SegBo: A Database of Borrowed Sounds in the World's Languages

Eitan Grossman[1], Elad Eisen[1], Dmitry Nikolaev[2], Steven Moran[3]

The Hebrew University of Jerusalem[1], University of Stockholm1[2], University of Zurich[3]

Jerusalem, Israel[1], Stockholm, Sweden[2], Zurich, Switzerland[3]

{eitan.grossman, elad.eisen}@mail.huji.ac.il, dmitry.nikolaev@ling.su.se, steven.moran@uzh.ch

Abstract

Phonological segment borrowing is a process through which languages acquire new contrastive speech sounds as the result of borrowing new words from other languages. Despite the fact that phonological segment borrowing is documented in many of the world's languages, to date there has been no large-scale quantitative study of the phenomenon. In this paper, we present SEGBO, a novel cross-linguistic database of borrowed phonological segments. We describe our data aggregation pipeline and the resulting language sample. We also present two short case studies based on the database. The first deals with the impact of large colonial languages on the sound systems of the world's languages; the second deals with universals of borrowing in the domain of rhotic consonants.

Keywords: linguistic database, language contact, typology, phonology

1. Overview

Phonological segment borrowing is a process through which languages acquire new speech sounds as the result of borrowing new words from other languages. All spoken languages have sound systems that are composed of a set of phonological segments (consonants and vowels) that are contrastive, i.e., they create lexical or grammatical distinctions, as well as suprasegmentals (e.g., stress, tone, intonation), which may also be linguistically contrastive. All languages can acquire new phonological segments in a number of ways. One well-known, but hitherto under-explored way a language can acquire new phonological segments, is via lexical borrowing.

Despite the fact that phonological segment borrowing is documented in many of the world's languages (e.g., as described by linguists in grammars and phonological descriptions), to date there has been no large-scale quantitative study of this phenomenon. This is mainly due to the lack of a cross-linguistic database that documents observations of phonological segment borrowing in the world's languages.

In this paper, we present SEGBO, a new cross-linguistic database of borrowed phonological segments, which includes information about the languages that borrow (and lend) their speech sounds (Grossman et al., 2020). First we describe how we create this open-source resource from different input data sources and then we describe its current worldwide coverage. Finally, to illustrate the research viability of SEGBO, we present two brief case studies that use the database. The first deals with the impact of large colonial languages[1] on the sound systems of the world's languages; the second describes universals of borrowing in the domain of rhotic consonants, i.e., r-sounds.

---

[1]The term 'colonial' may be arguable for some cases; what we mean is languages that spread out considerably from their original homelands and are in some sense socio-politically dominant with respect to local indigenous languages.

2. Background

All languages are influenced to some extent by the languages with which they are in contact. Furthermore, no domain of language is impervious to contact-induced change: semantics, lexicon, morphology, syntax, pragmatics and discourse, and phonology, can all be influenced by language contact (Matras, 2009). However, not all linguistic properties are equally likely to diffuse from one language to another (Haugen, 1950; Thomason and Kaufman, 1988; Muysken, 1999; Matras, 2009), and there is general agreement that the sociolinguistic nature of the contact situation plays a significant role in shaping the outcomes of language contact (Thomason and Kaufman, 1988).

Understanding the dynamics of language contact is crucial for understanding the historical processes that led to the current distribution of speech sounds and grammatical phenomena in the world's languages. First of all, the study of language contact sheds light on cognitive aspects of multilingualism and on the socio-historical embedding of languages. From the point of view of language typology, understanding language contact is particularly important because language contact is one of the major factors that make languages more similar (through convergence) or dissimilar (through divergence of genealogically related languages under the influence of different contact languages). When scaled up, language contact can lead to the areal patterning of linguistic properties of nearly any kind. For example, contact situations have made the Indo-European languages of western Europe more similar to each other in many respects, while making them more dissimilar to the Indo-European languages of Asia. As such, any attempt to uncover universals of language has to account for the role that language contact has played in shaping the distribution of linguistic properties in space and time. For example, Dryer (1989) argues that due to the existence of large and ancient linguistic areas, the Greenbergian word-order correlations can be understood

as linguistic universals in a narrow sense only to the extent that they are independent of areality. Subsequent research has shown that nearly all linguistic properties, and most constellations thereof, show areal patterning, which has been taken as evidence that language contact has played a pervasive role in shaping the distribution of linguistic properties in the world's languages.

