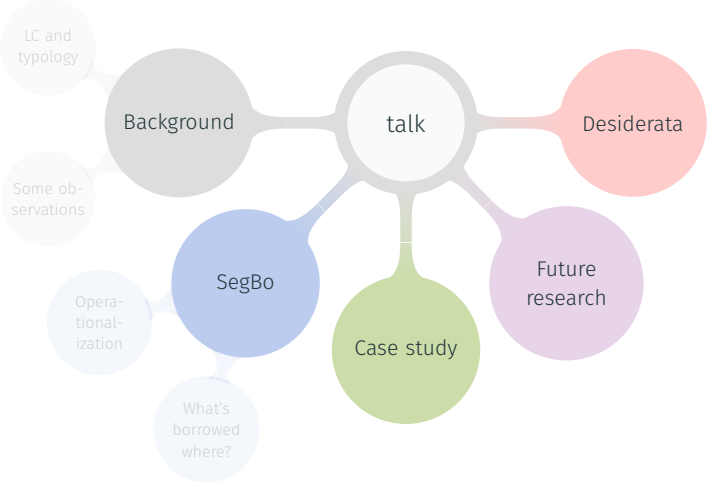# What contact typologists want from grammatical descriptions
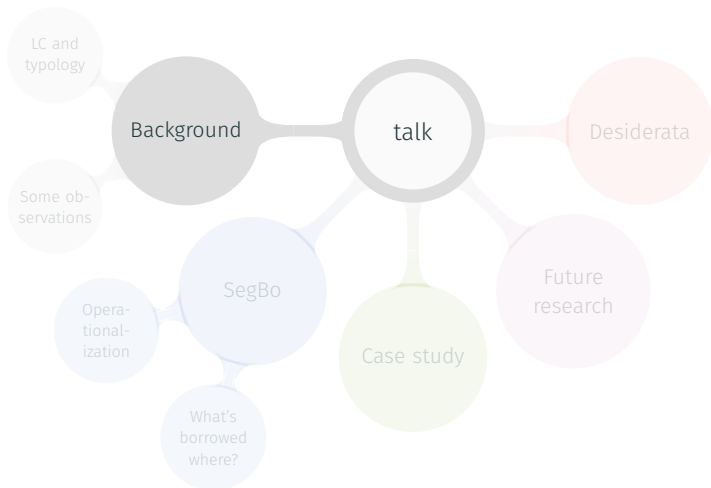
Eitan Grossman and Steven Moran
Hebrew University and University of Zurich

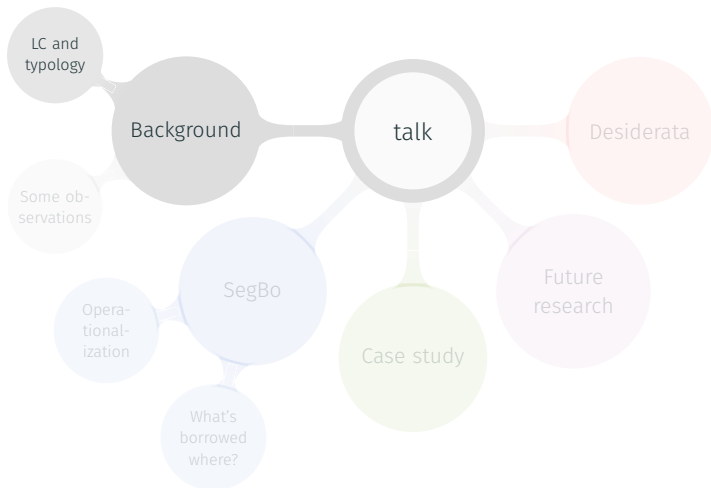Descriptive grammars and typology
Helsinki, 27–29 March 2019

The importance of language contact in typology

- All typologists have to be concerned with language contact, at least in some way.
- Why? Language contact can an lead to changes that make languages
  - more *similar* to each other
  - more *different* from each other
- This is important because typologists are interested in generalizations about cross-linguistic diversity.
  So we want to know about the sources for cross-linguistic similarities and differences.

There may be many sources for cross-linguistic (dis)similarities

- Domain-general cognitive biases or preferences
- Domain-specific cognitive biases or preferences ('Universal Grammar')
- Specific features of human anatomy
- Social and cultural factors
- Environmental factors
- Language-external events of human history (migrations, conquest, trade, etc.)

Three approaches to language contact in typology

1. Classical Greenbergian typology
2. Distributional Typology
3. Studying the typology of contact-induced change directly
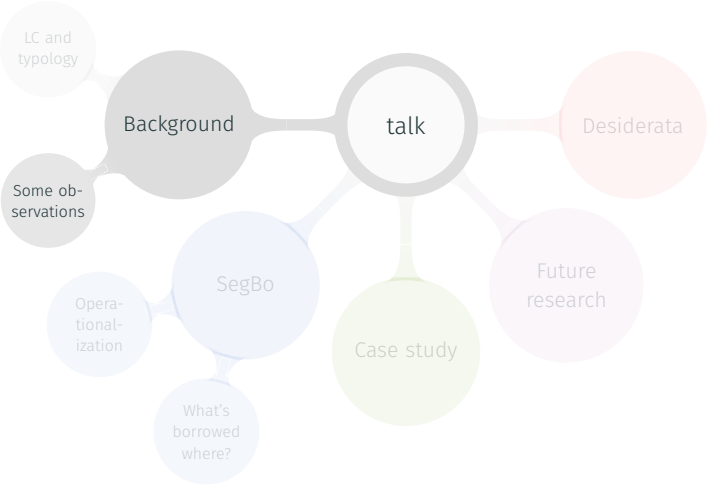
Some languages have small consonant inventories

	bilabian	alveolar	velar
voiceless	p	t	k
voiced	b	d	g

Central Rotokas (Non-Austronesian Papuan, Bougainville)

