



**Angus McIntosh Centre**  
for Historical Linguistics

# Allophony: the phonological status of spelling in historical linguistics

## **Session 2: Non-native Spellers**

*Benjamin Molineaux*



THE UNIVERSITY  
of EDINBURGH

ETHL Summer School  
Friday 26 June 2026

# Wednesday and today

---

- **Session 1: Wednesday**

- Native spellers (i.e. native speakers spelling their own language)
- Spelling as phonological knowledge wrt non-standard spelling
- Graphemes: the *referential* and the *autonomistic* approaches
- Grapho-phonological profiles: prodigal, economical & sub-optimal
- Case study: Older Scots dental fricatives

- **Session 2: Today**

- **Non-Native Spellers: one script two phonologies**
- **'General' and 'Proper' Alphabets: phonological knowledge & typology**
- **Case studies: Mapudungun dentals; Yahgan and Aymara vowels**

# A 'proper' spelling for every language?

*rítr sínum stǫfum hver þjóð sína tungu*  
(Old Icelandic 'First Grammarian' c.1150AD)

'Each nation writes its own language with letters of its own'

*Chaque peuple a son alphabet qui n'est pas celui d'un autre*  
(Charles de Brosses 1765: 15)

'Each people has its own alphabet, one which is not the same as that of other peoples'

*eaux hayle nawl!*

- Or are we missing something?
  - What is meant by '**stǫfum**' and '**alphabet**'?

# The Proper Alphabet Principle

---

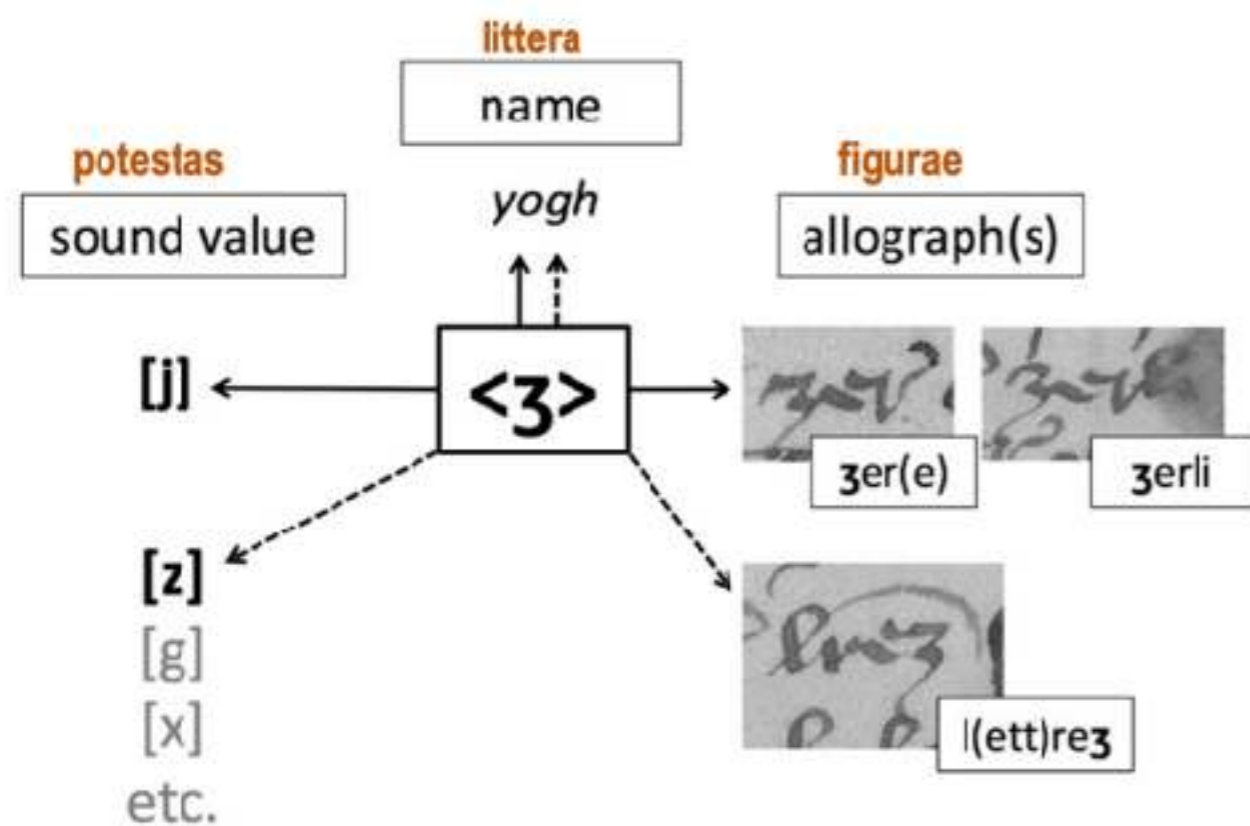
the term **stafróf** ('alphabet') implies that even if the Icelandic alphabet contains 'all those Latin letters that seemed to me to fit our language well', it is nevertheless a separate alphabet, ... even if both the shape (*figura*) and the name (*nomen*) of many letters remain the same as in the Latin alphabet, the fact that each letter – by the very fact of being used in a new language – necessarily has a value (*potestas*) of its own, proper to the new language, is sufficient to make it part of a separate alphabet. (Benediktsson 1972: 87)

There is no country which does not have a manner of speaking which is peculiar to itself, and which those from another country can never imitate successfully, or indeed cannot imitate at all. [...] We in Europe have **letters** which it is impossible for the Chinese to pronounce; and our best efforts are of no use to us when we attempt to copy the inflection of the Hottentots. (Charles de Brosses 1765: 15)

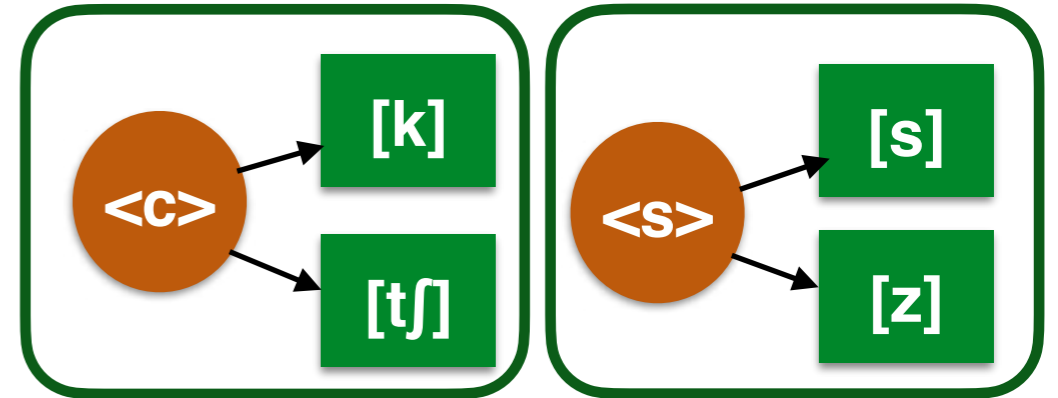
- 'Alphabet' here, he quite specifically means the set of abstract letters that each language has, not the set of graphs with which they are written.

# The doctrine of the *littera*

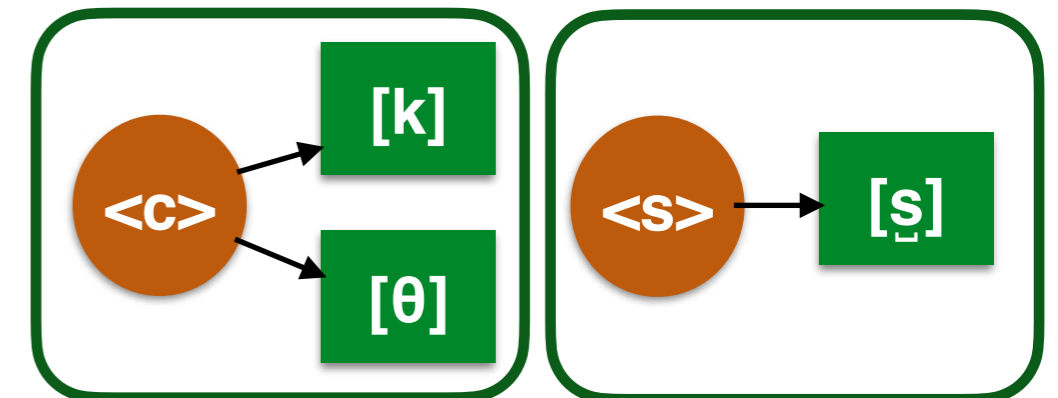
(one more time with feeling)



Italian



Castilian



- The alphabetic principle, until the advent of 'general' alphabets (the IPA and its predecessors) was language specific and not bi-unique
- Letters didn't have a generalised *potestas* but one that was language specific and hence *phonetically detailed*

# 'Proper' v. 'General' Alphabetic Principles

---

- **David Cram** (1999, 2018, 2024) identifies these two competing principles in the historiography of alphabetic writing systems
- ***The Proper Alphabet Principle***
  - *“Each language should have a writing system that matches its own alphabet (i.e.its set of abstract letters)”*
- ***The General Alphabet Principle***
  - *“For the purposes of linguistic analysis we need a system for transcribing the sounds of speech which is independent of any particular language and equally applicable to all languages”*

# Skepticism around a 'General' Alphabet

---

[D]ifferences in pronunciation occur in various languages which are not attributable so much to the individual letters, as to the whole style of speech of the community. For instance the English as it were push forward the whole of their pronunciation into the front of the mouth, speaking with a wide mouth cavity, so that their sounds are more distinct. The Germans, on the other hand, retract their pronunciation to the back of the mouth and the bottom of the throat, so that they have a stronger and more forceful pronunciation.

