

# *Estimating historical probabilities of natural and unnatural processes*

**Gašper Beguš**

University of California, Berkeley

---

## **Supplementary materials**

---

### **Appendix A: Lists of languages**

Languages are listed according to the sound change they feature, as reported in Kümmel (2007), and are divided according to the alternation the sound change produces. In addition, each subsection divides languages according to the number of places of articulation targeted by the relevant sound change. Languages featuring the same sound change (as reported and grouped in Kümmel 2007) are listed together, and separated from other languages by a semicolon.

#### **1 Postnasal voicing**

*Three places of articulation:* Milian, Lycian, Karian, Lydian; North-West Middle Indo-Aryan; New Indo-Aryan Dom-Lom-Rom, Sindhi, Lahnda, Punjabi, Western Pahari; Common East Iranian; Late Old Persian, Common West Middle Iranian; Middle Armenian, Common Armenian; Common Albanian; Middle Greek; Common South Italian; Spanish (Pyrenean); Scottish Gaelic, Manx; East Saami; East Saami (Skolt, Paatsjoki, Inari); Central Saami; Mari; Hungarian; Common Permic; Enets; North, Middle (dialectal), South Selkup

*Two places of articulation:* Proto-Irish; Milian, Lycian, Karian; Ormuri; New Babylonian; Late Old Umbrian; North Selkup

*One place of articulation:* Tumshuqese; Old Greek Pamphylian; Parochi, Ormuri

#### **2 Intervocalic voicing**

*Three places of articulation:* Sarmatian, Alanian, Common Ossetian; Chwaresmian; Khotanese, Tumshuqese; Pamir (Yazghulami, Sughni);

## 2 *Gašper Beguš*

Bactrian; Yidgha-Munji ; Pashto; Common North-West Middle Iranian; Middle Persian, Common South-West Iranian; Karelian (Proper, Olonets, Ludic), Ingrian (dialectal), Veps; South Estonian (dialectal), Livonian; Common North-West Middle Indo-Aryan; Common Middle Indo-Aryan; Central Italian Marche, Umbrian, Latio, New Corsican; Common West Romance; Middle Selkup; Proto-Brittonic; Manx; South Norwegian, Swedish (Bohuslän, West Västergötland); Common Old Danish; Common Permic; Kamassian, Koibalian; Pamir Sanglechi-Ishkashimi; Gorani, Azari, Sivendi; Common Nuristani; Mordvinian; Mari; South Selkup (dialectal)

*Two places of articulation:* South-West Iranian (Kumzari); North Khanty (Obdorsk); Sogdian; Common Parachi-Ormuri; Enets; Tundra Nenets; South-West Iranian (Bashkardi)

*One place of articulation:* Waxi; Common Parachi-Ormuri; Proto-Anatolian

## 3 **Final devoicing**

*Three places of articulation:* Dardic (Kashmiri); Dardic (Dameli); Nuristani (Askunu, Kati, Prasun); Common Rhaeto-Romance, Common French, Franco-Provençal, Common Occitan; Breton; Proto-Nordic; Old Low Franconian, Old High German (Middle Franconian), North Thuringian; New Dutch; Middle Low German, New Low German; Bulgarian; Serbo-Croatian West Chakavian (Istrian), Common Slovenian; Common Slovak; Russian; West Russian; Ethiopian Tigrinya; Dardic (Kashmiri); Common Catalan; Old Breton; Common Middle High German; New West Frisian

*Two places of articulation:* Sangsari

*One place of articulation:* Yaghnobi; Old Saxon; Old High German (South Rhine Franconian, East Franconian)

## 4 **Postnasal devoicing**

### 4.1 *First change*

*Three places of articulation:* Milian, Lycian, Karian; Common East Middle Iranian; Old Greek (Pamphylian, Central Krete, West Argolian, Laconia, Elis, Boiotic); Late Old Greek Hellenic; Common South Italian; Milian, Lycian; Young Avestan; Parachi; Late Old Persian, Common West Iranian; Late Parthian, Common North-West Iranian; Late Middle Persian, South-West Iranian (dialectal); Common Celtic; Common Finnic; Mari; Common Akkadian (dialectal); Middle Hebrew, Phoenician, Aramaic; East Balochi; Central Italian (Tuscany dialectal); Common Middle Indo-Aryan; New West Middle Indo-Aryan; Common Nuristani; Khotanese, Tumshuqese; Common Parachi-Ormuri; Common New Italian, Common Rhaeto-Romance; Common French, Franco-Provençal; Common Catalan, Common Spanish; Portuguese