However, contact-induced change has rarely been targeted as the object of cross-linguistic study in its own right. While some aspects of contact-induced change have been explored in numerous case studies, there is relatively little large-scale cross-linguistic research on most contact-induced change. Some exceptions include Haspelmath and Tadmor (2009), which provides an in-depth study of loanwords across a sample of 41 languages, and Seifart (2017), which investigates affix borrowing in a sample of 100 languages. However, there is to date no comprehensive study of contact-induced change in the domain of phonology. This paper reports on a project that addresses this lacuna in the domain of **phonological segment borrowing**.

Phonological segments are linguistically contrastive speech sounds abstracted from the speech stream, i.e., the consonants, vowels, and suprasegmentals (stress, tone, intonation) that a language employs to create lexical and grammatical contrasts (Moran, 2019). Every spoken language contains a set of segments that include both linguistically contrastive speech sounds (what linguists call *phonemes*) and non-contrastive speech sounds (*allophones*). In English, the words 'bad' and 'dad' are a minimal pair that illustrate that the speech sounds 'b' and 'd' are lexically contrastive. Compare the two different speech sounds for the different t-sounds in the words 'top' and 'stop'. The former is aspirated, i.e., it is produced with a puff of air; the latter is unaspirated. In English this distinction is not linguistically contrastive (speakers do not hear the difference in these t-sounds). In many other languages, however, an aspirated vs unaspirated sound would signal different words or grammatical functions for speakers.[2]

The field of phonological typology has uncovered many important generalizations about the structure and composition of phonological segment inventories. For example, all languages have a finite number of phonological segments in their inventories, but cross-linguistic variation is considerable, ranging from very small inventories of 11 contrastive speech sounds, as in Rotokas (North Bougainville, Papua New Guinea) and Pirahã (Mura, Brazil), to 141 in the click language !Xũ (K'xa, Southern Africa). However, most languages have 20–37 contrastive segments (Moran and McCloy, 2019). Typologists have discovered important structural factors that determine the makeup of segment inventories, such as the 'Size Predicts' generalization, i.e., the number of segments in an inventory largely determines its content: small systems recruit

few (and basic) dimensions, while larger systems entail additional (and secondary) dimensions (Maddieson, 1984). For criticism of typological studies based on phonological segments, see (Simpson, 1999; Vaux, 2009; Kiparsky, 2018).

Non-linguistic factors that shape cross-linguistic diversity in phonological inventories are a matter of intensive research in recent decades. Several studies investigate the potential effects of demography (Pericliev, 2004; Hay and Bauer, 2007; Donohue and Nichols, 2011; Moran et al., 2012; Greenhill, 2014), environment and climate (Everett, 2013; Everett et al., 2015; Maddieson, 1984; Everett et al., 2016), genetics (Dediu and Ladd, 2007; Creanza et al., 2015), geography and population dispersal (Atkinson, 2011)), culture (Labov et al., 2013; Blasi et al., 2019), and anatomy (Moisik and Dediu, 2015; Dediu et al., 2017)) on phonological inventories and their diversity and composition.

Beyond these language-external factors, individual language descriptions indicate that phonological inventories are also influenced by language contact. In Diegueño (Cuchimi-Yuman, California), for example, the sound /g/ is found only in words of foreign origin, such as *gaat* 'cat', from Spanish *gato* (Miller, 1990). In Khwarshi (Nakh-Daghestanian, Daghestan), the voiceless velar fricative /x/ is found only in loanwords from the related language Avar, such as *xul* 'intention' (Khalilova, 2009). In such cases, an influx of loanwords has led to the introduction of a new segment into the phonological inventory of the 'borrowing' language.

Lexical borrowing can also result in change of the contrastiveness of an already existing sound. For instance, [dˁ] in Alqosh Neo-Aramaic (Afro-Asiatic, Kurdistan) was an allophone of phoneme /d/ next to pharyngealized segments, but it gained phonemic status due to borrowing, and it appears also in other contexts (Coghill, 2004).

Borrowed phonemes are often classified as "marginal", or "marginally contrastive", deeming phonemic status not binary but rather gradient (Hall, 2013) (hence the cover-term segment). Therefore, we propose to use the following definition:

> **Phonological segment borrowing** is a process in which a certain sound becomes a contrastive segment in a language, or in which a marginal segment in a language becomes contrastive in more domains or environments, due to lexical borrowing.