Table 7: Consonant chart, using Nakagawa's (2006) orthography

Series		Lb	Dt	Dt-Af	Dt-Af-Cl	Al-Cl	Al-Af (Lt)-Cl	Pl	Pl-Cl	Vl	Uv	Gl
		Extended Place of Articulation										
Stop segments	plain	p	t	ts	|	!	‖	(c)	ǂ	k	q	?
	voiced	b	d	dz	g|	g!	g‖	(ɟ)	gǂ	g	(ɢ ?)	
	voiceless ejective		t'	ts'	|'		‖'	(c')	ǂ'	k'		
	voiceless aspirated		tʰ	tsʰ	|ʰ	!ʰ	‖ʰ	(cʰ)	ǂʰ	kʰ		
Stop cluster	plain + x		tx	tsx	|x	!x	‖x	(cx)	ǂx			
	plain + q				|q		‖q	(qy)	ǂq			
	plain + ɢ				|ɢ		‖ɢ		ǂɢ			
	plain + ?				|?	!?	‖?	(?y)	ǂ?			
Nasal	voiced	m	n		ŋ|		ŋ‖	(ɲ)	ŋǂ	ŋ		
Pre-nasalised		mb	nd		ŋ|g	n!g	ŋ‖g	(ɲy)	ŋǂg	ŋg		
Fricative	voiceless		s								x	h
Tap or Flap			(ɾ)									
Glides		w						j				

(Abbreviations used: Lb = labial, Dt = dental, Dt-Af = dental affricate, Dt-Af-Cl = dental affricate click, Al-Cl = alveolar click, Al-Af (Lt)-Cl = alveolar affricate lateral click, Pl = palatal, Pl-Cl = palatal click, Cl = velar, Uv = uvular, Gl = glottal)

Ts'ixa (Kalahari Khoe, Fehn 2014)

8

Most languages are somewhere in the middle

	bilabial	labio-dental	alveolar	post-alveolar	palatalized	palatal	velar	glottal
plosive	p		t		t^j		k	
nasal	m		n		n^j			
trill			r					
fricative		f v	s	ʃ	s^j			h
lateral			l		l^j			
approximant						j		

Estonian (Uralic)

Cross-linguistic distributions are skewed

Clicks, labiovelar stops, and and front rounded vowels are cross-linguistically rare, while bilabials are present in nearly every language.
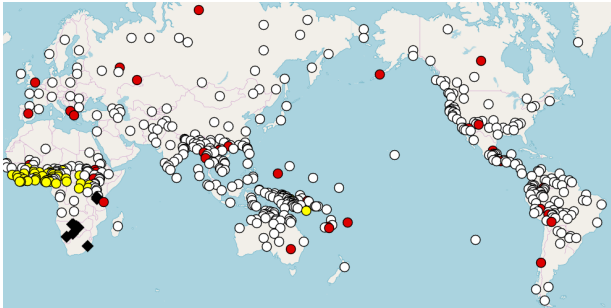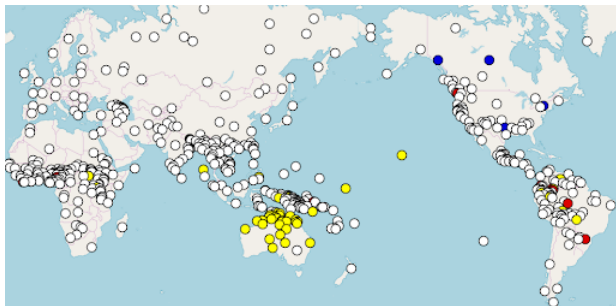

🔴 front rounded vowels   🟡 labiovelar plosives   ◆ clicks

MADDIESON 2013 (WALS Online)
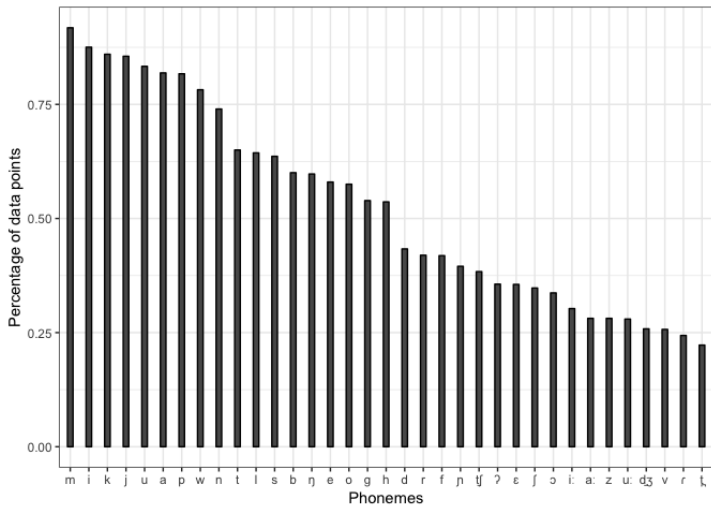
Cross-linguistic distributions are skewed

● no bilabials   ● no fricatives   ● no nasals

Maddieson 2013 (WALS Online)

Some sounds are very common

The 35 most frequent segments in PHOIBLE 2.0

tʼkxʼ dʲ̥ ntr cʼ ʔʍʲ tsʲː gǃxʼ

ʰm nd̤z̤ pkʰ pʷɣː

dẑ ʰkʲ mʷɣː cɕʼ ʂʷ

q̄ ŋǃʰ dʲ d̠ⁿ ʼɣ qχˤː pʼkxʼ

tʼː gˡ kl̩ ŋʰ ʔt̪ɬ n̥ d̠ ʔ kl̥ʔ l̩

ʃ w̩ wʲ̧ ʔņ̩ dzʲ̥ rʰ t̪ɬʼ ɲʼ tʃx

p̩ tʃʰˤ l̩ʰ sʲ bʼ dʰ r̥ nːː b̩ d̤ː

ʔd ŋm̥ ɮ̥ hᵠ ɮ ŋʲː tŝʰ mɓ

dz̩ wʼ lʼː ⁿtɕ ŋ̥ʷ ʰg qǁʼ ɭʲ kɬ nʲ

ts̩̥ qʰwʲ ʃʷː ʰwtɕ l̥ k‡ǃx nθ

ŋkx lᵠ ʷr qχʰ fʃ

- Ease of production?
- Ease of perception?
- Phonotactics?
- Phonetic naturalness (Gordon 2016)
- Properties of sound systems, e.g., symmetry, dispersion, feature economy (Clements 2009)?
- Universal grammar (e.g., markedness)?

Explanations for skewed distributions

All of these are potential functional triggers (BICKEL 2015), causal factors that can change transitional probabilities of change in a particular direction (or retention).

'Functional triggers are grounded in the biological/cognitive or social/communicative conditions of language, such as specific processing preferences (e.g. HAWKINS, 2004; CHRISTIANSEN and CHATER, 2008) or specific sociolinguistic constellations (e.g. TRUDGILL, 2011; LUPYAN and DALE, 2010) that systematically bias the way linguistic structures evolve'

(BICKEL 2017)

Synchronic present-day distributions of sound patterns.

