This hypothesis is based on an observation that dates back to at least Jakobson (1941), i.e., that contrastive speech sounds that are cross-linguistically common are good predictors of age of acquisition due to aspects of their universality. However, a language's phonology is a complex interplay of factors including relative segment frequency

but also its position in the syllable, morpheme, or word, as well as other structural and information theoretic factors. In this paper, the authors argue that the cross-linguistic frequency of consonant phonemes is predictive of their age of acquisition. They note some exceptions, but their general observations of speech sound production are in line with biological aspects of vocal tract development and other quantifiable aspects of phonology in terms of information theory.

3.2 Challenges

Much recent work addresses various challenges in phonological typology. In this volume, the contributions include:

- Canonical phonology: Resolving the four dilemmas of phonological typology (Erich **Round**)
- Frequent violation of the sonority sequencing principle in hundreds of languages: How often and by which segments (Ruihua **Yin**, Jeroen **van de Weijer** & Erich **Round**)
- On the comparability of prosodic categories: why 'stress' is difficult (Nikolaus **Himmelmann**)
- Are creole word-prosodic systems special? (Ana Lívia **Agostinho**)
- The foot: beyond iambs and trochees (Kirsten **Culhane**)
- Robust co-occurrence networks of consonants and the Basic Consonant Inventory (Dmitry **Nikolaev**)

3.2.1 Round

Round's contribution outlines four apparently unrelated dilemmas which complicate the effectiveness of cross-linguistic comparison in phonological typology. The paper proposes that all of these dilemmas, which are in fact unavoidable, arise from how linguists attempt to deal with criterial conflicts, cases in which the properties of a phonological phenomenon pull its analysis in different directions. Using segmentation as a case study, the author shows how the method of Canonical Typology can be applied to the analysis of phonological phenomena for the purpose of typologizing. He argues that normalization, the traditional strategy for dealing with criterial conflicts, can potentially lead to unreliable results, and that the more labor-intensive strategies of factorial analysis and multivariate analysis are better equipped to accurately showcase the diversity and inherently multidimensional nature of sound patterns. The paper identifies criterial conflicts as the area in which phonological theory would benefit most from the insights of typology, and proposes that

phonological typology focus on elucidating an inventory of known criterial conflicts so that more effective strategies for dealing with them can be developed.

3.2.2 Yin et al.

The paper by **Yin et al.** investigates the sonority sequencing constraint (SSC) in a cross-linguistic study of nearly 500 languages from 58 language families. It has long been noted that sonority sequence violations are common in languages (see Clements 1990 for discussion) and much research towards a reformulation, e.g., in terms of perceptual constraints has been undertaken (Wright 2004). Nevertheless, Yin et al. (this volume) show that the governing principle of syllable structure within the SSC remains contested. Therefore, the authors define sonority in terms of acoustic intensity and then probe their cross-linguistic dataset for SSC violations within consonant clusters word initially and finally. They find that nearly half of the languages in their sample violate this definition of the SSC and that the violations in onsets and codas are not symmetrical. This type of data-driven research in phonological typology highlights how the testing of long standing theories can be relevant for debates on universal phonotactics. Their data and methodology also show the importance of openly accessible data and reproducible methods, models, and code.

3.2.3 Himmelmann

Himmelmann's contribution adds to the debate on the cross-linguistic comparability of linguistic categories. He examines the prosodic category 'stress' in West Germanic languages and argues that it is a complex cluster concept comprised of at least six dimensions with numerous sub-dimensions. Given the complexity of stress in West Germanic languages, as a concept it is not useful for cross-linguistic comparison. Nevertheless, Himmelmann argues that at least one of the six dimensions provides a promising point for further typological enquiry – the acoustic and auditory prominence of stress. As such, he argues that the comparability of cluster concepts, such as stress, need not use arbitrary comparative concepts.