Despite the fact that such contact-induced phonological change is rampant in languages, there has been to date no systematic cross-linguistic research on phonological segment borrowing. As a result, still to be explored is the effect of such borrowing events, ultimately due to the historically contingent events of human history, on the cross-linguistic distributions of sounds and sound patterns

---

[2]Following standard practice in linguistics, we will denote phonemes with slashes, e.g., /tʰ/ and allophones with brackets, e.g., [t].

in the world's languages.

This paper presents a research program that aims to address this lacuna with SEGBO, the first large-scale database of borrowed phonological segments based on a genealogically- and areally-dense sample, and by analyzing the data to uncover the dynamics of segment borrowing and its impact on the distribution of speech sounds in present-day languages. The main goal of SEGBO is to provide an open-source and easily accessible resource for qualitative and quantitative research, which will allow language scientists to conduct exploratory and hypothesis-driven research on phonological segment borrowing.

3. Data Extraction and Aggregation

Borrowed segments were sourced from published descriptions (grammars of languages and phonological overviews) by the authors and by linguistics students at the Hebrew University of Jerusalem, the University of Helsinki, and the Higher School of Economics (Moscow, Russia). Sounds characterized as "borrowed from language X" or "exclusively/predominantly found in loan-words" in the data source were extracted. Each language description constitutes a *doculect*, i.e., the language variety in the linguistic source description (Good and Cysouw, 2013). Sometimes multiple doculects exist for a given language: they are processed and added independently and have unique inventory IDs (see below).

All the annotators participating in the project were highly trained in phonology and typological linguistics research, which ensured high quality of extracted data. A series of checks was performed, mostly prompted by instances of ambiguous or conflicting statements in the source flagged by the annotators. Annotators' judgements were found to be sound in the majority of cases.

For convenience of adding and aggregating data sources, we use Google Sheets as our working format for data entry. Google Sheets is a web-based distributed platform with a revision history in which edits are tracked by user. We use two main spreadsheets: a METADATA table and a PHONEMES table. The METADATA table has the fields:

- InventoryID (int): a unique identifier for each data point

- Glottocode (text): a unique and stable identifier for each languoid, i.e., any language family, language, or dialect

- LanguageName (text): the language name as specified by the source (e.g., a grammar)

- BibTexKey (text): the bibliographic reference identifier (links to a BibTeX field with the full reference citation)

- Filename (text): file name of the PDF from which the contributor extracted the data

- Contributor (text): who entered the data

- Comments (text): any comments about the source being added

And the PHONEMES table contains the fields:

- InventoryID (int): foreign key to the Metadata table

- BorrowingLanguageGlottocode (text): a Glottocode indicating the language that borrowed the sound

- BorrowedSound (text): the borrowed sound

- SourceLanguageGlottocode (text): a Glottocode indicating the language from which the sound was borrowed

- OnlyInLoanwords (text): a set of options (no, yes, unknown) of whether the borrowed sound occurs *only* in loanwords

- Result (text): a set of options (new phoneme, phonologization of allophone, other distributional change, unknown) that describe the result of the sound being borrowed

- NewDistinction (text): what new distinction, if any, did the borrowed sound create in the borrowing language? For example, 'voicing in plosives' might be a new distinction created by the borrowing of a sound like /g/

- Comments (text): any comments about the sound being borrowed

We use a unique identifier (an integer) for each data source added. As such, the two tables can be combined by the INVENTORYID and are in a one-to-many relationship from metadata (doculects, i.e., linguistic source descriptions from which we extract borrowing information) to phonemes (i.e., the sounds being borrowed).

We export the data from Google Sheets as Unicode UTF-8 CSV files. In our GitHub repository,[3] we have two R scripts that: 1) merge the two SEGBO tables (METADATA and PHONEMES) and 2) merge in additional linguistic (e.g., genealogical) and non-linguistic metadata (e.g., geographic location). We also merge in additional information in the form of phonetic feature vectors from PHOIBLE (Moran and McCloy, 2019). PHOIBLE includes distinctive feature data for every phoneme in every language in its sample. A distinctive feature in linguistic theory is the most basic unit of phonological structure that describes the major articulatory (i.e., physiological) features of the speech apparatus, including the major (e.g., syllabic, consonantal), laryngeal (e.g., voice, spread glottis), manner (e.g., nasal, continuant), and place features (e.g., labial, coronal).