(**John Wallis** [1652] trans. Kemp, 1972: 209–11)



# Skepticism around a ‘General’ Alphabet

If one wanted to count how many letters there are that are not precisely similar to any other, one would have to count how many different sounds that voices can make, and how many different inflexions there are in the manner in which each people on the earth varies the movement of each of the vocal organs: which would produce an infinite number of letters.  
(Charles de Brosses 1765: 106)



Wenn man von der Artikulationsbasis einer Sprache oder eines Dialekts spricht, meint man eben den Durchschnitt der Mundlagen der einzelnen Glieder der Sprachgenossenschaft, ebenso wie man unter einem ‘englischen t’ oder ‘französischen u’ den Durchschnittswert versteht. (Jespersen ([1904] 1933): 179)

When we speak of the **articulation basis** of a language or dialect, we are referring to the **average positioning of the mouth and tongue** among the members of the linguistic community—just as we understand the term to denote the average value when referring to an “English *t*” or a “French *u*”

# So, how to go about non-native writing?

---

- *Scripting*: the development of a system of writing for a language, usually on the basis of of a ‘donor’ script (Bunčić et al 2016)
- The orthography of the donor language is typically implemented as closely as possible for the adopting language *potestates*
- ‘Straightforward’ when the two languages share ~~phonemes~~.
- Economical because scripters do not have to master two completely different orthographies; “transfer reading knowledge from one well-learned language to the new system” (Hinton 2014: 145).
- Problems arise where the phonologies of the adopting and donor languages do not match. The scripters must decide, among other things, what to do with ‘superfluous’ graphemes (Stenroos & Smith 2020)

170. And precisely why it is helpful to include phonetic detail in the table is completely clear: the **Graphemicisation** hides linguistically-relevant information from us. (Hale, Tuesday)

# Case Study: Aymara vowels



- Competing spelling conventions for the vowels in the language (<"> marks length)
- Consonants:

		<b>Bilabial</b>	<b>Dental/ Alveolar</b>	<b>Palatal/ Postalveolar</b>	<b>Velar</b>	<b>Uvular</b>	<b>Glottal</b>
<b>Nasal</b>		m	n	ñ	nh		
<b>Plosive</b>	<b>voiceles</b>	p	t	ch	k	q	
	<b>aspirate</b>	ph	th	chh	kh	qh	
	<b>ejective</b>	p'	t'	ch'	k'	q'	
<b>Fricative</b>			s			ih	i
<b>Approximant</b>	<b>central</b>			y	w		
	<b>lateral</b>		l	ll			

- Assuming the two systems represent the same spoken form, is one better suited to the language than the other?

## AKAPACH JAQ<sup>E</sup> WALINKAÑAPATAKI IN<sup>O</sup>QAT ARU

### Qalltañataki

UNANCHASA, aka pachana jaq<sup>e</sup>jh munañanisina, ch<sup>e</sup>qpacha amuyasisa, ch<sup>e</sup>qa thakir sarjhatasa, jilan sullkanjama arnaqasa, jan nuwasisa utjañaru wayt'asiñ yati;

UNANCHASA, jaq<sup>e</sup>n walinkañapataki wakisir arunaka armasna, yaq<sup>h</sup>e<sup>p</sup> jaq<sup>e</sup>jh masinakaparu khuyañ autuyi, mututanakasti, uñatatasisa, wajcha kankañapa atipasna, jan khitirus ajhsarasa, ch<sup>e</sup>q-ch<sup>e</sup>qa qhana arumpi arustatasjhapjhe;

UNANCHASA, jaq<sup>e</sup>n walinkañapataki in<sup>o</sup>qat arujha arjhatatäñapaw jan jaq<sup>e</sup> ñanqhachirinakaru ch'ajhwa qalítasna waninchapjhañapataki;

UNANCHASA, taqpacha markanakajh kunti munapj<sup>e</sup> ukanaka jikjhatañatak<sup>e</sup>jh jupanakpacha sumañat arusthapisipjhañapaw;

UNANCHASA, Mathapit Mayachat Markanakan Arupa taqjhatasna akham markanakajh amuyasi: Taq<sup>e</sup> mayniwa, chachas warmis, jaq<sup>e</sup> kankañanísinjha, ch'amachasisna má ch'ulla chachjama jaq<sup>e</sup>n walinkañapa jikjhatañaru wayt'asipj<sup>e</sup>;

UNANCHASA, taq<sup>e</sup> mrakanakajh Matahapit Markanakamp chika akham amtapj<sup>e</sup>: sapaqat markanjha aniw jaq<sup>e</sup>jh inaki arknaqatäkaspati; jakañapans yänakapans arjhatatäñapaw;

UNANCHASA, Mathapit Markanakan amuyasitanakapa yatisina khuskat taqpachanipaw wayt'asipjhañapa ph<sup>o</sup>qasiñapataki.

## MAYACHAT MARKANAKAJH JACH'A MATHAPIWINAKAPAN JAQ<sup>E</sup>N WALINKAÑAPATAKI IN<sup>O</sup>QAT ARU YATIYI

Taqpacha jaq<sup>e</sup>naka, ayllunaka, markanaka, aka arunakaru iyau sasna ch'amachasipjhañapaw, isk'alalanakas, wayna tawaq<sup>o</sup>nakas, chuymani jaq<sup>e</sup>nakas, markapan q<sup>o</sup>rpap manqhana yatichatäp<sup>j</sup>hañapaw. Ukhamarus q<sup>o</sup>rpamarkanakampi jaya markanakampi arusthapisisa taqpacha markanakajh khuskat ch'amachasipjhatapampi alajh toq<sup>e</sup>ru wayt asitapjaniw.

## AKAPACH JAQ<sup>I</sup> WALINKAÑAPATAKI IN<sup>U</sup>QAT ARU

### Qalltañataki

UNANCHASA, aka pachana jaq<sup>i</sup>jh munañanisina, ch<sup>i</sup>qpacha amuyasisa, ch<sup>i</sup>qa thakir sarjhatasa, jilan sullkanjama arnaqasa, jan nuwasisa utjañaru wayt'asiñ yati;

UNANCHASA, jaq<sup>i</sup>n walinkañapataki wakisir arunaka armasna, yaq<sup>h</sup>i<sup>p</sup> jaq<sup>i</sup>jh masinakaparu khuyañ autuyi, mututanakasti, uñatatasisa, wajcha kankañapa atipasna, jan khitirus ajhsarasa, ch<sup>i</sup>q-ch<sup>i</sup>qa qhana arumpi arustatasjhapjhi;

UNANCHASA, jaq<sup>i</sup>n walinkañapataki in<sup>u</sup>qat arujha arjhatatäñapaw jan jaq<sup>i</sup> ñanqhachirinakaru ch'ajhwa qalítasna waninchapjhañapataki;

UNANCHASA, taqpacha markanakajh kunti munapj<sup>i</sup> ukanaka jikjhatañatak<sup>i</sup>jh jupanakpacha sumañat arusthapisipjhañapaw;

UNANCHASA, Mathapit Mayachat Markanakan Arupa taqjhatasna akham markanakajh amuyasi: Taq<sup>i</sup> mayniwa, chachas warmis, jaq<sup>i</sup> kankañanísinjha, ch'amachasisna má ch'ulla chachjama jaq<sup>i</sup>n walinkañapa jikjhatañaru wayt'asipj<sup>i</sup>;

UNANCHASA, taq<sup>i</sup> mrakanakajh Matahapit Markanakamp chika akham amtapj<sup>i</sup>: sapaqat markanjha aniw jaq<sup>i</sup>jh inaki arknaqatäkaspati; jakañapans yänakapans arjhatatäñapaw;

UNANCHASA, Mathapit Markanakan amuyasitanakapa yatisina khuskat taqpachanipaw wayt'asipjhañapa ph<sup>u</sup>qasiñapataki.

## MAYACHAT MARKANAKAJH JACH'A MATHAPIWINAKAPAN JAQ<sup>I</sup>N WALINKAÑAPATAKI IN<sup>U</sup>QAT ARU YATIYI

Taqpacha jaq<sup>i</sup>naka, ayllunaka, markanaka, aka arunakaru iyau sasna ch'amachasipjhañapaw, isk'alalanakas, wayna tawaq<sup>u</sup>nakas, chuymani jaq<sup>i</sup>nakas, markapan q<sup>u</sup>rpap manqhana yatichatäp<sup>j</sup>hañapaw. Ukhamarus q<sup>u</sup>rpamarkanakampi jaya markanakampi arusthapisisa taqpacha markanakajh khuskat ch'amachasipjhatapampi alajh tuq<sup>i</sup>ru wayt asitapjaniw.

# Case Study: Aymara vowels

- Five-vowel system v. three-vowel system

- <e> ~ <i>

- <o> ~ <u>

Allophonography, yay!

- Is the variation free? *eaux hayle nawl!*

- <i> → <e> before or after <q>, <q'>, <qh> or <jh> (uvular consonants)

- <u> → <o> in the same context

/i/  $\begin{cases} [i] \\ [e] \end{cases}$

/u/  $\begin{cases} [u] \\ [o] \end{cases}$

- This appears to be allophony:

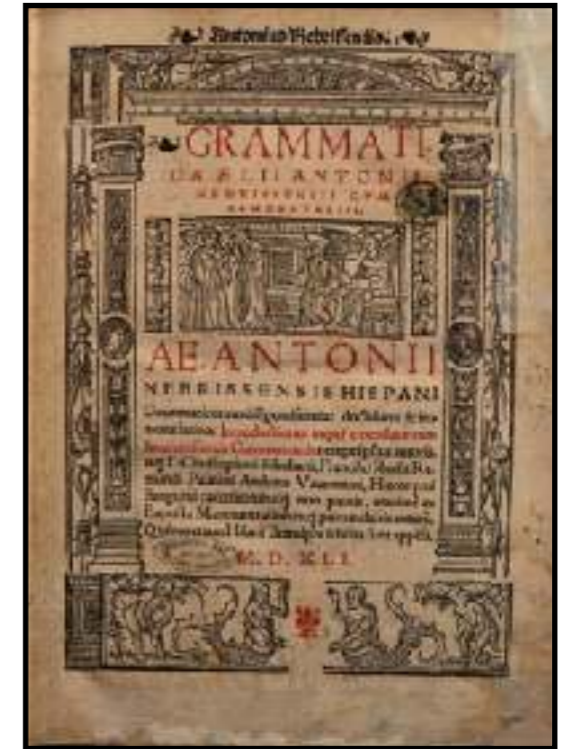
- Why encode allophones in an alphabetic system?