Linguists typically assume that present-day distributions of properties can reveal insights about the nature of Language, e.g., processing explanations of word order and case-marking, or inherent markedness of sounds.

(The Uniformatian Assumption)

Some problems

1. There is no guarantee that present-day distributions tell us anything very interesting about Language (Nichols 1992, Maslova 2000, Cysouw 2005, Bickel 2007 and subs., Pantadiosi and Gibson 2013).
2. Universal areality.

The conclusion: present-day distributions are 'historically grown' (Bickel 2015, 2017).

Areal skewing of frequent sounds in PHOIBLE 2.0
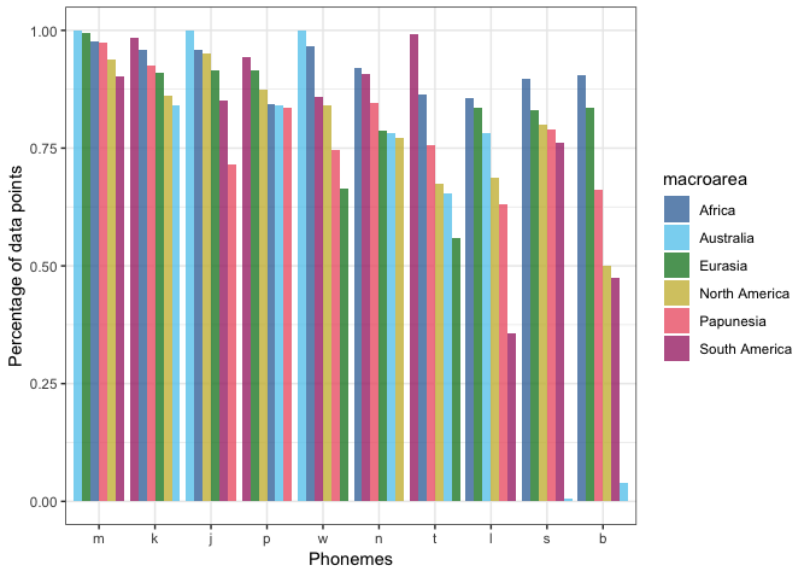
To what extent are current cross-linguistic distributions the result of events of human history?

Specifically, are some cross-linguistically frequent sounds frequent because of language contact?

Even more specifically, are some cross-linguistically frequent sounds the result of either borrowing or contact-induced sound change (e.g., BLEVINS' [2017] 'perceptual magnet')?

Consonants

	Labial	dental	Prelingual alveolar	palatal	Postlingual velar	uvular	Glottal
Stops							
voiceless	(p)	t		ch	k	q	
(voiced)	(b)	(d)			(g)		
Fricatives							
voiceless	(f)	*hd*		s	x	x̂	
voiced	(v)	d		z	g	ĝ	
Nasals							
aspirated	hm	hn			hng		
voiced	m	n			ng		
Approximants							
aspirated	hw	hl		hy			h
voiced	w	l	(r)	y			
Attuan							
Stops	(p)	t	tʳ	ch	k	q	
Fricatives	v			s	g/x	ĝ/x̂	
Nasals	m	n			ng		
Approximants		l	(r)	y			h

Consonants	Labial	dental	Prelingual alveolar	palatal	Postlingual velar	uvular	Glottal
Stops							
voiceless	(p)	t		ch	k	q	
(voiced)	(b)	(d)			(g)		
Fricatives							
voiceless	(f)	*hd*		s	x	x̂	
voiced	(v)	d		z	g	ĝ	
Nasals							
aspirated	hm	hn			hng		
voiced	m	n			ng		
Approximants							
aspirated	hw	hl		hy			h
voiced	w	l	(r)	y			
Attuan							
Stops	(p)	t	t'	ch	k	q	
Fricatives	v			s	g/x	ĝ/x̂	
Nasals	m	n			ng		
Approximants		l	(r)	y			h

Labial obstruents in Eastern and Atkan are found only in loanwords, while the Attuan *v* corresponds with E-A *-w-*, *-mg-* and preconsonantal *-m-*. The stops are bilabial, e.g. *A paltux̂* 'coat' (Russian *pal'tó*), *suupax̂* 'soup' (Russian *sup*), *baabuskax̂* 'midwife' (Russian bábushka).

In the charts, items in parentheses are found only in Russian or English loanwords, the one in italics only in Eastern Aleut and the underlined one only in Atkan and in loanwords, while the Attuan consonants are tabulated separately.
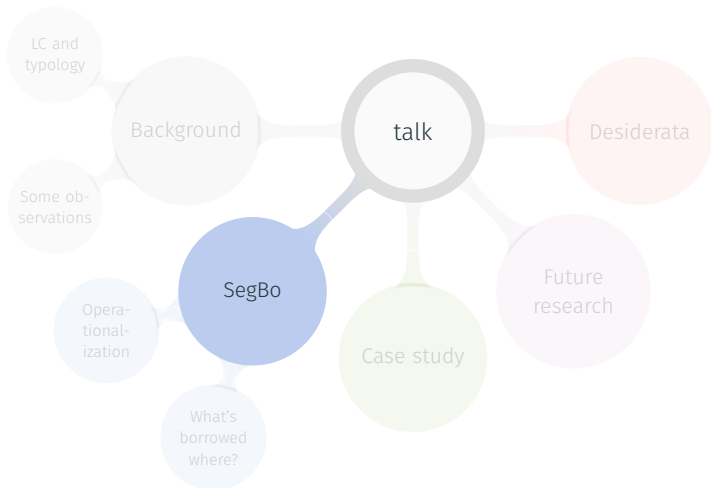
(BERGSLAND 1997)

What about event-based triggers, i.e., facts of human history that brought languages into contact...

thereby making them more similar (and possibly, more different from related languages)?

Are likely to be part of any explanatory story about cross-linguistic distributions.

LC and typology

Background

talk

Desiderata

Some observations

SegBo

Future research

Operationalization

Case study

What's borrowed where?

World Survey of Phonological Segment Borrowing [SegBo]
(GROSSMAN, NIKOLAEV, MORAN 2019+)

Currently 1573 observations (borrowed sounds), 494 borrowing languages (from 100+ families), 227 source languages, 220 unique borrowed sounds.