After we have merged the SEGBO tables with the additional linguistic and non-linguistic data sources from Glottolog[4] and PHOIBLE,[5] we run a set of software tests to check the integrity of the data in terms of factors including missing Glottocodes, Unicode-compliant segments, and

---

[3] https://github.com/segbo-db
[4] https://glottolog.org/
[5] https://phoible.org/

Africa	Australia	Eurasia	N America	Papunesia	S America
42	5	222	97	115	24

Table 1: Distribution of borrowing languages by macro-area

valid BibTeX entries. This procedure ensures that we have valid data and decreases the probability that mistakes, whether they are due to encoding or interpretation, are not introduced into the final SEGBO dataset. We make the final version available via a denormalized table in CSV format and as an R data object.

We did not document 'age of borrowing,' since this information is not given in the data sources (e.g., descriptive grammars). However, it is plausible that the majority of borrowed segments were borrowed rather recently. There are two main reasons for this. The first is that ancient borrowings are likely to be obscured by sound change or extension to native lexical items. The second is that for many data points we do have a *terminus post quem* based on the date of *initial* contact – which may in fact be considerably earlier than the date of borrowing. For example, since we know when speakers of Iberian Romance languages reached North and South America, we can be confident that this is the earliest possible date. These two considerations, taken in conjunction, point to a general date for the entire database of no more than 1000 years, and it is likely that in many (if not most) cases, the age of borrowing is much younger.

4. The Language Sample

At the time of writing, the sample includes data on 532 doculects that represent 514 different Glottocodes (languages and dialects) and 469 ISO 639-3 codes (languages) from 96 different language families. The worldwide distribution is shown in Fig. 1. The distribution of borrowing languages by macro-area is shown in Table 1. There are 144 different source languages. On average, a source language is connected to 7.6 borrowing languages. However, there is a chasm between the large source languages (such as English, Spanish, and Russian) influencing dozens of languages and local sources only influencing their immediate neighbors. The effect of this disparity on the distribution of borrowed segments is discussed in § 5.

5. Case study 1: The Effect of Large Donors

There is a huge disparity in the sample between a few large donor languages, such as English (which served as a source language for 71 target languages in the sample) and Spanish (65 target languages), and the majority of source languages, from which no more than 3 languages borrowed segments (89 languages were listed as a source language for one target language, 19 for two, and 11 for three). So-called "spheres of influence" of large donor-languages are shown in Fig. 1. As is well known, South and Central America are dominated by Spanish with a strong showing from Portuguese; Africa and Eurasia lean towards Arabic, Russian, and English; and Papunesia presents a dense

patchwork of different dominant colonial languages. Does this disparity have an undue influence on the distribution of borrowed segments, such that large donors have an outsized effect on segment borrowing on a global scale?

In order to test this question, we computed the ranking of the most-frequently borrowed segments originating from the five largest donor languages (Spanish, English, Arabic, Russian, and Indonesian) and again from all other languages. The results of this comparison are presented in Fig. 2. The results indicate that the distribution of the most commonly borrowed segments is relatively stable across large and small donors. For example, /f/ is the most commonly borrowed segment in both subsamples with slight permutations of ranks for other segments (except for /o/, which is rarely borrowed from small donors). One reason /f/ may be so heavily borrowed is due to historical-cultural pressures. It has been demonstrated that /f/ emerged late in the evolution of human speech, and became relatively frequent only after the advent and spread of food processing technologies and agriculture, because of their effect on the human bite configuration (Blasi et al., 2019). The findings of SEGBO indicate that the spread of /f/ (and other labiodental sounds) might be the result of phonological segment borrowing in most parts of the world, and its cross-linguistic frequency might therefore be even later than previously suggested. The segment /o/ requires further investigation – as do all classes of sounds. In all cases, however, historical events and other language-external factors have to be considered together with the phonetic (i.e., articulatory, acoustic, and perceptual) properties of phonological segments, as these phonetic properties have an impact on the transmission and borrowabilty of phonological segments.

To summarize, SEGBO points to two hypotheses about the role of colonial languages on the distribution of phonological segments in the world's languages. First, since many events of phonological segment borrowing are due to contact between large colonial languages and smaller indigenous languages, it may be that colonial languages have had a large-scale effect on the sound systems of the world's languages. Second, even though the effect of such languages might be large in terms of number, it may be that there is no significant difference between colonial and other languages in terms of which sounds were borrowed most often.