European; Manx; Common Old Danish; East Saami; Mordvinian; Permic

*Two places of articulation:* Central Kurdish (Warmawa); Pamir Sanglechi-Ishkashimi; Ormuri; Yidgha-Munji; Ethiopic Amharic; New High German (Middle German, West Alemannic, North Middle Bavarian, East Middle Bavarian); Common North-West Middle Indo-Aryan; Common Middle Indo-Aryan; Dalmatian; Common Sardinian; Common West Romance

*One place of articulation:* Tigre, Tigrinya, Harari, Gurage; West Tocharian (dialectal); Common Middle Indo-Aryan; Ossetian; Late Baktrian; Old Greek Pamphylian; Vulgar Latin, Common Romance; New High German (Common Middle Bavarian); Central Italian (Tuscany dialectal); Common Old Swedish; Middle Dutch; Low German (dialectal); New North Frisian (Sylt, Friesland); Enets (Karassin); New High German, South and East Moselle Franconian, Hessian, Palatine; Common Albanian; Pashto; Central Kurdish (Sulaymaniyah); Umbrian; Yidgha, Munji (dialectal); Central Italian (Tuscany dialectal); New North Frisian (dialectal); New Dutch

#### 4.2 Second change

*Three places of articulation:* New Persian, North-West Iranian (dialectal); Scottish Gaelic, Manx; Common New Icelandic; Common New Danish (dialectal); Old High German (East Franconian), Upper German; Proto-Armenian; Proto-Phrygian; Proto-Germanic; Hittite, Palaic, Luwian; Old Indo-Aryan (dialectal); New High German (dialectal); Common Anatolian; Common Tocharian

*Two places of articulation:* South Italian (South Latio, North Camoanian)

*One place of articulation:* Old High German (East Franconian), Upper German; Old High German (Rhine Franconian); South Italian (East Apulia, East Sicilian dialectal)

#### 4.3 Third change

*Three places of articulation:* Slovenian Upper Carniolan (dialectal)

*Two places of articulation:* New Persian (Modern); Italian New Greek (Terra d'Otranto, Bova); Common Rhaeto-Romance; Low German (Ostfalian, Central Low Saxon); Common Old High German

*One place of articulation:* West New Aramaic; New Low German Westphalian (dialectal); New High German, South Bavarian, West Cimbrian; North-East New Aramaic (Jewish Azerbaijan); Arabic Bahrain Shia (dialectal); Ugaritic; Common South-West Iranian, Old Persian; Pamir Ishkashimi; Yagnobi; Waxi; Middle Norwegian; Middle Swedish; New Swedish; Common West Germanic; New English (dialectal); New North Frisian (Heligoland); Middle Low German, Middle Dutch; East Saami (Ter); South Saami (dialectal); Veps,

Livonian; Morvinian; South New Irish dialectal (Munster); Ormuri (Baraki Barak); Pashto (North-East, North-West); Common Old Armenian; New Greek (Bova); Occitan (dialectal); New Low German (Mecklenburg); Arabic (Egyptian, Levantine, Maghrebi, Maltese); New South Arabic (Soqotri); Common Aramaic; North-East New Aramaic; Ossetian; Albanian (Eastern dialectal); Late Old Frisian (dialectal), Common New West Frisian; New East Frisian (Saterland); New East Frisian (Wangerooge); Lombardic

## 5 Intervocalic devoicing

### 5.1 First change

*Three places of articulation:* Milian, Lycian; Young Avestan; Parachi; Late Old Persian, Common West Iranian; Late Parthian, Common North-West-Iranian; Late Middle Persian, South-West Iranian (dialectal); Common Celtic; Common Finnic; Mari; Common Akkadian (dialectal); Middle Hebrew, Phoenician, Aramaic; East Balochi; Central Italian (Tuscany dialectal); Common Middle Indo-Aryan; New West Middle Indo-Aryan; Common Nuristani; Khotanese and Tumshuqese; Common Parachi-Ormuri; Central Kurdish (Warmawa); Common New Italian, Common Rhaeto-Romance; Common French, Franco-Provençal; Common Catalan, Common Spanish; Portuguese European; Manx; Common Old Danish; East Saami; Mordvinian; Permic

*Two places of articulation:* Pamir Sanglechi-Ishkashimi; Ormuri; Yidgha-Munji; Ethiopic Amharic; New High German, West Alemannic, Nord, East Middle Bavarian; Common North-West Middle Indo-Aryan; Common Middle Indo-Aryan; Dalmatian; Common Sardinian; Common West Romance