3.2.4 Agostinho

The paper by **Agostinho** examines the word prosodic systems of Afro-European creole languages and shows a correlation between the prosodic patterns and the lexical origin (African vs. European). In this regard, Agostinho's contribution to phonological typology shows how socio-historical processes can shape the prosody of phonological systems in creoles. Also, it raises an important issue at the heart of linguistic theory and creolistics, i.e., whether creoles are fundamentally different

from non-creoles. In doing so, this paper also highlights that the study of creoles in phonological typology has largely been neglected.

3.2.5 Culhane

Culhane's contribution critiques traditional typologies of foot structure, including the classifications "iambic" and "trochaic", which are centered around prominence and syllable weight. Through a pilot survey of 30 languages, the author showcases the cross-linguistic diversity of non-prominence-related evidence for foot structure, including segmental and tonal processes, phonotactics, affixation phenomena, reduplication, and word games. A major finding of this work is that much of the cross-linguistic diversity in foot structure must be explained by referencing the segmental structure of the foot. This paper is an example of current approaches in phonological typology which challenge problematic traditional classifications with evidence from diverse and understudied languages.

3.2.6 Nikolaev

Nikolaev's paper revisits the findings of Nikolaev and Grossman (2020), which used a novel statistical method to investigate co-occurrence classes, or groups of segments that tend to be found together in consonant inventories. In the current study, Nikolaev attempts to replicate the general results of Nikolaev and Grossman by using a simpler statistical method borrowed from the study of gene regulatory networks, and to refine the results by extending the analysis to macroareas. While the simpler statistical method largely replicates the earlier results regarding co-occurrence classes, the cluster structure of the cross-linguistically most frequent consonants /p t k m n l r/ (called the Basic Consonant Inventory in Nikolaev and Grossman 2020) cannot be replicated either globally or areally; instead, smaller clusters are found. This paper demonstrates how our field can benefit from both replicability studies and statistical methods developed in other disciplines.

3.3 Micro-typology

How to design the data and model the typological research question at hand are at the heart of typological research. In this volume, the contributions include:

– Khoisan Phonological Typology Database: towards a comprehensive phonological typology of the Kalahari Basin Area (Hirosi **Nakagawa**, Alena **Witzlack-Makarevich**, Daniel **Auer**, Anne-Maria **Fehn**, Linda **Gerlach**, Tom **Güldemann**,

Sylvanus **Job**, Florian **Lionnet**, Christfried **Naumann**, Hitomi **Ono** & Lee J. **Pratchett**)
- Phylotypology of implosives in Indo-Iranian (Qandeel **Hussain** & Jeff **Mielke**)
- Controlling for contact effects in phonological typology (Matías **Guzmán Naranjo** & Miri **Mertner**)
- Typology and universal tendencies in evolutionary phonology: Implications of rare sound patterns (Juliette **Blevins** & Ander **Egurtzegi**)

3.3.1 Nakagawa et al.

Nakagawa et al. contribute a paper which reports on initial results from the KBA Phonological Typology Database, the first large-scale comprehensive survey of the phonological systems of the Tuu, Kx'a, and Khoe(-Kwadi) languages of the Kalahari Basin Area (KBA) linguistic area. Here the authors focus on identifying similarities in the consonant inventories and phonotactics of these languages. They find that when a moderate cluster analysis of click structures is undertaken, the consonant inventories of KBA languages, often described as extreme outliers in terms of their size, fall more in line with other attested large inventories such as those of the Caucasus region. This study provides a quantitative analysis of the place and manner properties of consonant phonotactics in lexical roots in these languages, which are characterized by a skewed distribution of clicks and dorsal consonants in the initial onset of the root and non-dorsal consonants in other root consonants. The authors also find that synchronic frequency distributions of root structures support a hypothesis of historical simplification of bimoraic root templates in the region.

3.3.2 Hussain & Mielke

The contribution by **Hussain & Mielke** examines the distribution, place typology, and development of implosives in Indo-Aryan. In a survey of the consonant phoneme inventories of 130 Indo-Aryan languages, the authors find that 10% of these have implosive consonants. These languages, which are geographically concentrated in a hotspot around Sindh Province, are rich in velar and retroflex implosives, which are extremely rare in other implosive inventories worldwide. Using historical and phonetic supporting evidence, the authors argue that the origin of implosives in Indo-Aryan can be found in geminate consonants, and that the cross-linguistically rare change from geminates to implosives started with velars in these languages before spreading to other places of articulation. The high frequency of retroflex implosives reflects the frequency of retroflex consonants in the region in general.