Based on reports of borrowed sounds in grammars and secondary literature.

Eitan Grossman, Steven Moran and Dmitry Nikolaev, as well as Elad Eisen, Einav Levanon, and other contributors from HUJI, Helsinki, and Moscow. Erich Round (Queensland) contributed Australian data.

Pre-existing data sources

- Stanford Phonology Archive (SPA) (Crothers et al. 1979)
- Marginal/borrowed sounds from PHOIBLE (MORAN et al. 2014, http://phoible.org/)
- Database of Eurasian Phonological Inventories (NIKOLAEV et al. 2015, http://eurasianphonology.info/)
- Kurdistan Phonological Inventory Database Online (NIKOLAEV et al. 2017, http://kurdistanphonology.info/)
- Plus a lot of data collection and analysis on our part!
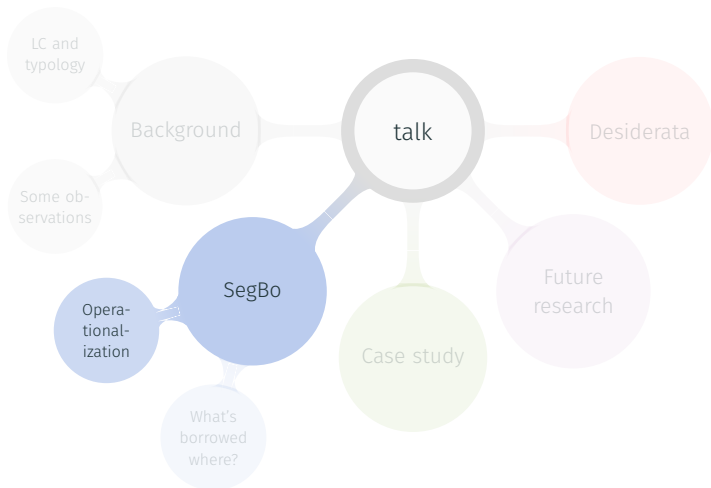
- Reports of borrowing
- Reports of segments that are limited to loanwords
- Reports of segments that are *mostly* limited to loanwords

*We're talking about 'doculects,' not 'languages.'

What are the variables?

- Classified as:
  1. new phoneme
  2. phonologization of allophone
  3. other distributional change

What are the variables?

- Classified as:
    1. new phoneme
    2. phonologization of allophone
    3. other distributional change
- New distinctions are recorded (basically, new place/manner distinctions, or other classes, e.g., laterals or rhotics)

What are the variables?

- Classified as:
    1. new phoneme
    2. phonologization of allophone
    3. other distributional change
- New distinctions are recorded (basically, new place/manner distinctions, or other classes, e.g., laterals or rhotics)
- Donor languages are recorded to the extent that they are known

What are the variables?

- Classified as:
    1. new phoneme
    2. phonologization of allophone
    3. other distributional change
- New distinctions are recorded (basically, new place/manner distinctions, or other classes, e.g., laterals or rhotics)
- Donor languages are recorded to the extent that they are known
- Other metadata have been recorded from sources, but most are extracted automatically from Glottolog
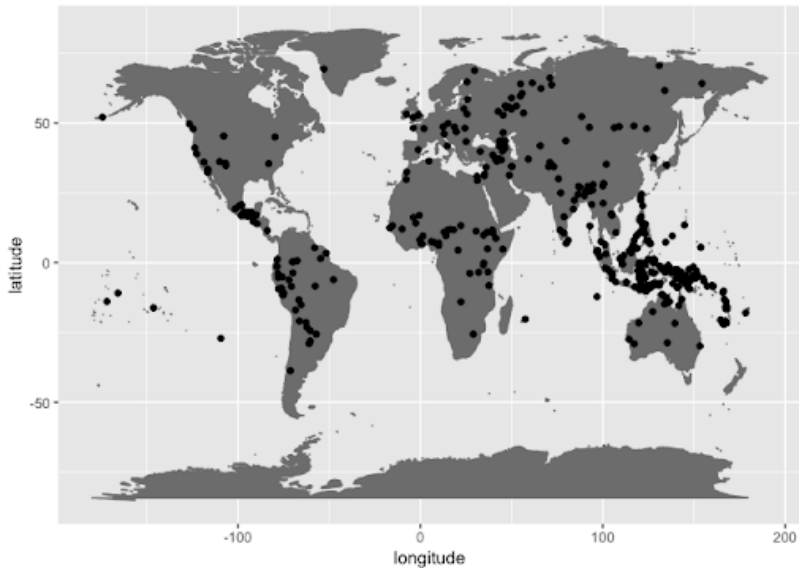
Coverage so far

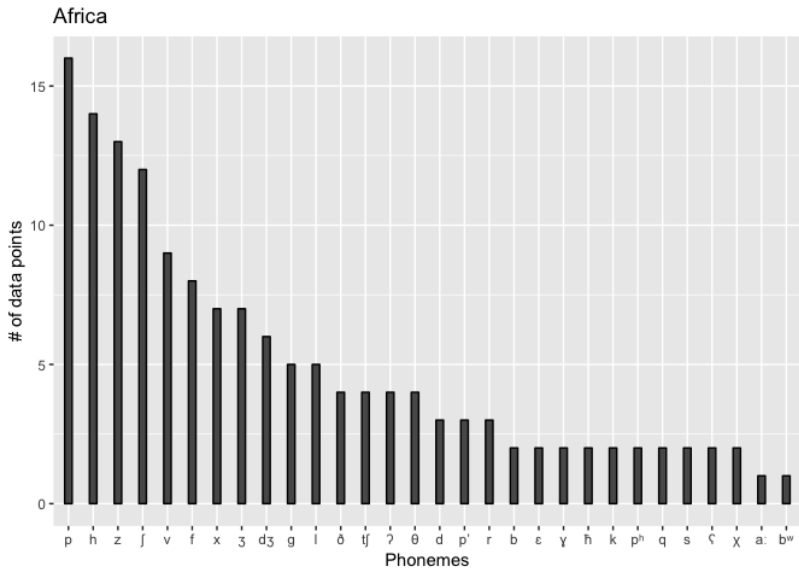| macroarea	glottolog.count	segbo.count	coverage
Africa	2338	46	0.019
Australia	383	17	0.044
Eurasia	1966	113	0.057
North America	784	40	0.051
Papunesia	2197	195	0.088
South America	703	39	0.055

What's borrowed?