- Who is doing the writing and for whom?

- Most writers are fluent in Spanish, where mid vowels are phonemic

# Excursus: Early linguistic description in the Americas

- Pre-conquest linguistic description is unknown
- In 1492 few grammars of European languages existed
- Descriptions of American languages emerge in tandem with ‘grammatization’ of languages in Europe (Oesterreicher 2019)
- In both, Latin grammars persisted as the model
- A descriptive tradition emerged in Catholic empires



We, who, though unworthy, exercise on earth the power of our Lord and seek with all our might to bring those sheep of His flock who are outside into the fold committed to our charge, consider, however, that the Indians are truly men and that they are not only capable of understanding the Catholic Faith but, according to our information, they desire exceedingly to receive it. ... the said Indians and all other people who may later be discovered by Christians, are by no means to be deprived of their liberty or the possession of their property, even though they be outside the faith of Jesus Christ...

*Sublimis Deus*, 1537 – Pope Paul III

# Excursus: Early linguistic description in the Americas

- A first grammar (Tarascan/Purépecha) was written in 1558 by Maturino Gylberti
- In South America, Domingo St. Tomás published a grammar of Quechua in 1560
- These were thought of as pedagogical tools
- They were often followed by a series of Christian doctrinal texts to be used in evangelisation
- A key objective of the Church was to ‘reduce’ these languages and their speakers (Hanks 2010)

*Arte de la Lengua Mexicana (Nahuatl),*  
Antonio de Rincón, 1595

diferētes preceptos. De manera q̄ é aq̄  
llo q̄ me è podido aprouechar d̄ la gra  
matica latina siēpre meyre arrimado a

ella pero é las úmas cosas, é q̄ esta légua  
se diferēcia d̄ la latina por ser ellas nue  
uas a sido forçoso reducirlas anuevas  
reglas, cō el nuevo estilo q̄ se require.



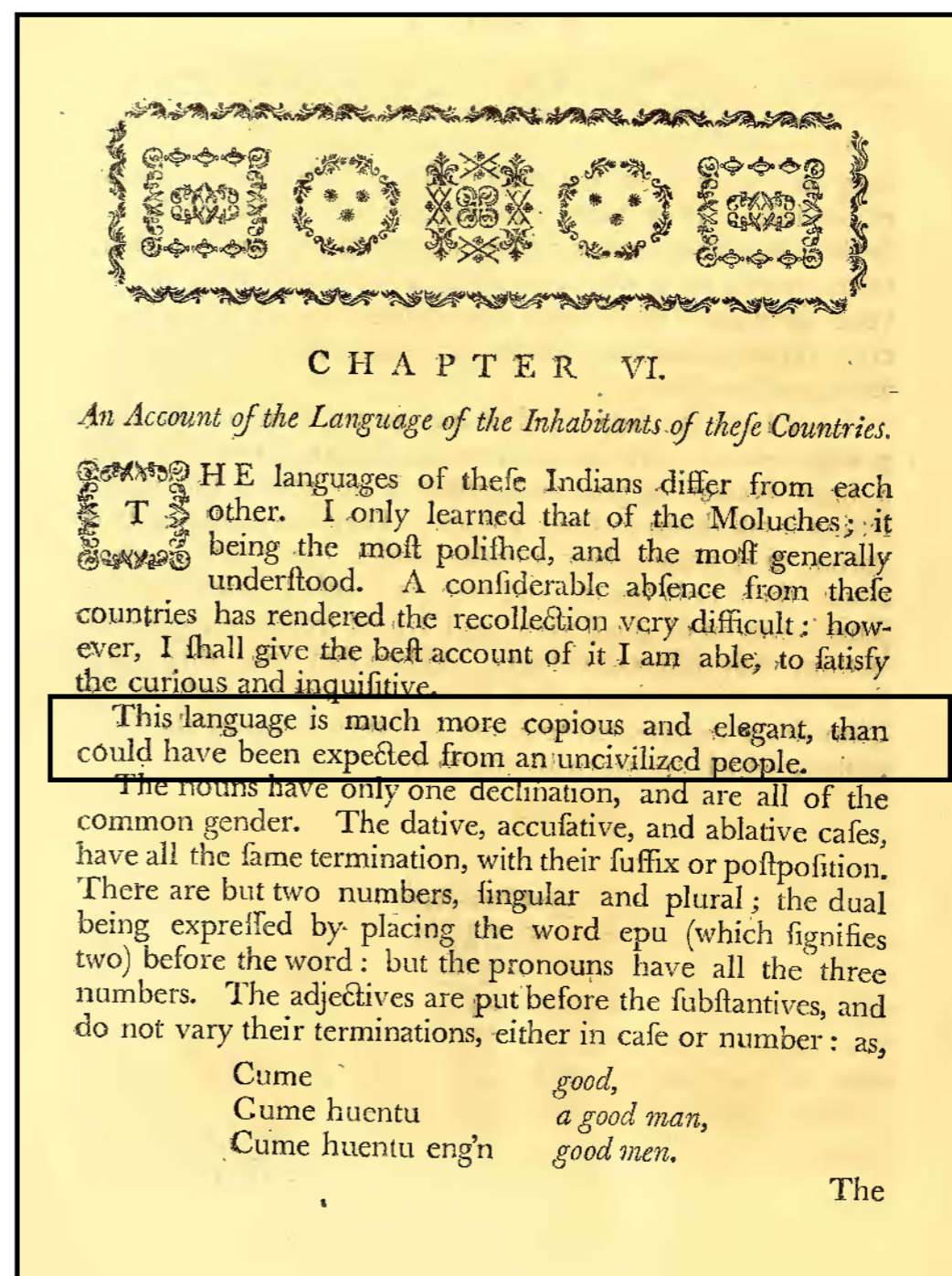
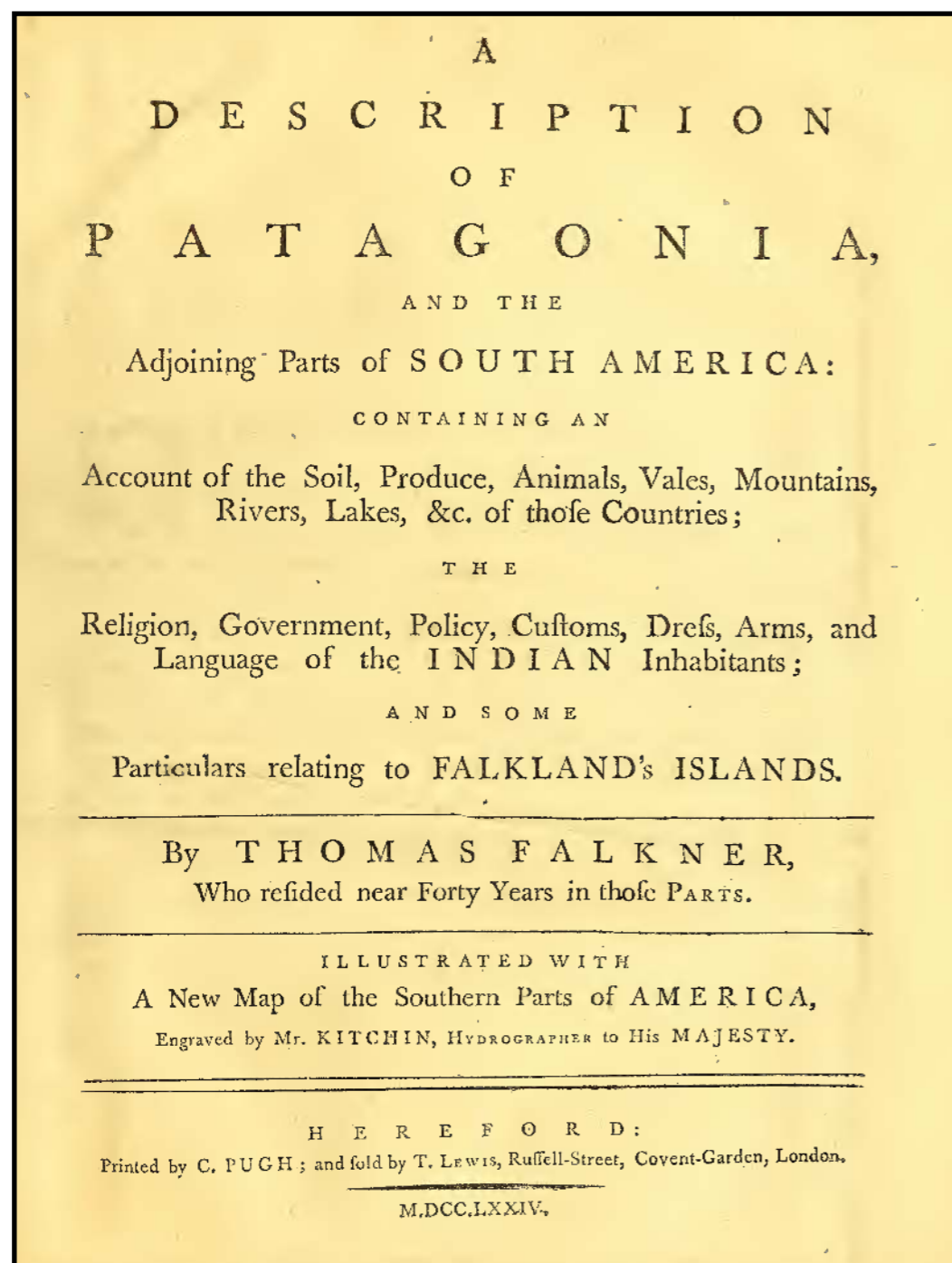
So it is that, where I could take advantage of the Latin grammar I did approximate it, but in other things, in which this tongue is different from the Latin, for they are new, I have by force had to ‘reduce’ these to new rules, with the new style which they require.