*One place of articulation:* Tigre, Tigrinya, Harari, Gurage; West Tocharian (dialectal); Common Middle Indo-Aryan; Ossetian; Late Baktrian; Old Greek Pamphylian; Vulgar Latin, Common Romance; New High German Common Middle Bavarian; Central Italian (Tuscany dialectal); Common Old Swedish; Middle Dutch; Low German (dialectal); New North Frisian (Sylt, Festland); Enets (Karassin); New High German, South and East Moselle Franconian, Hessian, Palatine; Common Albanian; Pashto; Central Kurdish (Sulaymaniyah); Umbrian; Yidgha, Munji (dialectal); Central Italian (Tuscany dialectal); Selkup (dialectal); New North Frisian (dialectal)

### 5.2 Second change

*Two places of articulation:* New Dutch (dialectal); Spanish Aragonese Castillian, Andalusian; New Danish (South Jutland); Palatine, East Franconian, Upper Saxon (dialectal)

*One place of articulation:* Galician; Catalan (Apixtat), Ribagorza; Young Avestan, West Iranian (dialectal); New High German (Middle

Franconian); New Low German (dialectal); East and South Italian (dialectal)

### *5.3 Third change*

*Three places of articulation:* Waxi; Balochi

*Two places of articulation:* Central Italian (Tuscany dialectal)

*One place of articulation:* North-East New Aramaic; Arabic dialectal (Egyptian, Levantine, Maghrebi, Maltese); New South Arabic (Soqotri); Common Aramaic; North-East New Aramaic; Ossetian; Albanian (Eastern dialectal); Late Old Frisian dialectal, Common New West Frisian; New East Frisian (Saterland); New East Frisian (Wangerooze); Lombardic; Pamir Sanglechi; New Greek Anatolian; Central Sardinian, Logudorese; Middle Norwegian, Middle Swedish, Common Old Danish; New English (Shetland, Orkney, Manx); South Khanty, North Khanty (Nizjam, Sherkaly); Mansi; Serbo-Croatian Montenegrin; New Greek dialectal (Bova); Parochi

## **6 Final voicing**

### *6.1 First change*

*Three places of articulation:* Hebrew; Amharic; Old Saxon, Old Dutch, Old High German

### *6.2 Second change*

*Three places of articulation:* Sarmatian, Alanian, Common Ossetian; Chwaresmian; Khotanese and Tumshuqese; Bactrian; Yidgha-Munji; Pashto; Common North-West Middle Iranian; Middle Persian, Common South-West Iranian; Karelian (Proper, Olonets, Ludic), Ingrian (dialectal), Veps; South Estonian (dialectal), Livonian; Middle Selkup; South Norwegian, Swedisch (Bohuslän, West Västergötland); Common Old Danish; Kamassian, Koibalian; Gorani, Azari, Sivendi; Mordvinian; Mari; South Selkup (dialectal)

*Two places of articulation:* South-West Iranian (Kumzari); Sogdian; Common Parachi-Ormuri; Enets; South-West Iranian (Bashkardi); Tundra Nenets

*One place of articulation:* Common Parachi-Ormuri; Proto-Anatolian

### *6.3 Third change*

*Three places of articulation:* Ethiopian Gurage; Milian, Lycian; Common Dardian; Sindhi; New Greek; Common Albanian; Romanian; Common North Italian; Common Rhaeto-Romance, Common Gallo-Romance; Common Ibero-Romance; Old Irish; Common Brittonic; Danish Jutlandic, Insular; Middle English; Common New Frisian, Middle Low German, Middle Dutch; Mordvinian, Common Mari; Common Permic; Common Khanty, Mansi; Hungarian; Proto-Samoyedic; Common New High German

**Appendix B: Synchronic typology****1 Postnasal voicing**

The list of languages is based on a query for a change from [–voiced] to [+voiced] with the left environment conditioned on [+nasal] in the PBase database in Mielke (2019).

Mixe, Lowland (Coatlán variety); Maasai; Binumarien; Delaware (Unami); Pitjantjatjara/Western Desert Language; Guatuso (Maúku Jaíka); Bemba; Khmuṛ; Xhosa; Totonac, Misantla; Mixe, North Highland (Totontepec Mixe); Kui; Quichua, Ecuador (Puyo Pongo variety); Mwera; Passamaquoddy-Maliseet (Malecite-Passamaquoddy); Tamil; Muruwari; Malayalam; Ciyao (Yao); Nyangumata; Si-Luyana; Kpelle; Tiriyó (Trió); Pero (Gwandum dialectal); Quechua, Ecuadorean Highland (Bolivar dialectal); Limbu; Ojibwa, Eastern; Pero

**2 Intervocalic voicing**

The list of languages is based on a query for a change from [–voiced] to [+voiced] with the left environment conditioned on [+vowel] in the PBase database in Mielke (2019).