34

What's borrowed?

All borrowed sounds

Africa

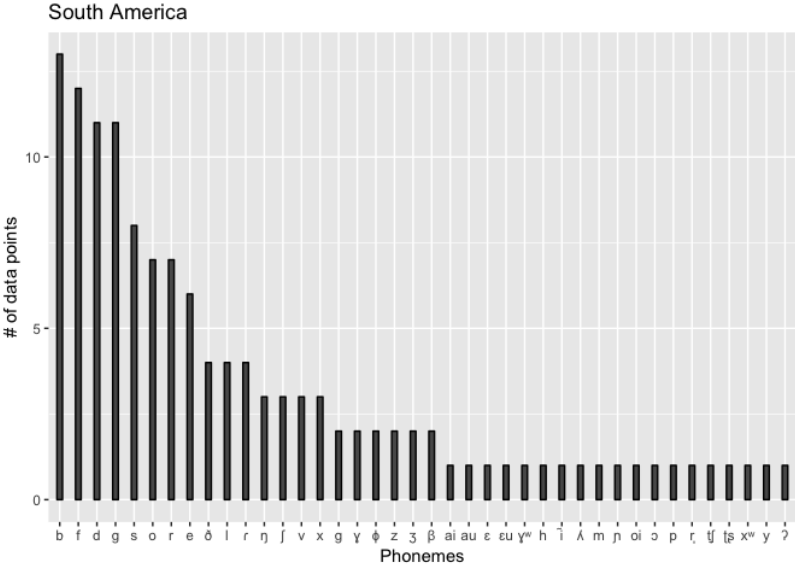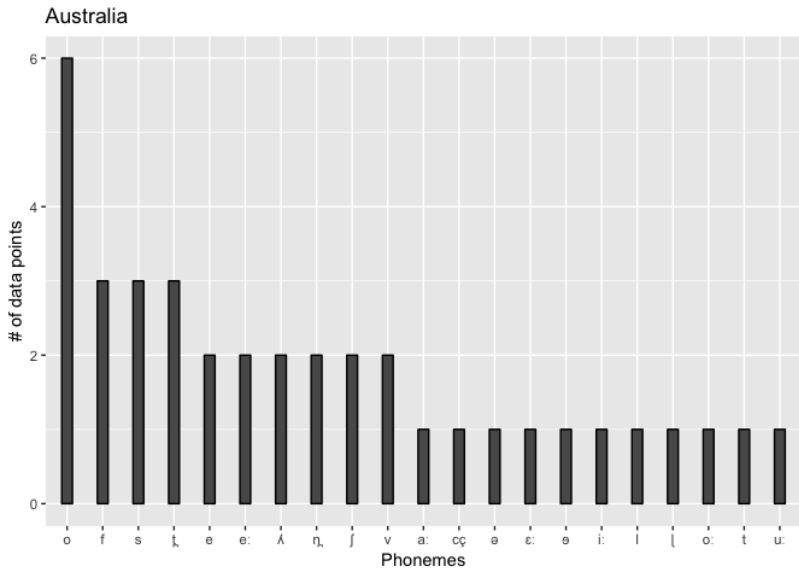Africa

Eurasia

Top 50 Eurasia
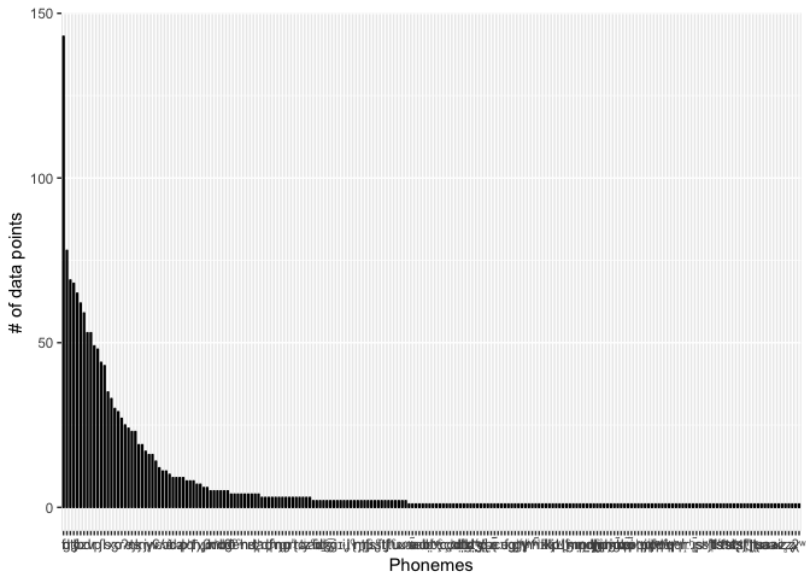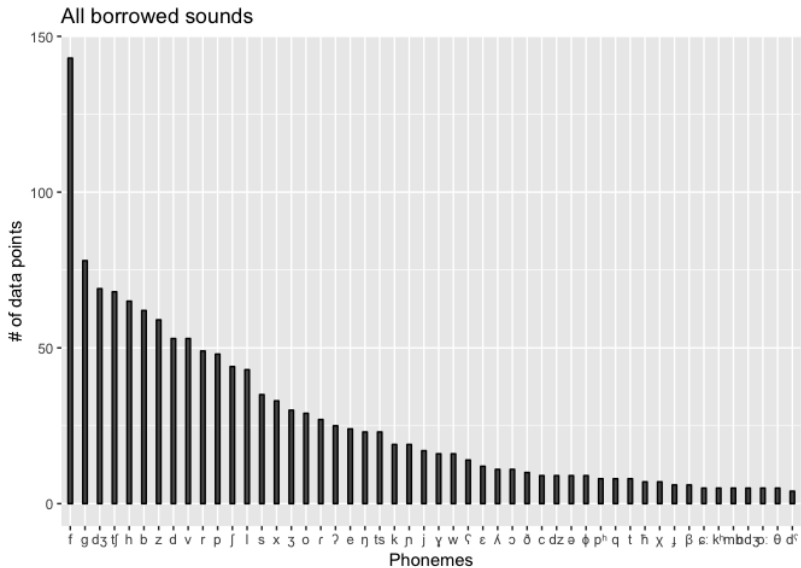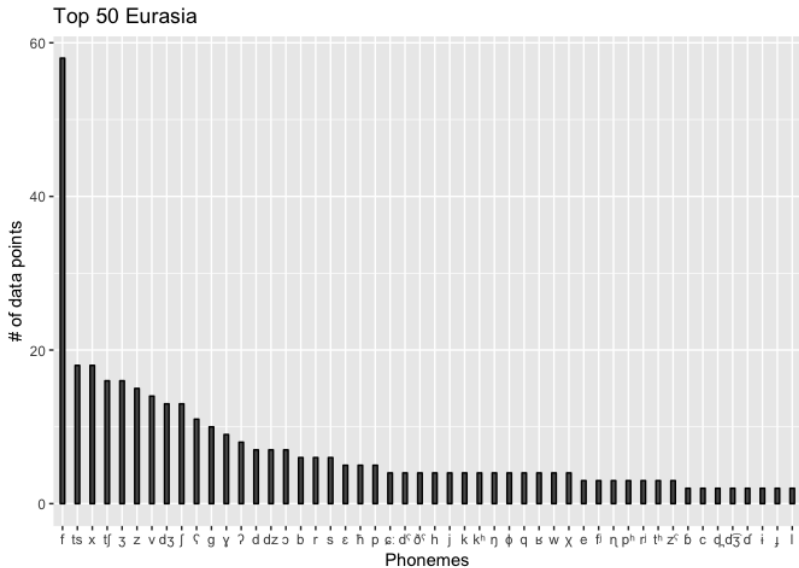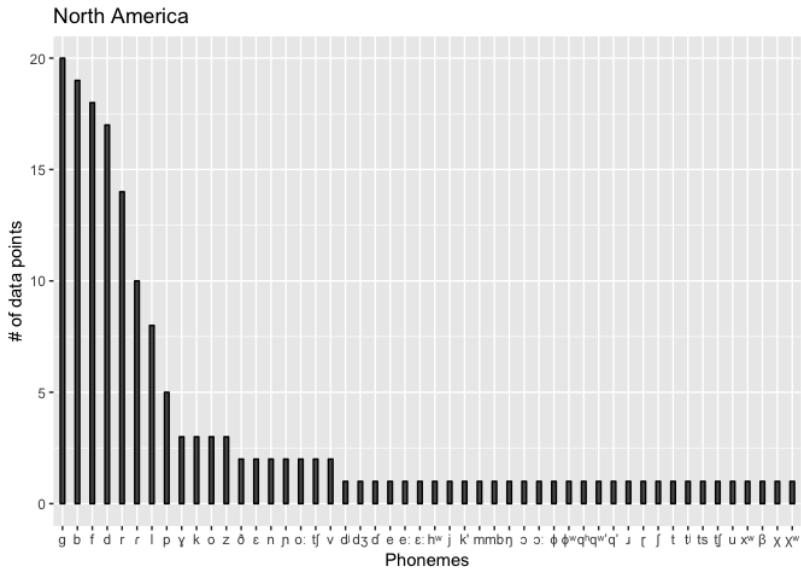
North America