# Mapudungun

- A.K.A. *Araucanian*, *Mapuche* is the ancestral tongue of the Mapuche people
  - ±200,000 speakers (Chile/Argentina)
  - Considered “definitely endangered” (UNESCO, 2010), mostly due to poor transmission
- Presumed to be a language isolate
- Documented from the early 17c onwards
  - writing by Jesuits
  - then Franciscans
  - linguists/anthropologists
  - natives speakers



# Admiration and othering



# Admiration and othering



**H**abet autem Lingua Chilensis quam plurima omnino præclara, rara & inaudita: unicam nempe Declinationem, Conjugationem unicam; & utramque perfacilem & obviam; unicum genus, nullum Nomen aut Verbum anomalum; Regulas non nisi universales ac sine exceptione, atque semper eadem vox paucis literis mutatis, demptis aut adjectis, agit vices & substantivi, & verbi, & adjectivi, & adverbii; adjectisque vel insertis variis & sæpe permultis Particulis format omnia & quæcunque sua composita, nec raro sensum integrum & bene longum:

Unde fit, ut Lingua Chilensis sit Lingua multo facillima; & tamen Barbarorum, non solum non barbara; sed aliis Linguis tanto melior; ut sicuti montes Andes alios montes; ita hæc alias Linguas usque eo supereminet; ut qui callet Chilense Idioma, alias Linguas velut e specula longe infra se despiciat, clare patenterque agnoscens quantum sit in illis superflui, quam multa quoque desint, &c. &c. & quam jure merito unicuique non Chileno exprobrari possit:

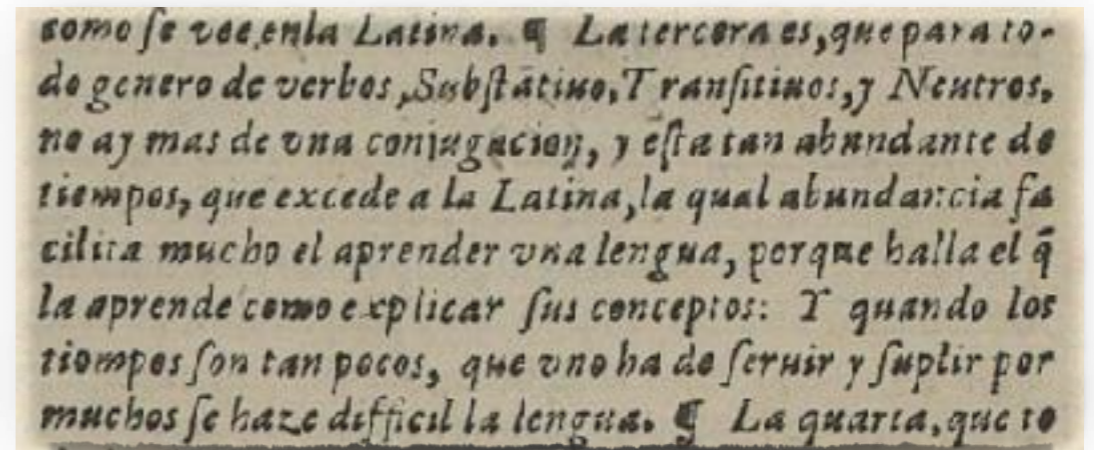
Si tua Lingua bona est, tamen est præstantior illâ Lingua Chilena, brevis quam docet iste liber. vide. P. 5. sub finem.

The language of Chile is a very easy tongue; although it is a language of barbarians, not only is it not a barbarian tongue, but it is so much better than other tongues, as the Andes are greater than other mountains, so it towers over other languages. They who know the Chilean tongue will look down upon other tongues as far below themselves, as if from a watchtower, knowing clearly and patently how much is superfluous in them, how much is lacking, etc. and to any non-Chilean they can boast: if your tongue is good, the Chilean tongue is superior.

# Structural descriptions

- Luys de Valdivia (1606)

*The third [advantage of this tongue] is, that among all its verb types, substantive, transitive, and neutral, there is only one conjugation, and this is so abundant in tenses [‘tiempos’], that it exceeds the Latin, this said abundance facilitates the learning of a language, because he who learns it finds a way to explain his concepts, and when the tenses are so few that one has to serve and be used in place of many, the language becomes difficult.*



como se ve en la Latina. ¶ La tercera es, que para todo genero de verbos, Substantivo, Transitivos, y Neutros, no ay mas de una conjugacion, y esta tan abundante de tiempos, que excede a la Latina, la qual abundancia facilita mucho el aprender una lengua, porque halla el q̄ la aprende como explicar sus conceptos: Y quando los tiempos son tan pocos, que uno ha de servir y suplir por muchos se haze difficil la lengua. ¶ La quarta, que to

The verbs have three numbers, singular, dual, and plural; and as many tenses as in the Greek tongue; all of which they form by interposing certain particles before the last letter of the indicative, and before the last syllable of the subjunctive: as,

Present tense,	Elun,	<i>to give.</i>
Imperfect,	Elubun,	
Perfect,	Eluyeen,	
Preterperfect,	Eluyeebun,	
First Aorist,	Eluabun,	
Second Aorist,	Eluyeeabun,	
First Future,	Eluan,	
Second Future,	Eluyean.	

- Thomas Falkner (1774)

- Contemporary descriptions claim that the language has a very minimal tense-marking system – it only has explicit future marking
- However, the language has abundant marking for modality, aspect, evidentiality and mood marking
- The admiration here is not unwarranted, but it is misplaced: value is only assigned in relation to pre-existing parameters and usefulness

# Mapudungun consonants

	Labial	Dental	Alveolar	Retroflex	Postalveolar/ Palatal	Velar
Stop/affricate	p	t̪	t	ʈʂ	tʃ	k
Fricative	f	θ	s	ʂ	ʃ	
Nasal	m	n̪	n		ɲ	ŋ
Lateral		l̪	l		ʎ	
Approximant	w				j	ɥ

**Present-day Central Mapudungun Consonants**  
(based on Sadowsky et al. 2013)

- Key distinction between *peripheral* and *coronal*
- Rich coronal inventory, including a robust dental-alveolar contrast

# Dental-alveolar contrast in Mapudungun

- This is backed up by a number of more recent phonetic studies:
  - Sadowsky *et al.* 2013, Fasola *et al.* 2015, Painequeo *et al.* 2018, Mena and Salamanca 2018, Figueroa *et al.* 2019,

**/mi<sub>ɲ</sub>na/**  
‘cross-cousin’



**/mi<sub>n</sub>na/**  
‘enough’

Stops	Nasals	Laterals	Fricatives
[t̪ən] ‘head louse’	[mɐ.ɲa] ‘cousin’	[ki.ɭa] ‘bamboo’	[θa.kel] ‘pact/agreement’
[tən] ‘high sound’	[mɐ.na] ‘much’	[ki.la] ‘three’	[sa.ku] ‘sack’(<Spa.‘saco’)

# Historical evidence for Mapudungun dentals

- Dental/alveolar contrasts in the earliest grammar:

Luys de Valdivia: Arte y Gramática (1606)

se ha de notar que estas tres letras siguientes l. n. t. de mas de la pronunciacion que tienen con nuestra Española en estas syllabas (la, na, ta) que también se usa en esta lengua; hazen otra pronunciacion con ellas, que es necesario saberla para evitar la equiuocacion q̄ ay en muchos vocablos, cuya significaciō depende de esta pronunciaciō, la qual es desta manera, que **arrimando la punta de la lengua a los dientes pronuncia (la, le, li, &.) y (na, ne, ni, &.)** lo qual no tiene la primera pronunciacion nuestra

[l̪] <l'an> 'die'

[l̪] <lanta> 'testicle'

[n̪] <n̪oyn> 'eat too much'

[n̪] <n̪on> 'win'

- But no comments on [t̪] and [θ]...