This study illustrates how densely sampled family and areal studies can complement broad global surveys and thereby enrich our understanding of the mechanisms behind the evolution and cross-linguistic patterning of inventory structure.

3.3.3 Guzmán Naranjo & Mertner

In their contribution, **Guzmán Naranjo & Mertner** address the issue of how to establish contact effects in large-scale cross-linguistic studies. They propose a method for disentangling various known interdependencies in cross-linguistic data which combines three statistical techniques: Gaussian processes to identify contact effects, phylogenetic regression to identify genetic effects, and multivariate probit models to identify correlations among phonemes which tend to co-occur in inventories. The authors test their method on the consonant phoneme inventories of 645 languages of Africa, seeking to determine whether the model yields six phonological zones which have previously been proposed for the continent. Their results are promising, yielding estimated areas of phonological convergence which map quite closely to phonological areas proposed in the literature. Their approach is an important addition to the methodological toolkit of those who explore typology from quantitative and distributional perspectives.

3.3.4 Blevins & Egurtzegi

Blevins & Egurtzegi contribute a paper which includes a survey and critique of current and widely-used broad phonological databases which employ macro-typologies: broad classifications with limited encoding of phonological properties and little or no phonetic detail, and which may rely on sampling techniques that miss rare sound patterns. The authors discuss case studies of rare sound patterns: the contrast between a voiceless glottal fricative and a nasalized voiceless glottal fricative in segment inventories, the phenomenon of final obstruent voicing, and regular cluster-splitting epenthesis in word-initial sequences of an oral stop, a sonorant consonant, and a vowel (#TRV patterns). The authors use these studies to demonstrate that more targeted typologies, which involve language samples and databases constructed to capture precise language properties, are not only complementary to macro-typologies but necessary in order to empirically ground phonological theories such as Evolutionary Phonology. In addition, the case studies in this paper underline the importance of including rich phonetic detail in phonological databases, representing diverse dialects in language sampling, and carefully considering language contact histories in phonological typological work.

4 Concluding remarks

The articles in this special issue are a good representation of the current work being done in phonological typology today. Segment inventories, prosodic systems, and phonotactics are the focus of most of the papers (with the study of tone largely absent in this volume, despite its importance in phonology). The perspectives adopted include synchronic, diachronic, areal, phylogenetic, acquisitional, and theoretical-methodological. Almost all include a quantitative dimension. Some exploit existing databases and others are based on newly-curated data; some leverage sophisticated analytical techniques. All have findings that increase our knowledge. In fact, phonological typologists theorize, observe, generalize and even experiment within a roughly agreed-upon paradigm. As a result, our knowledge of the sound patterns of the world's languages and the factors that shape them is steadily and incrementally increasing, case study by case study.

But linguists thrive on new questions, of the sort that can lead to received ideas and methods being questioned or even jettisoned. Despite a narrative of progress from classical typology to present-day typology, with a possible paradigm shift from Greenberg's Question (what is a possible or impossible language?) to Nichols-Bickel's Question (what's where why?),[8] we see substantial conceptual continuity from the early days of phonological typology until today. It is not clear whether the way forward will emerge from new methods or combinations thereof, from new sources of data, or from a rethinking of foundational concepts like 'phonological segment' or 'syllable.' Will our problems be solved once we can induce phonological generalizations at multiple levels of abstraction directly from speech recordings? Will an ideal multidisciplinary constellation of traditional typology, experimental phonetics, computational modeling, and other sciences, tell us what's where why? Perhaps, but it is possible that as we hone our existing tools and create new ones, truly new thoughts are out there waiting to surface and to be put into practice. It may even be, as one of *Linguistic Typology*'s editors suggests, that what's where why has been the wrong question all along; this kind of suggestion is exactly the kind of stimulus that we need, pushing us to frame new questions that orient and contextualize our research.