South America

Australia

Some preliminary observations

- We see an overall preference for phonologically simple sounds.
- Consonants > vowels - by far!
- In many areas, the most highly borrowed sounds are fricatives, affricates, and voiced stops, as well as rhotics and laterals.
- Clear areal preferences as well, often related to 'gaps' (e.g., African /p/, Eurasian /f/, American rhotics and voiced stops, Australian mid vowels)
- Note the overwhelming predominance of /f/ borrowing nearly everywhere, cf. recent paper on the evolutionary lateness of labiodentals in human speech (Blasi, Moran et al. 2019).

- **Affricates**: Abau, Abun, Asilulu, Dusner, East Makian, Koki, Konai, Maybrat, Nimboran, and another 40 languages.
- **Approximants**: Lewotobi, Blagar, Helong, Baukeno, Laiyolo…
- **Labiodentals**: Bierebo, Madurese, Kwaza, Sundanese, Sahu, Savosavo…
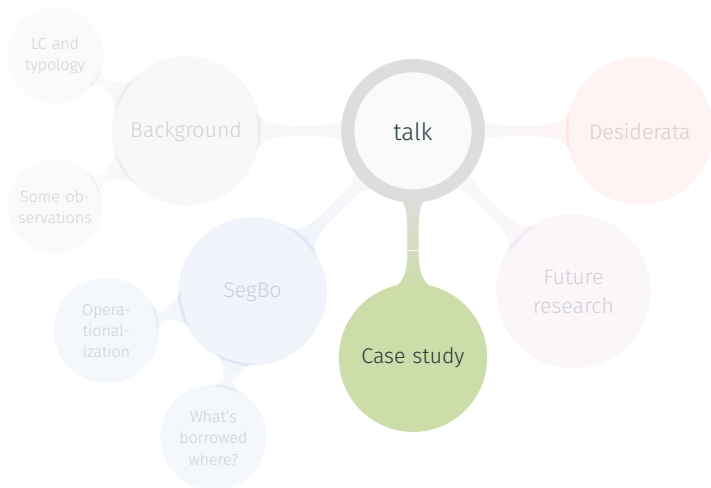
Case study:

When and why we
borrow rhotics?

[Based on work by Elad Eisen (HUJI)]

Rhotics

- Not a well-defined group of sounds
- Often represented with symbols such as r, ɾ, ɹ, ʀ, ʁ, ɽ or ɻ
- Do not share any single feature, but rather have a family resemblance (Ladefoged & Maddieson 1996)
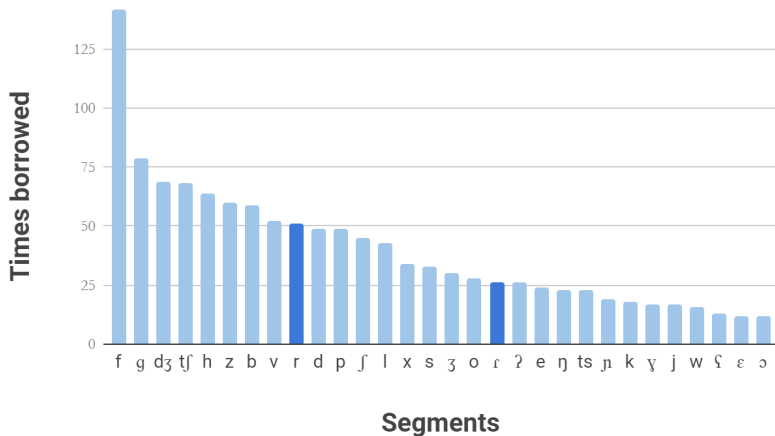
Rhotic phonemes in the world's languages