We now turn to our second case study, which examines a single class of phonological segments in light of the data in SEGBO.

6. Case study 2: Rhotic Borrowing

This second case study aims to examine the circumstances in which rhotic consonants are borrowed in the languages of the world. Rhotics are a group of sounds, informally called "r-sounds", and usually represented with the International Phonetic Alphabet (IPA) symbols: r, ɾ, ɹ, ʀ, ʁ, ɽ or ɻ. Rhotics vary in their manner and place of articulation, as well as in their phonation, but they are similar in their
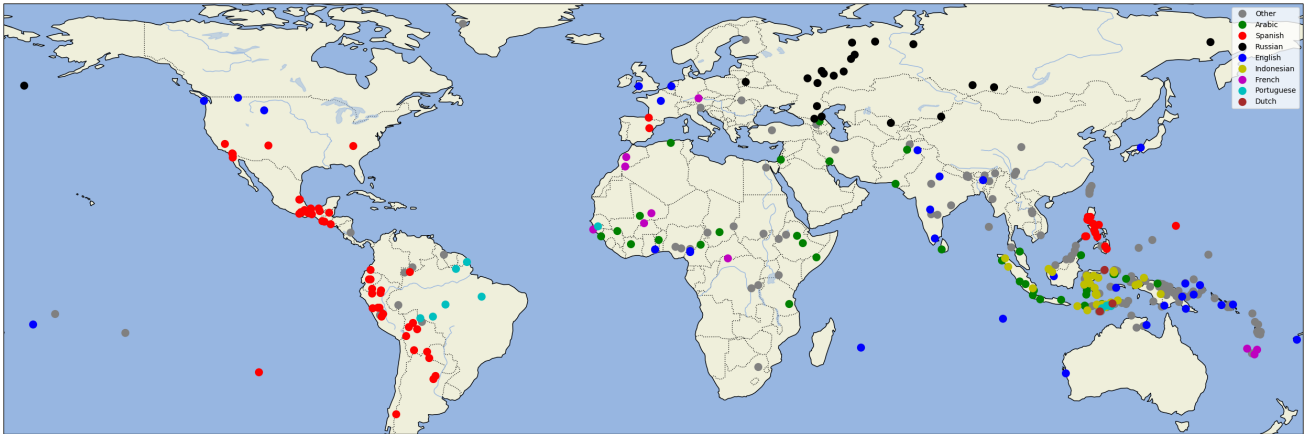
Figure 1: The geographical coverage of the SEGBO language sample. Languages are color-coded for donor languages.
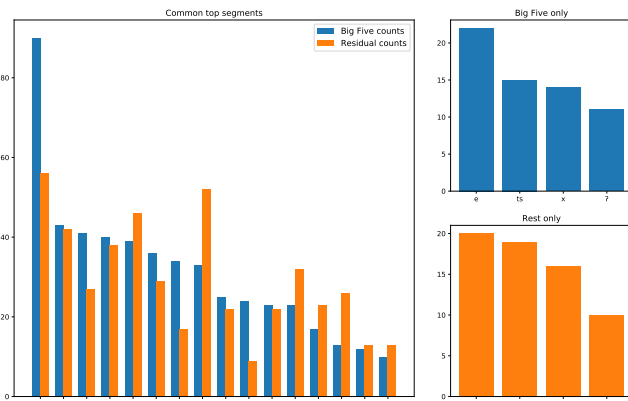

Figure 2: Rankings of segments borrowed from Spanish, English, Arabic, Russian, and Indonesian vs. all other source languages (left). Rankings of segments borrowed only from large donor vs. all other source languages (right).

# of borrowed rhotics	# of languages	Percentage
0	459	86.3%
1	66	12.4%
2	7	1.3%

Table 2: Number of borrowed rhotics in SEGBO

phonological behavior across languages (Ladefoged and Maddieson, 1996). They are quite common typologically: 80% of the languages in PHOIBLE have at least one rhotic, and more than 30% of PHOIBLE's languages differentiate between two or more rhotic segments (Moran and McCloy, 2019).