So is phonological typology a fish swimming incessantly forward, sometimes awake and sometimes asleep, or is it a hitherto unknown amphibian, waiting to take on a new body-plan? We do not know. But we take this opportunity to call upon practitioners of phonological typology to continue improving our data, our methods, and to develop new theories, while being awake to the possibility of asking

8 Part of which is what Grossman and Noveck (2015) called Greenberg's Second Question – 'Why do languages change the way they do?' – a question at the heart of historical linguistics.

new and unsettling questions that can break our established patterns of thought and action.

Author contributions: SM & EG conceived of the special issue. SM & SE developed the special issue. SM drafted the paper. All authors contributed to the final paper.
Research funding: Swiss National Science Foundation, Award Identifier/Grant number: PCEFP1_186841.

References

Anderson, Cormac. 2017. The place of phonetics in phonological typology. Presentation given at 12th meeting of the Association of Linguistic Typology, ANU, Canberra, Australia, 12–14 December.

Atkinson, Quentin D. 2011. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332(6027). 346–349.

Bakker, Dik. 2011. Language sampling. In Jae Jung Song (ed.), *Handbook of linguistic typology*, 100–127. Oxford: Oxford University Press.

Bateman, Nicoleta. 2007. *A crosslinguistic investigation of palatalization*. San Diego: University of California PhD dissertation.

Bateman, Nicoleta. 2011. On the typology of palatalization. *Language and Linguistics Compass* 5. 588–602.

Bell, Alan. 1978. Language samples. In Joseph H. Greenberg (ed.), *Universals of human language*, *volume 1: Method and theory*, 123–156. Stanford: Stanford University Press.

Bergstrom, Carl T. & Jevin D. West. 2021. *Calling bullshit: The art of skepticism in a data-driven world*. New York: Random House.

Bickel, Balthasar. 2007. Typology in the 21st century: Major current developments. *Linguistic Typology* 11. 239–251.

Bickel, Balthasar. 2014. Linguistic diversity and universals. In Nick J. Enfield, Paul Kockelman & Jack Sidnell (eds.), *The Cambridge handbook of linguistic anthropology*, 101–124. Cambridge: Cambridge University Press.

Bickel, Balthasar. 2015. Distributional typology: Statistical inquiries into the dynamics of linguistic diversity. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 2nd edn., 901–923. Oxford: Oxford University Press.

Bickel, Balthasar. 2017. Areas and universals. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 40–54. Cambridge: Cambridge University Press.

Bickel Balthasar & Joanna Nichols. 2020. There is no significant typological difference between hunter-gatherer and other languages. In Tom Güldemann, Patrick McConvell Rhodes Richard (eds.), *The language of hunter-gatherers: Global and historical perspectives*, 67–75. Cambridge: Cambridge University Press.

Black, Alan W. 2019. CMU wilderness multilingual speech dataset. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5971–5975. Brighton, UK: IEEE.

Blasi, Damian E, Steven Moran, Scott R. Moisik, Paul Widmer, Dan Dediu & Balthasar Bickel. 2019. Human sound systems are shaped by post-neolithic changes in bite configuration. *Science* 363. 20200198.

Blevins, Juliette. 2017. Areal sound patterns. In Raymond Hickey (ed.), *The handbook of areal linguistics*, 88–121. Cambridge: Cambridge University Press.

Clements, George N. 1990. The role of the sonority cycle in core syllabification. In John Kingston & Mary Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*, vol. 1, 283–333. Cambridge: Cambridge University Press.

Clements, George N. 2003. Feature economy in sound systems. *Phonology* 20. 287–333.

Clements, George N. 2009. The role of features in phonological inventories. In Eric Raimy & Charles Cairns (eds.), *Contemporary views on architecture and representations in phonology*, 19–68. Cambridge, MA: MIT Press.