No rhotic
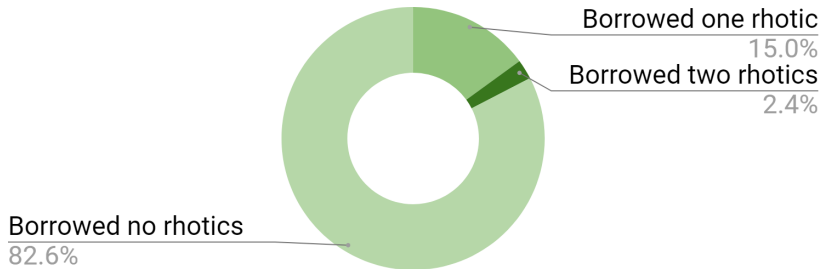19.6%

Several rhotics
32.2%
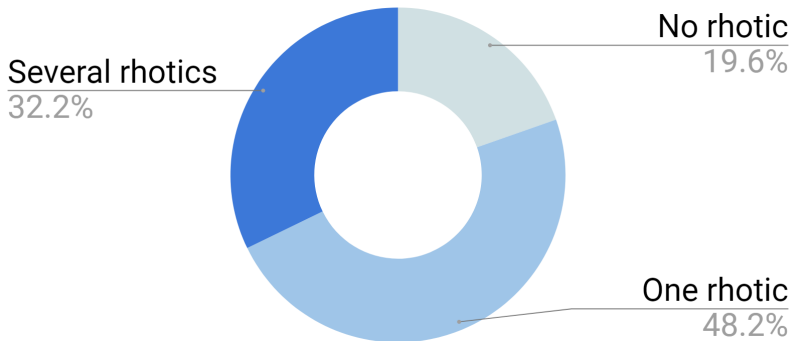
One rhotic
48.2%

Source: PHOIBLE

Rhotics in SegBo
(within 30 most frequently borrowed segments)

Rhotic borrowers in SegBo

Borrowed one rhotic
15.0%

Borrowed two rhotics
2.4%

Borrowed no rhotics
82.6%

Three examples:

- Hiw (Austronesian)
- Modern Hebrew (Semitic)
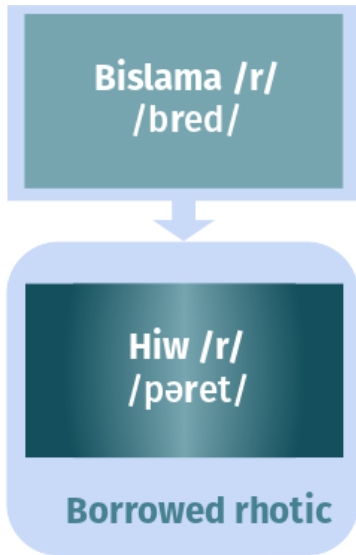- Komi-Yazva (Uralic)

The circumstances of rhotic borrowing

Three examples:

- Hiw (Austronesian)
- Modern Hebrew (Semitic)
- Komi-Yazva (Uralic)

Example 1: Hiw, rhotic borrowed

- Hiw has no native rhotic phoneme
- Borrowed /r/ from Bislama

Bislama /r/
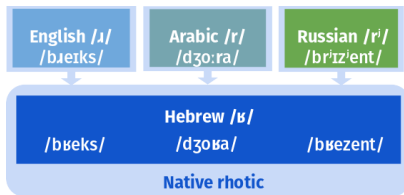**/bred/**

↓

Hiw /r/
/pəret/

Borrowed rhotic

The circumstances of rhotic borrowing

Three examples:

- Hiw (Austronesian)
- Modern Hebrew (Semitic)
- Komi-Yazva (Uralic)

- Modern Hebrew has one native rhotic phoneme, /ʁ/
- The native rhotic is used to render all foreign rhotics
- No rhotic is borrowed

The circumstances of rhotic borrowing
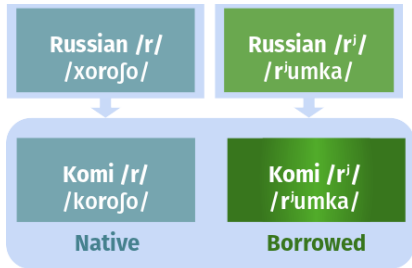
Three examples:

- Hiw (Austronesian)
- Modern Hebrew (Semitic)
- Komi-Yazva (Uralic)

- Komi-Yazva has one native rhotic, /r/
- Contact with Russian, which distinguishes between /r/ and /rʲ/
- Russian /r/ is rendered with the native /r/, but Russian /rʲ/ is borrowed

Russian /r/
/xoroʃo/

Russian /rʲ/
/rʲumka/

Komi /r/
/koroʃo/
Native

Komi /rʲ/
/rʲumka/
Borrowed

Hypothesis

A language tends not to borrow a rhotic phoneme unless:

1. it has no native rhotic phoneme (like Hiw);
   or:
2. the contact language distinguishes between a rhotic
   native to the target language and the rhotic borrowed (like
   Komi-Yazva).

The findings seem to support the hypothesis.

- A language will be more likely to borrow a foreign sound if the source language distinguishes between this sound and a similar sound that is native to the target language.

A possible generalization?

- A language will be more likely to borrow a foreign sound if the source language distinguishes between this sound and a similar sound that is native to the target language.
- That is: languages tend to borrow distinctions rather than sounds.

A possible generalization?

- A language will be more likely to borrow a foreign sound if the source language distinguishes between this sound and a similar sound that is native to the target language.
- That is: languages tend to borrow distinctions rather than sounds.
- This generalization may be tested with other distinctions, e.g. voicing.