Many languages do not borrow rhotics. American English speakers, for example, use their native /ɹ/ to pronounce the Hebrew /ʁ/ in *krav maga* (an Israeli martial art), the Arabic /r/ in *Quran* and the Russian /rʲ/ in *tsar*. Some languages, however, do borrow rhotics: Hiw (Austronesian, Vanuatu) lacked a native rhotic and borrowed the Bislama (English-based creole, Vanuatu) /r/ (François, 2010). A language might borrow a rhotic even when it already has a native one: Komi-Yazva (Uralic, Russia) has native /r/, which is used to pronounce Russian /r/, but it nonetheless borrowed

the other Russian rhotic /rʲ/ (Lytkin, 1961).

On the basis of examples such as these, we hypothesize that:

> A language tends not to borrow a rhotic segment unless: (1) It has no native rhotic segment, or (2) The source language distinguishes between a rhotic native to the target language and the rhotic borrowed.

Our findings support our hypothesis: 70% of the rhotic borrowers have no native rhotic (compared to only 20% in PHOIBLE's sample). Of the languages that borrowed rhotics even though they already had a native rhotic phoneme, 83% (i.e., another 25% of the rhotic borrowers) borrowed from source languages that distinguish between the rhotic native to the target language and the rhotic borrowed. All in all, our hypothesis accounts for 95% of rhotic borrowings in our sample. This is shown in Fig. 3.
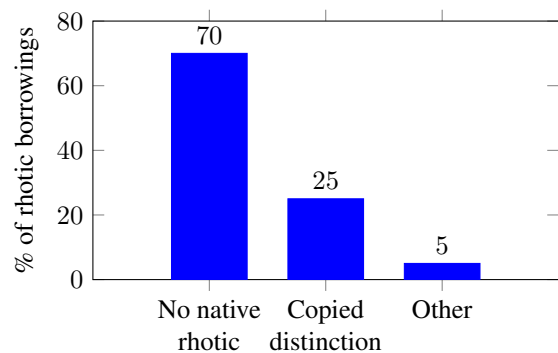

Figure 3: Rhotic borrowing in SEGBO

7. Summary

In this paper, we introduce SEGBO, a new database of borrowed speech sounds, aka phonological segments, in the world's languages. At present, our language sample consists of nearly 500 languages that have undergone phonological segment borrowing, i.e., whose phonological segment inventory has changed as the result of

language contact. We also briefly explain why investigating phonological segment borrowing – like any type of contact-induced change – is important for understanding the distribution of linguistic properties in the languages of the world. Databases like SEGBO allow language scientists to explore the dynamics of contact-induced language change with a breadth of coverage that has hitherto been inaccessible. For example, based on SEGBO, we are able to show that the borrowing of rhotics is not random, but rather follows a relatively clear pattern: languages borrow rhotics if they either had no native rhotic or if they are copying a distinction made in a contact language.

SEGBO also allows language scientists to explore the role of historical events on the distributions of phonological segments in the languages of the world. In this paper, for example, we show that while large colonial languages may have shaped the sound systems of many languages, thereby influencing cross-linguistic distributions of phonological segments, it is also possible that for the most frequently borrowed segments, there is not much of a difference between large and small donor languages.

These case studies are just the tip of the iceberg, meant to illustrate the usefulness of SEGBO for linguistic research. In future research, we will investigate these case studies in more detail and also investigate other areas of contact-induced sound change in the world's languages.

8. Acknowledgements

9. Author Contributions

All authors contributed to the data collection, analyses, and the writing of this paper.

10. Bibliographical References

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science*, 332:346–359.

Blasi, D. E., Moran, S., Moisik, S. R., Widmer, P., Dediu, D., and Bickel, B. (2019). Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science*, 363(6432):eaav3218.

Coghill, E. (2004). *The Neo-Aramaic dialect of Alqosh*. Ph.D. thesis, University of Cambridge.

Creanza, N., Ruhlen, M., Pemberton, T. J., Rosenberg, N. A., Feldman, M. W., and Ramachandran, S. (2015). A comparison of worldwide phonemic and genetic variation in human populations. *Proceedings of the National Academy of Sciences*, 112(5):1265–1272.

Dediu, D. and Ladd, R. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences*, 104(26):10944–10949.

Dediu, D., Janssen, R., and Moisik, S. R. (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.

Donohue, M. and Nichols, J. (2011). Does phoneme inventory size correlate with population size? *Linguistic Typology*, 15(2):161–170.

Dryer, M. S. (1989). Large linguistic areas and language sampling. *Studies in Language*, 13(2):257–292.

Everett, C., Blasi, D. E., and Roberts, S. G. (2015). Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences*, 112(5):1322–1327.