Cohen Priva, Uriel, Emily Strand, Shiying Yang, William Mizgerd, Abigail Creighton, Justin Bai, Rebecca Mathew, Allison Shao, Jordan Schuster, & Daniela Wiepert. 2021. The cross-linguistic phonological frequencies (XPF) corpus. Available at: https://www.urielcohenpriva.com/resources/xpf_manual07.pdf.

Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology*. Chicago: University of Chicago Press.

Cotterell, Ryan & Jason Eisner. 2017. Probabilistic typology: Deep generative models of vowel inventories. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, vol. 1: Long Papers*, 1182–1192. Vancouver: Association for Computational Linguistics.

Coupé, Christophe, Egidio Marsico & François Pellegrino. 2009. Structural complexity of phonological systems. In François Pellegrino, Egidio Marsico, Ioana Chitoran & Christophe Coupé (eds.), *Approaches to phonological complexity*, 141–169. Berlin & New York: Mouton de Gruyter.

Creanza, Nicole, Merritt Ruhlen, Trevor J. Pemberton, Noah A. Rosenberg, Marcus W. Feldman & Sohini Ramachandran. 2015. A comparison of worldwide phonemic and genetic variation in human populations. *Proceedings of the National Academy of Sciences* 112. 1265–1272.

Crothers, John H., James P. Lorentz, Donald A. Sherman & Marilyn M. Vihman. 1979. *Handbook of phonological data from a sample of the world's languages: A report of the Stanford Phonology Archive*. Stanford: Department of Linguistics, Stanford University.

Cysouw, Michael, Dan Dediu & Steven Moran. 2012. Comment on 'Phonemic diversity supports a serial founder effect model of language expansion from Africa'. *Science*, 335(6069). 657.

Dediu, Dan, Rick Janssen & Scott R. Moisik. 2017. Language is not isolated from its wider environment: Vocal tract influences on the evolution of speech and language. *Language and Communication* 54. 9–20.

Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13. 257–292.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68. 81–138.

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: https://wals.info.

Dunbar, Ewan & Emmanuel Dupoux. 2016. Geometric constraints on human speech sound inventories. *Frontiers in Psychology* 7. 1061.

Easterday, Shelece. 2019. *Highly complex syllable structure: A typological and diachronic study*. (Studies in Laboratory Phonology 9). Berlin: Language Science Press.

Everett, Caleb. 2013. Evidence for direct geographic influences on linguistic sounds: The case of ejectives. *PLoS One* 8. e65275.

Everett, Caleb. 2017. Languages in drier climates use fewer vowels. *Frontiers in Psychology*, 8. 70–84.

Everett, Caleb, Damián E. Blasi & Seán G. Roberts. 2015. Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences* 112. 1322–1327.

Everett, Caleb, Damián E. Blasi & Seán G. Roberts. 2016. Language evolution and climate: The case of desiccation and tone. *Journal of Language Evolution* 1. 33–46.

Gordon, Matthew K. 2016. *Phonological typology*. Oxford: Oxford University Press.

Greenberg, Joseph H., Charles A. Ferguson & Edith Moravcsik (eds.). 1978. *Universals of human language, Volume 2: Phonology*. Stanford University Press.

Grossman, Eitan & Ira Noveck. 2015. What can historical linguistics and experimental pragmatics offer each other? *Linguistics Vanguard* 1(1). 145–152.

Grossman, Eitan, Elad Eisen, Dmitry Nikolaev & Steven Moran. 2020. SegBo: A database of borrowed sounds in the world's languages. In *Proceedings of the 12th Conference on Language Resources and Evaluation*, 5316–5322. Marseille: European Language Resources Association. Available at: https://www.aclweb.org/anthology/2020.lrec-1.654.

Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2022. *Glottolog 4.6*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.

Heath, Jeffrey. 2016. Type-ology or typ-ology? *Linguistic Typology* 20(3). 479–495.

Hockett, Charles F. 1955. *A manual of phonology*. Baltimore: Indiana University.

Hockett, Charles F. 1967. The quantification of functional load. *Word* 23(1–3). 300–320.

Hyman, Larry M. 2007. Where's phonology in typology? *Linguistic Typology* 11. 265–271.