“... it should be noted that these three letters <l, n, t>, aside from the pronunciation they have in our Spanish syllables ‘la, na, ta’, have in this tongue another pronunciation which should be known in order to avoid mistakes, since the meaning of many words depends on this pronunciation, which is thus: that **nearing the tip of the tongue to the teeth**, they pronounce *la, le, li* etc. and *na, ne, ni* etc., which is different from our first pronunciation” (p.9.r)



# Historical evidence for Mapudungun dentals

- Coronal stops in Valdivia's *Arte y Gramática* (1606):
  - When spelled as <̄t> 'they move the tip of the tongue towards the high palate' (9r)
  - Stops spelled as <t> are pronounced 'like we pronounce them'

[t] <̄tecan> 'wander' > [tʂ]ekan  
 [t] <tica> 'adobe' > [t]ika  
 [ʔ] <tue> 'earth' > [t̪]ue

Present day Central Mapudungun

	Labial	Dental	Alveolar	(Alveolo-) Palatal	Retroflex	Velar
Stop/Affricate	p	t̪	t	tʃ	tʂ	k
Fricative	f	θ	s	(ʃ)	ʂ	
Nasal	m	n̪	n	ɲ		ŋ
Liquid		l̪	l	ʎ		
Approximant	w			j		ɥ

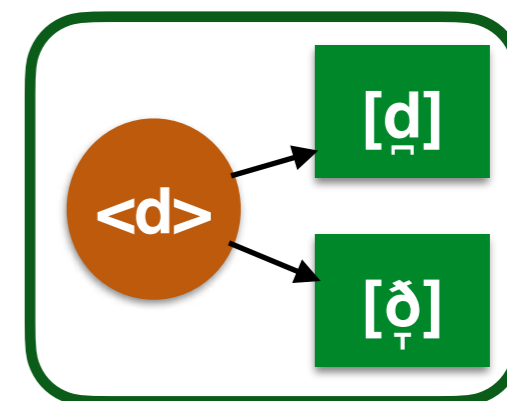
- It is unclear whether there is a dental-alveolar contrast here
  - Spanish stops were already denti-alveolar (Penny 2002), which probably implies they were laminal and post-dental
  - Mapudungun has interdentals which are laminal and alveolars which are apical

# Historical evidence for Mapudungun dentals

- What about the fricatives?
- In Valdivia's grammar, PDCM words with /θ/ are spelled with <d>

cuđañ,  
cude,

'testicle'  
'old woman'



[d̄]ios  
'god'

ca[ð̄]a  
'each'

- 16c Spanish <d> had two principal *potestates* (Penny 2002)
  - The implication would be that the dental fricative was voiced
  - Indeed, Valdivia (1606) represents all fricatives with **voiced** spellings

<v/b>=[β/v]      Volil, 'root'  
labquen 'sea'  
<r>=[ʒ]      Ruca 'house'  
Cura 'stone'  
<z>=[z]      Mizqui 'honey'  
Pozco 'yeast'

	Labial	Dental	Alveolar	Retroflex
	p	t̄	t	t̄
	β/v	ð̄	(z)	ʒ
	m	n̄	n	
		l̄	l	

The few instances of words containing <z> are either Quechua or Spanish borrowings containing [s] in the donor language (7 entries in the 3,100-entry Vocabulary)

# Dental fricatives in historical Mapudungun

- Valdivia (1606), Havestadt (1777), Febrés (1764) provide similar descriptions for northern dialects
- Independently, Thomas Falkner's 'Description of Patagonia' (1774) covers 18c central varieties further south
- PDCM /θ/ =<z>; /f/ = <f>



Az,

a[θ]

PDCM a[θ] 'face'

Cofque,

ko[f]ke

PDCM ko[f]ke 'bread'

Dialectology, yay!

- Use of <hu> for /w/ (cf. <huenuy> [wenuj] 'friend') suggests Spanish spelling practices. Hence <z> is not [z] but voiceless [θ] for the transcriber
- The key point is that fricatives appear to be voiceless in this area. This even applies to the retroflex (elsewhere <r>)

Labial	Dental	Alveolar	Retroflex
p	t̪	t	ʈ
f/ɸ	θ	(s)	ʂ
m	n̪	n	
	ɳ	ɳ	

Cufa

ku[ʂ]a

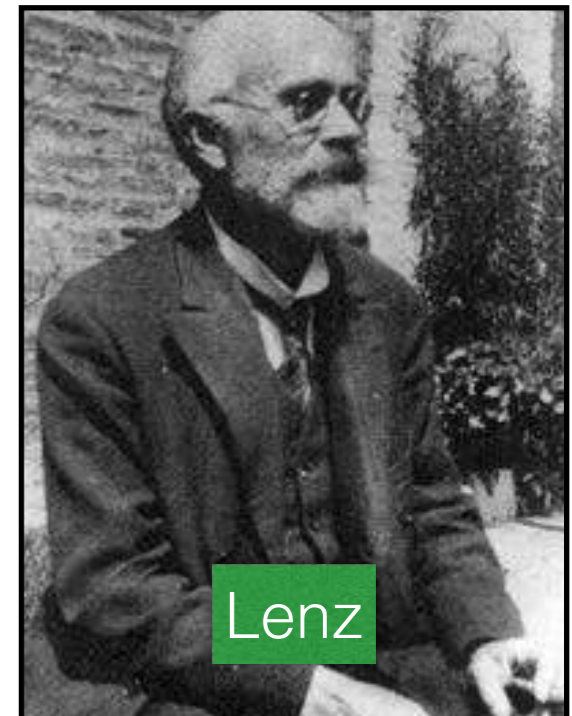
PDCM ku[z]a 'stone'

# Rudolf Lenz and Kalfün

- German-born linguist Rudolf Lenz (1863–1938) describe the dental-alveolar contrast in Mapudungun only in 1897
- He recalls how Kalfün (Segundo Jara), his main Mapuche collaborator would patiently turn to him when pronouncing the key sounds:

“Calvún, docile as ever, finally lifted his head at each *fent·e* [fɛnt̪ɛ], *mət·e* [mɛt̪ɛ], etc. in order to show me the tip of the tongue peaking between his teeth” (1897: 130)

- Only then did he eventually understood to be systematic.



Chumngechi wirintükungekey Mapuzungun mew  
Escrituras del Mapudungun

Dd

Variante: Zz

Kimeltuwe



Unificado · Raguileo · Azümchefe  
dewü · zewv · zewü

Chumngechi wirintükungekey Mapuzungun mew  
Escrituras del Mapudungun

Tt

Kimeltuwe



Unificado · Raguileo · Azümchefe  
tapül · tapvl · thapül

Chumngechi wirintükungekey Mapuzungun mew  
Escrituras del Mapudungun

Ll

Kimeltuwe



Unificado · Raguileo · Azümchefe  
loyka · boyka · lhoyka

Chumngechi wirintükungekey Mapuzungun mew  
Escrituras del Mapudungun

Nn

Kimeltuwe



Unificado · Raguileo · Azümchefe  
namün · hamvh · nhamünh

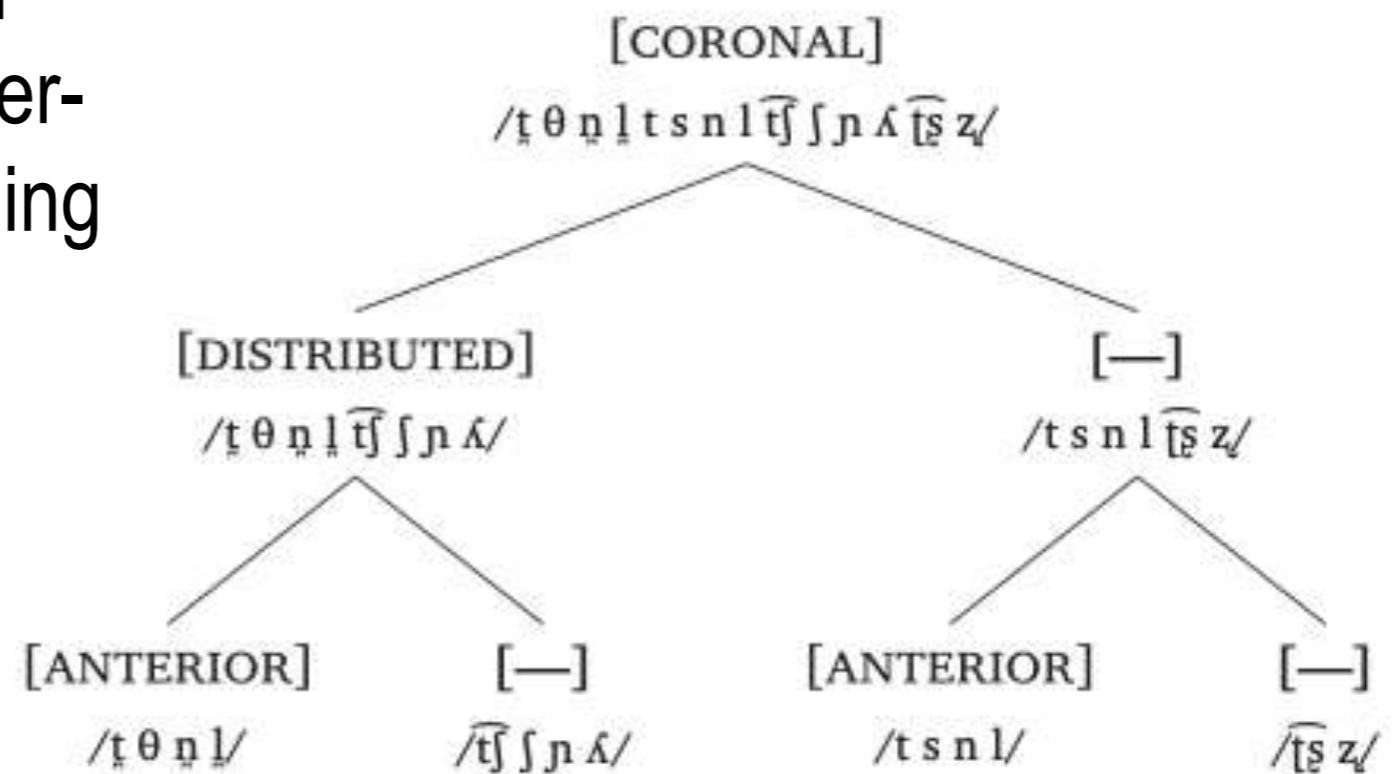
# So, what have we learned?