LC and typology

Background

talk

Desiderata

Some observations

SegBo

Future research

Operationalization

Case study

What's borrowed where?
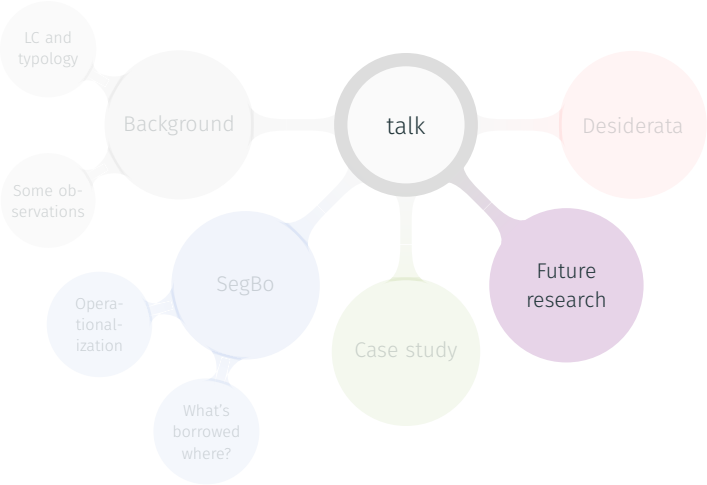
Future research: questions and directions

- Teasing apart (i) empirical frequency of borrowing and (ii) borrowability per se (a basically undefined notion).
- What factors, whether structural (e.g., phonetics, phonology) or language-external, predict the outcomes of contact?
- Individual areal histories.
- Ultimately, what is the contribution of segment borrowing to present-day distributions of segments and sound systems? (> revising assumptions about 'naturalness' based on empirical frequencies in the world's languages)
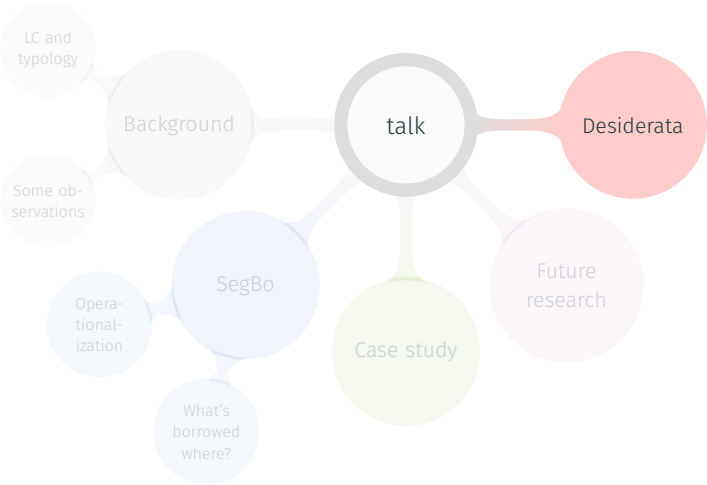
LC and typology

Background

talk

Desiderata

Some observations

Operationalization

SegBo

Case study

Future research

What's borrowed where?

What do contact typologists want from descriptive grammars?

- For them to explicitly mention language contact phenomena.

What do contact typologists want from descriptive grammars?

- For them to explicitly mention language contact phenomena.
- To discuss the contact situation as explicitly as possible: what languages, what contexts, different types of variation, etc.

What do contact typologists want from descriptive grammars?

- For them to explicitly mention language contact phenomena.
- To discuss the contact situation as explicitly as possible: what languages, what contexts, different types of variation, etc.
- For them to be explicit about sources for borrowed items, even hypothetical, possible, likely, or multiple ones.

What do contact typologists want from descriptive grammars?

- For them to explicitly mention language contact phenomena.
- To discuss the contact situation as explicitly as possible: what languages, what contexts, different types of variation, etc.
- For them to be explicit about sources for borrowed items, even hypothetical, possible, likely, or multiple ones.
- For them to use IPA, even if other transliterations/transcriptions are used. If not, then please describe the sounds (e.g., the y-/j-problem).

What do contact typologists want from descriptive grammars?

- For them to explicitly mention language contact phenomena.
- To discuss the contact situation as explicitly as possible: what languages, what contexts, different types of variation, etc.
- For them to be explicit about sources for borrowed items, even hypothetical, possible, likely, or multiple ones.
- For them to use IPA, even if other transliterations/transcriptions are used. If not, then please describe the sounds (e.g., the y-/j-problem).
- To be as explicit as possible about distributions.

What do contact typologists want from descriptive grammars?

- For them to explicitly mention language contact phenomena.
- To discuss the contact situation as explicitly as possible: what languages, what contexts, different types of variation, etc.
- For them to be explicit about sources for borrowed items, even hypothetical, possible, likely, or multiple ones.
- For them to use IPA, even if other transliterations/transcriptions are used. If not, then please describe the sounds (e.g., the y-/j-problem).
- To be as explicit as possible about distributions.
- Sample texts with borrowed items highlighted.

There is currently no usable proposal for **a typology of sociolinguistic parameters relevant for language contact** that can be used off-the-shelf or adapted for typology or fieldwork.

A really important desideratum.

Thanks! Kiitos!

Three approaches to language contact in typology

1. Classical Greenbergian typology
2. Distributional Typology
3. Studying the typology of contact-induced change directly

Typically aims to establish cross-linguistic generalizations on the basis of language samples.

Samples are often balanced or stratified with respect to area (as a proxy for contact-induced similarity) and for family (as a proxy for inheritance-based similarity).

Typologists are often concerned about how to deal with areal biases (Bell 1978, Bakker 2011).

Problems with the classical approach

The main problem is the representativeness of the current population of the world's languages vis-a-vis 'natural human language' in general.

Moreover, it has been claimed that there are not enough genetically- and areally-independent cases to allow statistical testing of universals (PIANTADOSI and GIBSON 2013).

It has been argued that inheritance and areality (as a proxy for contact-induced similarity) should not be treated as confounds.

Rather, we should try to target them directly in typological research (e.g., NICHOLS 1992, 1998; BICKEL 2007, 2012, 2015).

Three approaches to language contact in typology

1. Classical Greenbergian typology
2. Distributional Typology
3. Studying the typology of contact-induced change directly

Answering the 'what's where why?' question, an interest in explaining past and present linguistic diversity in its own right (BICKEL 2007, 2015).

Basically, contact is treated as a **predictor**, with areality as a proxy.

Three approaches to language contact in typology

1. Classical Greenbergian typology
2. Distributional Typology
3. Studying the typology of contact-induced change directly

3. A third family of approaches

Studying the typology of contact-induced change directly.

- Loanwords (Haspelmath & Tadmor 2009)
- 'Grammatical borrowing' (Matras & Sakel 2007)
- Affixes (Seifart 2015)

These studies all share the aim of directly targeting the typology of contact-induced change on an empirical cross-linguistic basis, in order to help explain cross-linguistic distributions and to make generalizations.