Hyman, Larry M. 2008. Universals in phonology. *The Linguistic Review* 25. 83–137.

Hyman, Larry M. 2014. What is phonological typology? In *UC Berkeley Phonology Lab Annual Report*, 101–118. Available at: https://escholarship.org/uc/item/3hr9r9cd.

Hyman, Larry M. 2018. What is phonological typology? In Larry M. Hyman & Frans Plank (eds.), *Phonological typology*, 1–20. Berlin/Boston: De Gruyter Mouton.

Hyman, Larry M. & Frans Plank (eds.). 2018. *Phonological typology*. Berlin/Boston: De Gruyter Mouton.

Jakobson, Roman. 1941. *Kindersprache, Aphasie und allgemeine Lautgesetze [Child language, aphasia and phonological universals]*. Translated by Allan R. Keiler in 1968. The Hague: Mouton Publishers.

Janssen, Dirk, Balthasar Bickel & Fernando Zúñiga. 2006. Randomization tests in language Typology. *Linguistic Typology* 10. 419–440.

Kiparsky, Paul. 2018. Formal and empirical issues in phonological typology. In Larry M. Hyman & Frans Plank (eds.), *Phonological typology*, 54–106. Berlin/Boston: De Gruyter Mouton.

Mackie, Scott & Jeff Mielke. 2011. Feature economy in natural, random, and synthetic inventories. In Nick G. Clements & Rachid Ridouane (eds.), *Where do phonological features come from? Cognitive, physical and developmental bases of distinctive speech categories*, 43–63. Amsterdam: John Benjamins.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Maddieson, Ian. 2018. Is phonological typology possible without (universal) categories? In Larry Hyman & Frans Plank (eds.), *Phonological typology*, 107–125. Berlin/Boston: Walter de Gruyter.

Maddieson, Ian, Tanmoy Bhattacharya, D. Eric Smith, and William Croft. 2011. Geographical distribution of phonological complexity. *Linguistic Typology*, 15(2), 267–279.

Maddieson, Ian, Sébastien Flavier, Egidio Marsico, Christophe Coupé & François Pellegrino. 2013. LAPSyD: Lyon-Albuquerque phonological systems database. In *Proceedings of Interspeech 2013*, 25–29, Lyon, France.

Marsico, Egidio. 1999. What can a database of proto-languages tell us about the last 10,000 years of sound changes? In *Proceedings of the XIVth International Congress of Phonetic Sciences*, 353–356, San Francisco, CA, 1–7 August 1999.

Marsico, Egidio, Sebastien Flavier, Annemarie Verkerk & Steven Moran. 2018. BDPROTO: A database of phonological inventories from ancient and reconstructed languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 1654–1658. May 7–12, Miyazaki, Japan. Available at: http://www.lrec-conf.org/proceedings/lrec2018/pdf/534.pdf.

Martinet, André. 1955. *Économie des changements phonétiques*. Berne: A. Francke.

McNew, Garland, Curdin Derungs & Steven Moran. 2018. Towards faithfully visualizing global linguistic diversity. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 805–809, May 7–12, Miyazaki, Japan. https://aclanthology.org/L18-1129/.

Michael, Lev, Tammy Stark, Emily Clem & Will Chang. 2015. *South American phonological inventory database v2.1.0. Survey of California and other Indian languages digital resourc*e. Berkeley: University of California. Available at: https://linguistics.berkeley.edu/saphon/en/.

Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Mielke, Jeff. 2013. P-base. Available at: https://pbase.phon.chass.ncsu.edu.

Moran, Steven & Daniel McCloy (eds.). 2019. *PHOIBLE 2.0*. Jena: Max Planck Institute for the Science of Human History. Available at: https://phoible.org.

Moran, Steven, Daniel McCloy & Richard Wright. 2012. Revisiting population size vs. phoneme inventory size. *Language* 88(4). 877–893.

Moran, Steven, Eitan Grossman and Annemarie Verkerk. 2020. Investigating diachronic trends in phonological inventories using BDPROTO. *Language Resources and Evaluation*, 55. 79–103.