Af Tunni Somali (Tunni); Alabama; Ao; Auyana; Berbice Dutch Creole; Boruca; Bribri; Burmese; Danish; Efik; Estonian; Faroese; Faroese (in some districts); Guatuso (Maléku Jaíka); Kalenjin, Nandi; Kui; Kwamera; Lele; Loni; Mangap-Mbula; Martuthunira; Mikasuki; Mixe, Lowland (Coatlán variety); Mixe, Lowland (Guichicovi variety); Mixe, Lowland (San Juan el Paraíso variety); Mixe, North Highland (Totontepec Mixe); Mixe, South Highland (Mixistlán variety); Mixe, South Highland (Tepantlali variety); Mixe, South Highland (Tepuxtepec variety); Mixe, South Highland (Tlahuitoltepec variety); Mohawk; Mupun; Mupun (Jipari dialect); Ngura; Nyangumata; Ojibwa, Eastern; Oneida; Palauan; Passamaquoddy-Maliseet (Malecite-Passamaquoddy); Pech (Paya); Pero; Popoluca, Sayula; Purik; Quechua, Ecuadorean Highland (Bolivar dialectal); Senoufo, Supyire; So (Soo); Tangkhul; Tsimshian, Coast; Turkish; Tyvan (Tuvín); Xakas (Khakas)

**3 Final devoicing**

The list of languages is based on a query for a change from [+voiced] to [–voiced] with the right environment conditioned on word boundary in the PBase database in Mielke (2019).

Fe’Fe’-Bamileke; Ejagham; Slovene; Shilluk; Czech; Lithuanian; O’odham (Papago); Polish; Pero; Afrikaans; Kirghiz; Tigre; Turkish; Russian; Af Tunni Somali (Tunni); Amele; Ingessana; Boruca; Dutch; Armenian, Standard Eastern; Dhaasanac (Daasanach); Bulgarian;

Arbore; Tirmaga; Serbo-Croatian (Cres Čakavian); Wolof; Faroese; Sepečides-Romani; Slovak; Nigerian English (Nigerian Pidgin); Afar

## Appendix C: R code

The R code for the four functions listed below is available at <https://doi.org/10.1017/S0952675720000263>.

### 1 *bsc()*

The function *bsc()* takes two vectors of equal length as arguments: a vector with counts of languages with sound changes required for an alternation  $A_k$ , and a vector of languages surveyed for each sound change. The function internally transforms the vectors with counts into a binomial distribution of successes and failures for each sound change in the count. It returns R bootstrap replicates of the historical probability of  $A_1$ , computed according to (1)–(4). Stratified non-parametric bootstrapping is performed based on the *boot* package: the output of *bsc()* is an object of the class ‘boot’. The output of *bsc()* should be used as an argument of *summary.bsc()* (see §3 below), which returns the observed  $P_\omega$  and 95%  $BC_a$  CIs. Two optional arguments of *bsc()* are *order* (if True, historical probabilities are divided by  $n!$ ) and *R*, which determines the number of bootstrap replicates.

### 2 *bsc2()*

The function *bsc2()* compares the historical probabilities of two processes. It takes as input the output of *bsc()* for the process in question. The function transforms the counts into a binomial distribution of successes and failures. It returns R bootstrap replicates of the difference in historical probability between the two alternations, computed according to (1)–(5). Stratified non-parametric bootstrapping is performed based on the *boot* package: the output of *bsc2()* is an object of the class ‘boot’. The output of *bsc2()* should be used as an argument of *summary.bsc2()* (see §4 below), which returns the observed  $\Delta P_\omega$  and 95%  $BC_a$  CIs for the difference. If 95%  $BC_a$  CIs fall above or below zero, the difference is significant, and otherwise is not. Two optional arguments of *bsc2()* are *order* (if True, historical probabilities are divided by  $n!$ ) and *R*, which determines the number of bootstrap replicates.

### 3 *summary.bsc()*

The function *summary.bsc()* computes the 95%  $BC_a$  CI for the bootstrap replicates based on the *bsc()* function (see §1 above), using the *boot.ci()* function from the *boot* package and returns the observed and estimated historical probabilities. For details, see §1.

**4 *summary.bsc2()***

The function *summary.bsc2()* computes the 95%  $BC_a$  CI for the bootstrap replicates based on the *bsc2()* function (see §2 above), using the *boot.ci()* function from the *boot* package, and returns the observed and estimated differences in historical probabilities of two alternations. For details, see §1.