---

- The scripting of Colonial Mapudungun dentals is predicted to be problematic:
  - There is a phonemic contrast in the target language (Mapudungun) which is not present in the donor language (Spanish)
  - Where there is no major phonetic overlap between the sounds of the two languages, a new spelling is devised:
    - <ñ> for [ɲ], where Spanish has only <n> for [n]
    - <l'> for [ʎ], where Spanish has only <l> for [l]
  - Use of <d> for [ð] is evidence for *littera* not being graphemic, but rather representing a constellation of *potestates*
    - <d> is used without comment, meaning it is an unremarkable mapping
    - Graphemic assumptions would have <d> as /d/

# So, what have we learned?

- The lack of a contrast in the scripting of the dental stop seems to imply that:
  - For Valdivia, Spanish <t, n, l> are not just /t, n, l/, but more concrete potestates [t̪], [n], [l], where the stops are not specified in the same way as the nasal and lateral
  - Nevertheless, <t> remains a bad match for the Mapudungun interdental [t̪] and the apical [t̪], failing to reflect a more systematic contrast in the phonology of Mapudungun



# So, what have we learned?

---

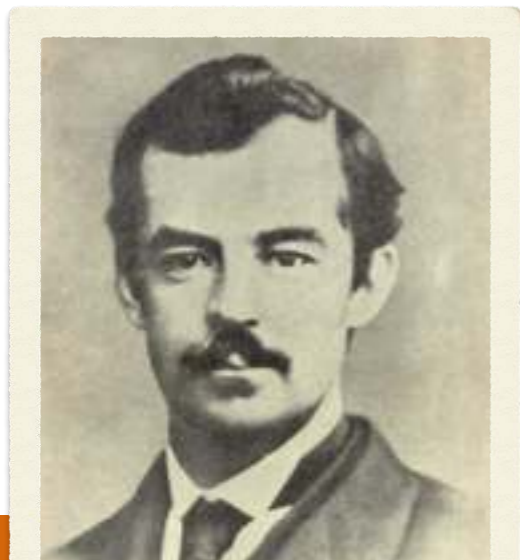
- Early attempts to script a language largely rely on expanding a ‘proper alphabet’ to represent a new language
- This means that acceptable variation is given by the potestatic substitution sets of the donor language, not by a phonemic analysis of the donor language and later application of a biuniqueness principle to the target

# A 'General Alphabet', at last? – Yahgan Vowels

- Yahgan is the southernmost native language in the world
- It is recently dormant, the following the death of Cristina Calderón in 2022
- The language was mostly documented by Thomas Bridges, of the South American Missionary Society, an Anglican group, in the late 1800s
- He published translations from the New Testament, wrote a grammar and a dictionary all using A. J. Ellis' 'Phonotypic Alphabet'

GOSPL L4C ECAMANĀCI.

THE GOSPEL OF S. LUKE  
TRANSLATED INTO THE  
YAHGAN LANGUAGE.



Rev. Thomas Bridges  
(~1842-1898)



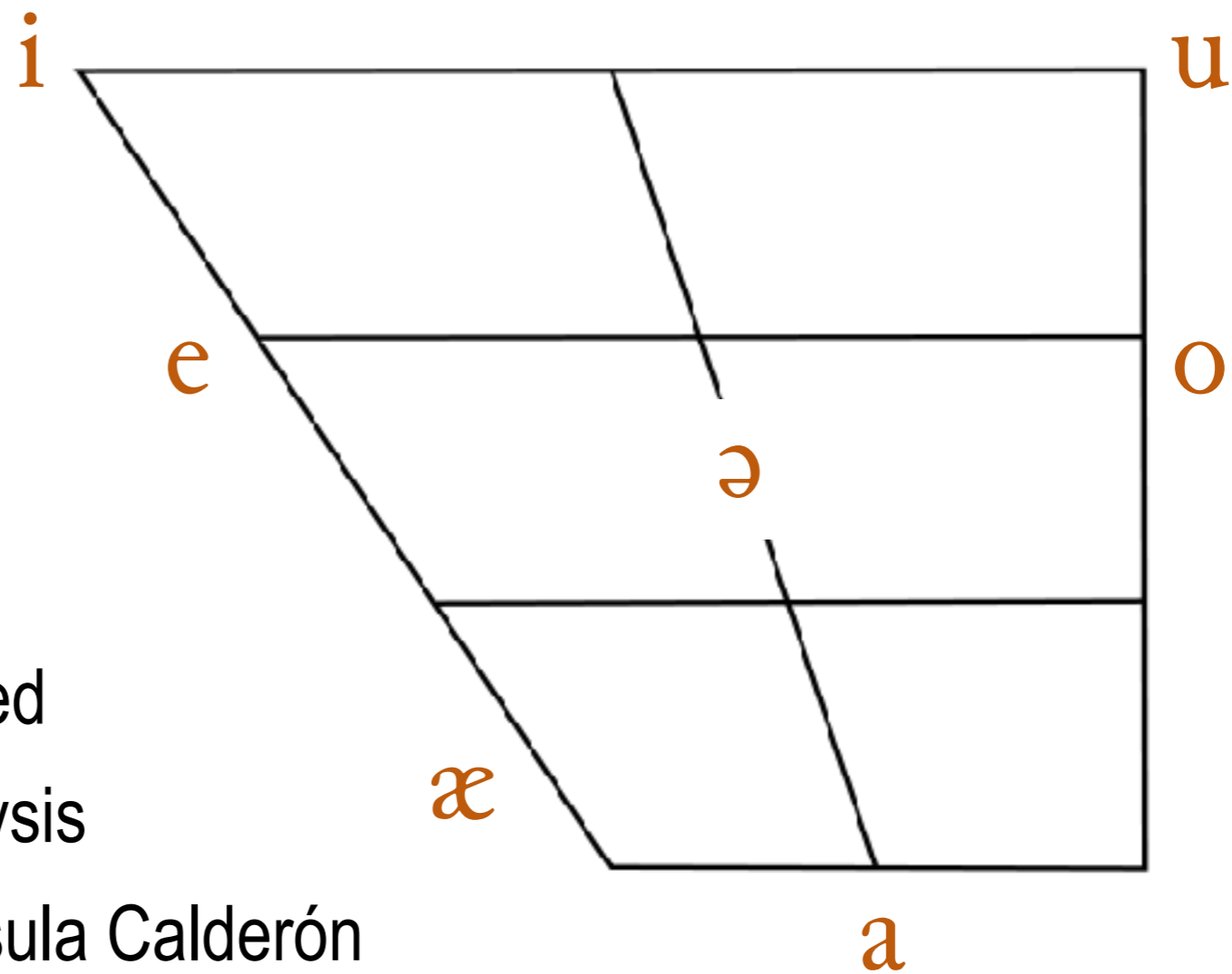
Cristina

Úrsula



# 20th-century Yahgan documentation

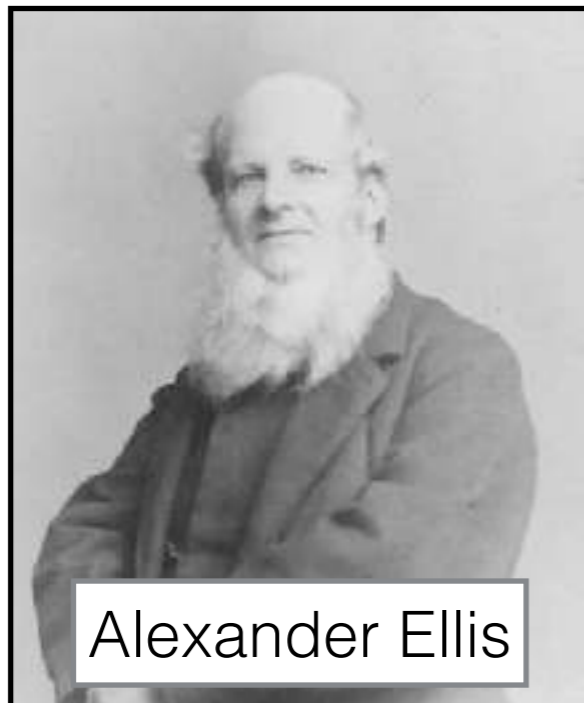
- There are three main studies done in the late 20th/ early 21st century
  - Salas & Valencia 1990; Poblete & Salas 1999; Aguilera 2000



- Seven vowels are described
- Traditional phonemic analysis
- Data from Cristina and Úrsula Calderón

# Ellis' *Phenotypic Alphabet*

- Developed by Alexander Ellis and Isaac Pitman
- It's main object was the teaching of English
- A precursor to the IPA
- Predates phonemic theory
- Was not a **General Alphabet**
- It strived for **bi-uniqueness for English**



Alexander Ellis

## THE ENGLISH PHONOTYPIC ALPHABET.

(1846)

VOWELS.				CONSONANTS.					
No.	Type.	Example of its sound.		No.	Type.	Example of its sound.			
LONG.									
1	Ɛ	e	eel	al	20	P	p	rope	rop
2	ŋ	a	ale	al	21	B	b	robe	rob
3	Å	u	alms	qmz	22	T	t	fate	fat
4	Θ	e	all	el	23	D	d	fade	fad
5	Ω	o	ope	op	24	Ɔ	g	etch	eg
6	U	u	food	fund	25	J	j	edge	ej
SHORT.									
7	I	i	ill	il	26	C	c	leek	lec
8	E	e	ell	el	27	G	g	league	leg
9	Λ	a	am	am	28	F	f	safe	saf
10	O	o	olive	oliv	29	V	v	sare	sav
11	U	u	up	up	30	T	t	wreath	ret
12	U	u	foot	fat	31	Ǝ	d	wreathc	ret
DIPHTHONGS.									
13	Ɔ	j	isle	il	32	S	s	hiss	his
14	Ɔ	e	oil	el	33	Z	z	his	hiz
15	Ɔ	ɔ	owl	xl	34	Σ	f	vicious	vifus
16	U	u	mule	mūl	35	Ɔ	g	vision	vigun
COALESCENTS.									
17	Y	y	yea	ya	36	R	r	for	fer
18	W	w	way	wa	37	L	l	fall	fel
ASPIRATE.									
19	H	h	hay	ha	38	M	m	seen	sem
					39	N	n	seen	sen
					40	Ɔ	g	sing	siŋ