Everett, C., Blasi, D. E., and Roberts, S. G. (2016). Language evolution and climate: The case of desiccation and tone. *Journal of Language Evolution*, 1:33–46.

Everett, C. (2013). Evidence for direct geographic influences on linguistic sounds: The case of ejectives. *PLoS ONE*, 8(6):e65275.

François, A. (2010). Phonotactics and the prestopped velar lateral of Hiw: Resolving the ambiguity of a complex segment. *Phonology*, 27(3):393–434.

Good, J. and Cysouw, M. (2013). Languoid, doculect, and glossonym: Formalizing the notion 'language'. *Language documentation & conservation*, 7:331–359.

Greenhill, S. J. (2014). Demographic correlates of language diversity. In Claire Bowern et al., editors, *The Routledge Handbook of Historical Linguistics*, pages 557–578. Routledge.

Hall, K. C. (2013). A typology of intermediate phonological relationships. *The Linguistic Review*, 30(2):215–275.

Haspelmath, M. and Tadmor, U. (2009). *Loanwords in the world's languages: A comparative handbook*. De Gruyter Mouton.

Haugen, E. (1950). The analysis of linguistic borrowing. *Language*, 26(2):210–231.

Hay, J. and Bauer, L. (2007). Phoneme inventory size and population size. *Language*, 83:388–400.

Khalilova, Z. (2009). *A grammar of Khwarshi*. Ph.D. thesis, Rijksuniversiteit te Leiden.

Kiparsky, P. (2018). Formal and empirical issues in phonological typology. In Larry M. Hyman et al., editors, *Phonological Typology*, pages 54–106. De Gruyter Mouton.

Labov, W., Rosenfelder, I., and Fruehwald, J. (2013). One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language*, 89(1):30–65.

Ladefoged, P. and Maddieson, I., (1996). *The sounds of the world's languages*, chapter 7, pages 215–245. Blackwell Oxford.

Lytkin, V. I. (1961). *Komi-jaz'vinskij dialekt*. Izdatel'stvo Akademii Nauk SSSR. (Russian).

Maddieson, I. (1984). *Patterns of sounds*. Cambridge University Press.

Matras, Y. (2009). *Language contact*. Cambridge University Press.

Miller, A. W. (1990). *A Grammar of Jamul Diegueño*. Ph.D. thesis, University of California at San Diego.

Moisik, S. R. and Dediu, D. (2015). Anatomical biasing and clicks: Preliminary biomechanical modelling. In Hannah Little, editor, *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015) Satellite Event: The Evolution of Phonetic Capabilities: Causes, Constraints, Consequences*, pages 8–13.

Moran, S. and McCloy, D. (2019). *PHOIBLE 2.0*. Max Planck Institute for the Science of Human History, Jena.

Moran, S., McCloy, D., and Wright, R. (2012). Revisiting population size vs. phoneme inventory size. *Language*, 88(4):877–893.

Moran, S. (2019). Phonological inventories. *Oxford Research Encyclopedia of Linguistics*.

Muysken, P. (1999). Three processes of borrowing: Borrowability revisited. In Guus Extra et al., editors, *Bilingualism and Migration*, pages 229–246. Berlin/New York: Mouton de Gruyter.

Pericliev, V. (2004). There is no correlation between the size of a community speaking a language and the size of the phonological inventory of that language. *Linguistic Typology*, 8(3):376–383.

Seifart, F. (2017). Patterns of affix borrowing in a sample of 100 languages. *Journal of Historical Linguistics*, 7(3):389–431.

Simpson, A. P. (1999). Fundamental problems in comparative phonetics and phonology: Does UPSID help to solve them? In John Ohala, et al., editors, *Proceedings of the 14th International Congress of Phonetic Sciences*, volume 1, pages 349–352. University of California.

Thomason, S. G. and Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. University of California Press.

Vaux, B. (2009). The role of features in a symbolic theory of phonology. In Eric Raimy et al., editors, *Contemporary Views on Architecture and Representations in Phonology*, pages 75–97. MIT Press.

11. Language Resource References

Grossman, Eitan and Eisen, Elad and Nikolaev, Dmitry and Moran, Steven. (2020). *segbo-db/segbo: v0.0.0*. Zenodo.

Moran, Steven and McCloy, Daniel. (2019). *PHOIBLE 2.0*. Max Planck Institute for the Science of Human History.