Nettle, Daniel. 1999. *Linguistic diversity*. Oxford, UK: Oxford University Press.

Newmeyer, Frederick J. 2010. On comparative concepts and descriptive categories: A reply to Haspelmath. *Language* 86(3). 688–695.

Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.

Nikolaev, Dmitry. 2018. The database of Eurasian phonological inventories: A research tool for distributional phonological typology. *Linguistics Vanguard* 4(1). 20170050.

Nikolaev, Dmitry & Eitan Grossman. 2018. Areal sound change and the distributional typology of affricate richness in Eurasia. *Studies in Language* 42(3). 562–599.

Nikolaev, Dmitry & Eitan Grossman. 2020. Consonant co-occurrence classes and the feature-economy principle. *Phonology* 37(3). 419–451.

Pericliev, Vladimir. 2008. Implicational phonological universals. *Folia Linguistica* 42(1–2). 195–225.

Pimentel, Tiago, Brian Roark & Ryan Cotterell. 2020. Phonotactic complexity and its trade-offs. *Transactions of the Association for Computational Linguistics* 8. 1–18.

Rijkhoff, Jan & Dik Bakker. 1998. Language sampling. *Linguistic Typology* 2. 263–314.

Roberts, Sean G. 2018. Robust, causal, and incremental approaches to investigating linguistic adaptation. *Frontiers in Psychology* 9. 20–40.

Round, Erich. 2019. *Australian phonemic inventories contributed to PHOIBLE 2.0: Essential explanatory notes (1.0)*. Geneva: Zenodo.

Salesky, Elizabeth, Eleanor Chodroff, Tiago Pimentel, Matthew Wiesner, Ryan Cotterell, Alan W. Black & Jason Eisner. 2020. A corpus for large-scale phonetic typology. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4526–4546. Association for Computational Linguistics.

Sapir, Edward. 1925. Sound patterns in language. *Language* 1(2). 37–51.

Seifart, Frank, Ludger Paschen & Matthew Stave (eds.). 2022. *Language Documentation Reference Corpus (DoReCo) 1.1*. Berlin & Lyon: Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

Sherman, Donald. 1975. Stop and fricative systems: A discussion of paradigmatic gaps and the question of language sampling. In *Working Papers on Language Universals*, vol 17. 1–31. Stanford University.

Sherman, Don & Marilyn Vihman. 1972. The language universals phonological archiving project: 1971–1972. *Working Papers on Language Universals*, vol. 9. 163–179. Stanford University.

Simpson, Adrian P. 1999. Fundamental problems in comparative phonetics and phonology. Does UPSID help to solve them? In *Proceedings of the XIVth ICPhS*, San Francisco, CA.

Sproat, Richard. 2011. Phonemic diversity and the out-of-Africa theory. *Linguistic Typology* 15(2). 199–206.

Stolz, Thomas & Nataliya Levkovych. 2021. *Areal linguistics within the phonological atlas of Europe. Loan phonemes and their distribution*. Berlin: De Gruyter Mouton.

Trubetzkoy, Nikolai. 1939. *Grundzüge der Phonologie*. Prague: Travaux du cercle linguistique de Prague 7.

van der Hulst, Harry. 2017. Phonological typology. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *The Cambridge handbook of linguistic typology*, 39–77. Cambridge, UK: Cambridge University Press.

Vaux, Bert. 2009. The role of features in a symbolic theory of phonology. In Eric Raimy & Charles E. Cairns (eds.), *Contemporary views on architecture and representations in phonology*, 75–98. Cambridge, MA: MIT Press.

Wang, Chuan-Chao, Qi-Liang Ding, Huan Tao & Hui Li. 2012. Comment on 'Phonemic diversity supports a serial founder effect model of language expansion from Africa'. *Science* 335(6069). 657.

Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 34–57. Cambridge University Press.

Ylonen, Tatu. 2022. Wiktextract: Wiktionary as machine-readable structured data. In *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, 1317–1325, Marseille, France.