# Bridge's system

- The *Yahgan Phonetic Alphabet* is a slight adaptation of Ellis' 1846 alphabet
- Treats the *Phonotypic Alphabet* as a **General Alphabet**
- The system is conceived as a series of bi-unique relationships
- His son, Lucas Bridges also adopts the system to work with Selk'nam, another Fuegian language (but that's another story)

## The Yahgan Phonetic Alphabet.

The letter is always			The letter is always			The letter is always		
written	printed	sounded as	written	printed	sounded as	written	printed	sounded as
A a	A a	a in at	O o	O o	oa in oat	Ō Ō	Ō Ō	ō in thō
Ā ā	Ā ā	ā - alms.	Ō Ō	Ō Ō	ō - far.	Ō Ō	Ō Ō	ō in so
Ō Ō	Ō Ō	ō - all.	<b>Consonants.</b>			Ō Ō	Ō Ō	ō - is
O o	O o	o - olive	C c	C c	c in cot	Ō Ō	Ō Ō	ō - show
U u	U u	oo - foot	G g	G g	g - go	Ō Ō	Ō Ō	ō - pleasure
U u	U u	oo - food	T t	T t	t - to	Ō Ō	Ō Ō	ō - low
I i	I i	i - ill	D d	D d	d - do	Ō Ō	Ō Ō	ō - ulsh
E e	E e	ee - eel	P p	P p	p - put	M m	M m	M - me
E e	E e	e - ell.	B b	B b	b - but	N n	N n	n - no
A a	A a	ai - aim.	F f	F f	f - foe	N n	N n	n - alman.
I i	I i	i - isle	V v	V v	v - vase	Ō Ō	Ō Ō	ō - wing
O o	O o	oi - oil	Ĉ ĉ	Ĉ ĉ	ch - chin	R r	R r	r - row
U u	U u	u - us.	J j	J j	j - jade	R r	R r	r - welsk
Ō Ō	Ō Ō	ow - owl.	Ĥ ĥ	Ĥ ĥ	th - thin	K k	K k	ch - german.



Beside the above letters there are these three vowel marks, ' = h, ` = y, ˘ = w only found over vowels, hence their name. These marks are often found united over vowels, thus ūā = uhuwa sij. ashes; ūā = wuhya sij. a door; ˘ ūā = wuhwa sij. tinder of birds' down. The effect of these marks always precedes the vowels over which they are found.

# The 19c data is hard to get at...

*Éléna. i. a To build, to make a wigwam, or house.*

29 scja acimus çesa satumuaçua ácupj ácwán; sina  
eloc s̄inatqmus cutupj sa bqv meàmèa sina cot  
ácwán.

- There is no Unicode for Ellis' Phonotypic Alphabet
- Traditional OCR fails
- Trained a model using 'dingbats' in Transkribus
- Works for print text, but not very well for handwritten text (yet!)



# The 19c data is hard to get at...

Yahgan / Bridges 1881 Ya... / p016.jpg Ground Truth

Yahgan / Yahgan Bridges ... / 0010.JPG In Progress

1 To look about for search to look.

2 Alagunsna.

3 To look at for ourself.

4 Alagunggin.

5 To look about, To chose with the eyes

6 Alaguetulli

7 St laqueméini. Alagenmeinten 8.6 To look at standing, sitting.

8 Hooking at.

9 commence or complete looking at take and look at

10 Alaguinata To us to get a sight of to sight.

11 An edible Ehellish

12 Mamuluk-Cuginsin

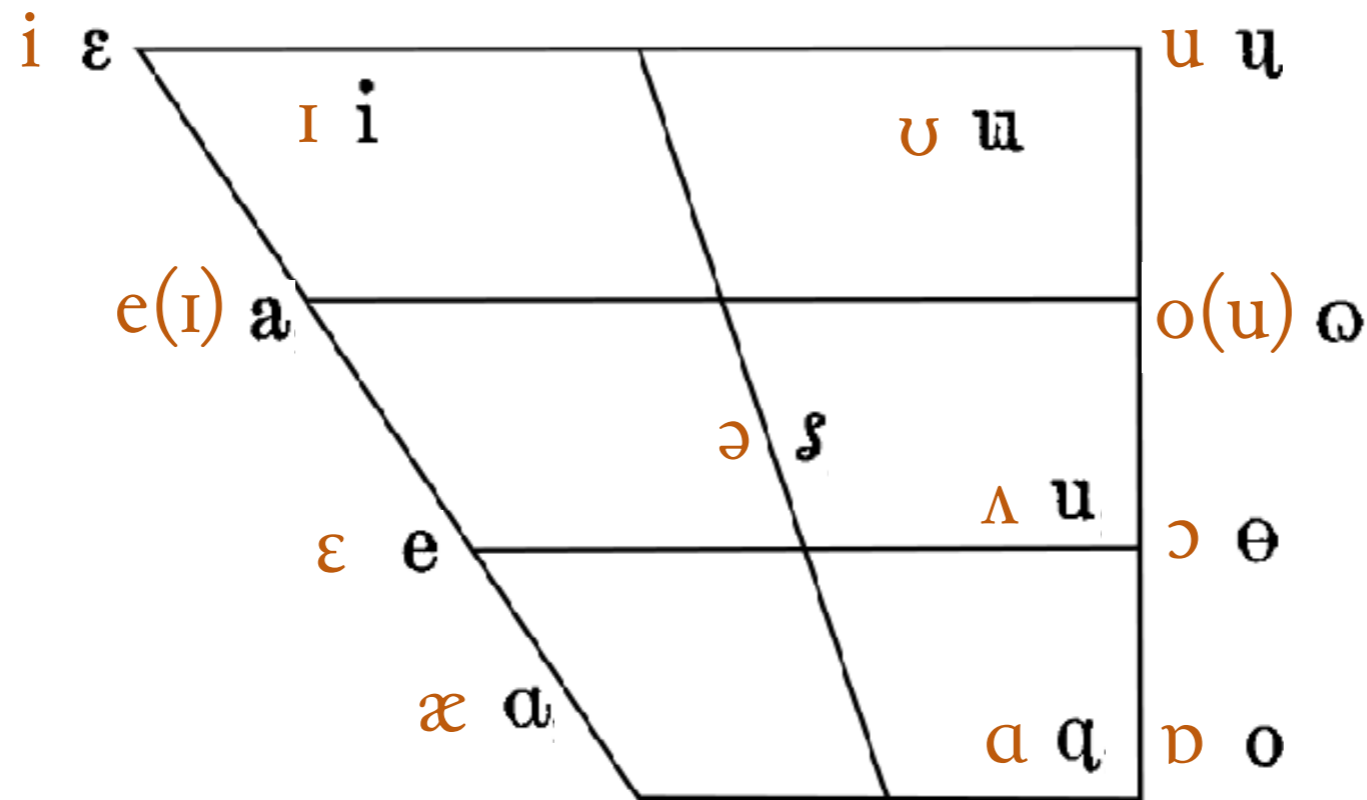
13 To fall into fall water

14 Alapuna

15 Alasggue. An appearance of enke To be—:—: The idewein. To be 18à-

# Bridge's vowel system (rough IPA in orange)

'Long and short vowels':



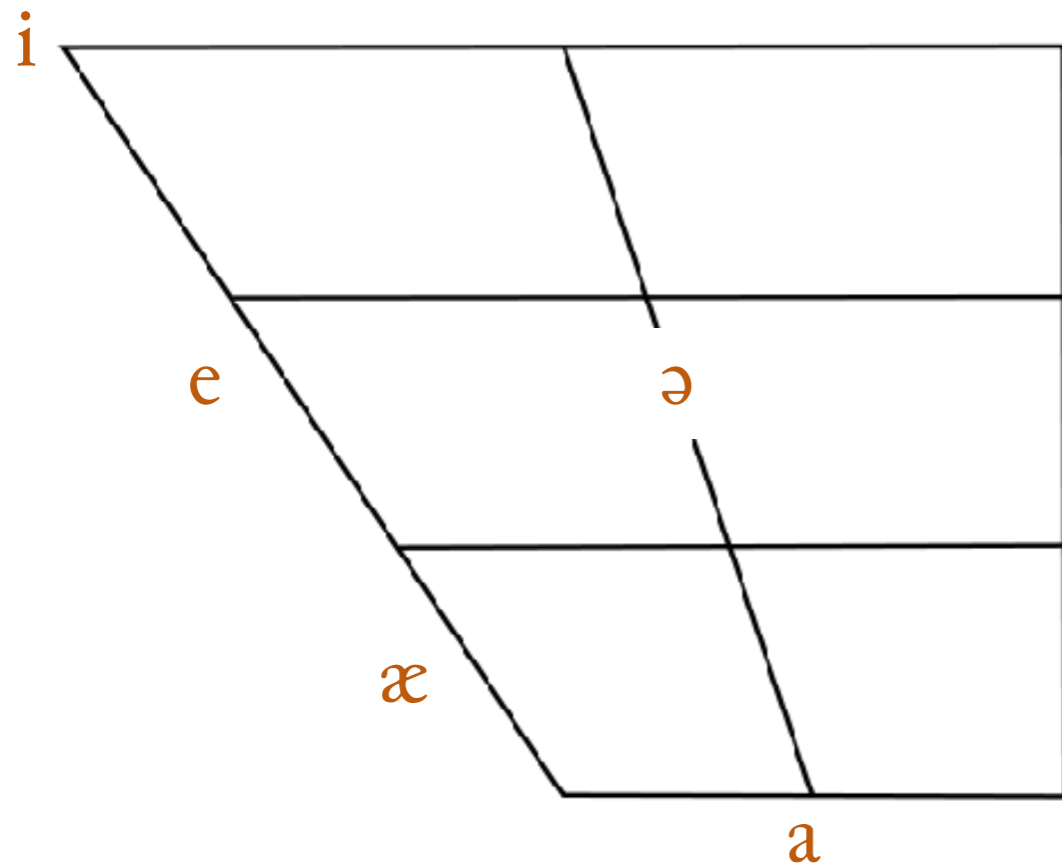
'Diphthongs':

ai j̥ ɔɪ ɔ̃ au ɝ

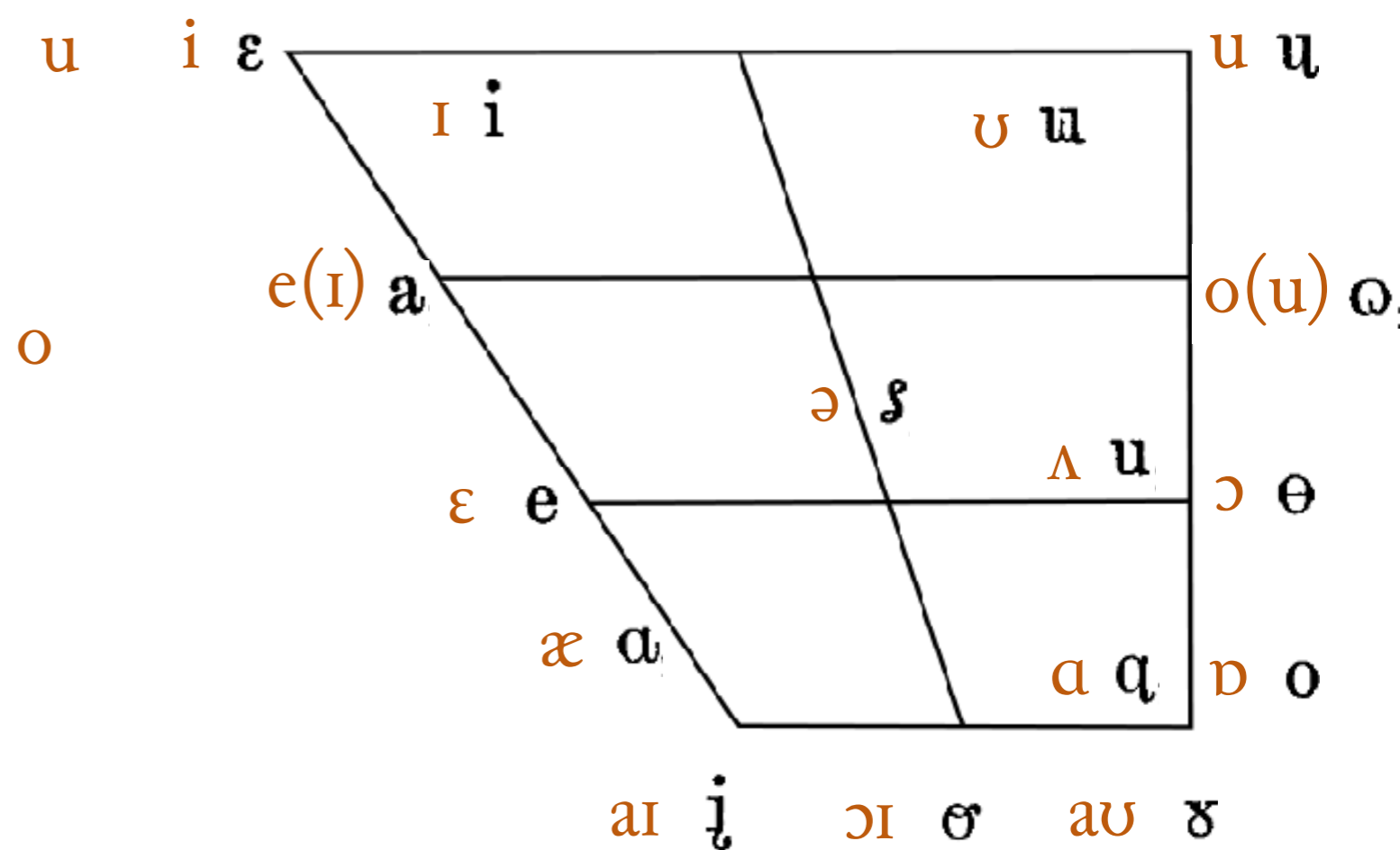
soljs  
tɜs  
anjɪ  
Gɔd

A a, a in <i>haze</i> .	E e, e in <i>men</i> .	Ɛ ε, ea in <i>each</i> .	I i, i in <i>ill</i> .
ʌ ɑ, a in <i>flask</i> .	A ʌ, a in <i>at</i> .	θ e, a in <i>all</i> .	O o, o in <i>ox</i> .
ʊ u, oo in <i>food</i> .	ʊ u, oo in <i>foot</i> .	ɔ̃ σ, oi in <i>oil</i> .	ɥ j̥, i in <i>isle</i> .
ʌ ɑ, oe in <i>hoe</i> .	U u, u in <i>us</i> .	ɝ ɝ, ou in <i>our</i> .	S s, er in <i>herd</i> .

# A clear discrepancy!



Late 20c: 7 vowels



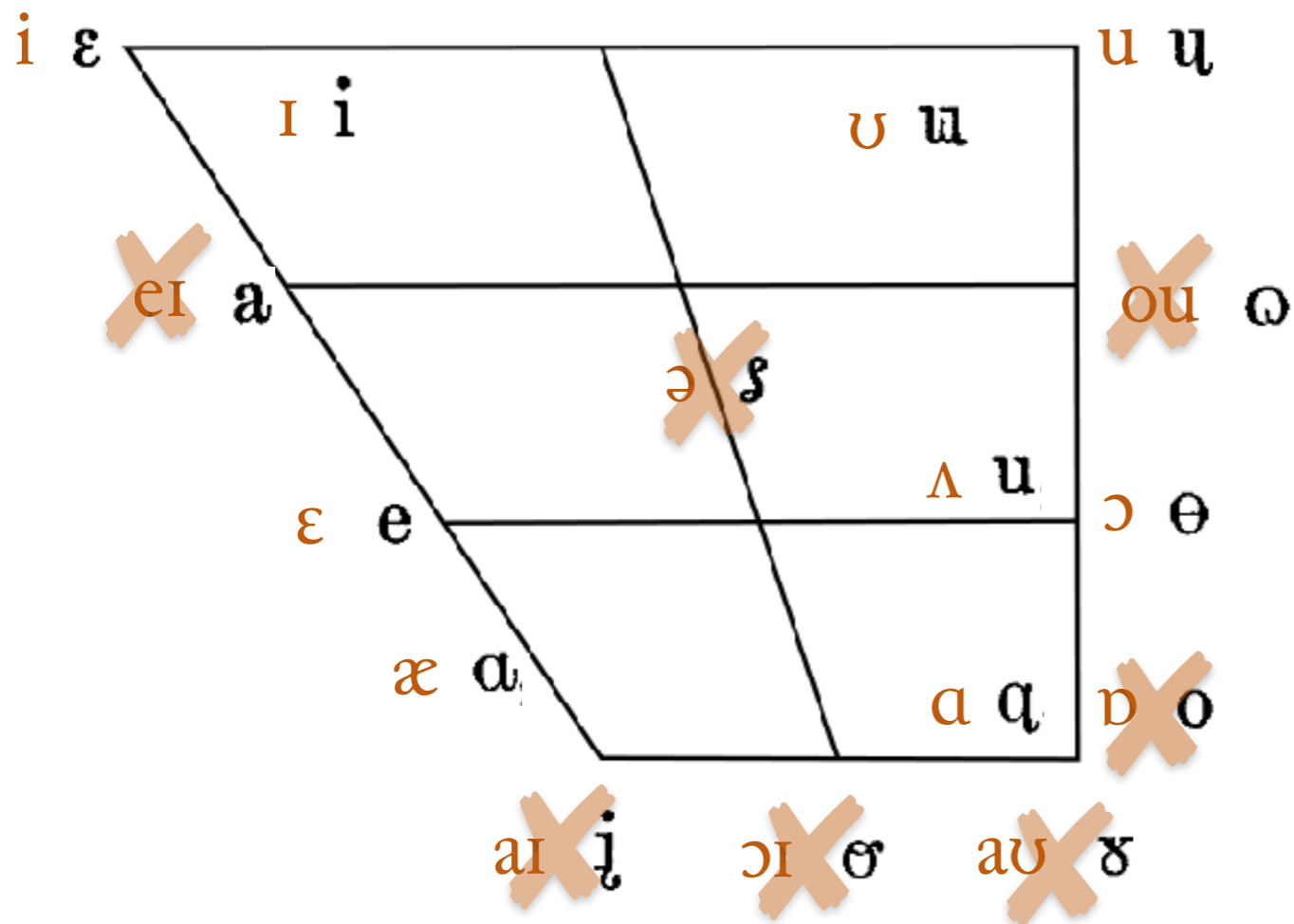
Late 19c: 16 vowels

- Is this sound change? Contact-induced? Result of loss of vitality?
- Descriptive differences?

# Attestations of Bridges' vowels

Ellis	IPA	Native Vocab	English Vocab
<b>a</b>	<b>e</b> i	me <b>e</b> iækλnætikæ 'fear not'; we <b>e</b> i 'may/can'	<b>a</b> ngel, <b>G</b> abriel
<b>ε</b>	<b>i</b>	ikæmænæ 'draw/write'; ki <b>i</b> tu 'self'	Jude <b>a</b> , prie <b>s</b> t
<b>ι</b>	<b>a</b> i	ka <b>a</b> iide 'swiftly bring'; ha <b>a</b> i 'I'	Christ, <b>f</b> ive
<b>ω</b>	<b>ou</b>	jo <b>ou</b> idag 'strong'; mo <b>ou</b> λnna 'repent'	Clo <b>a</b> k, Jo <b>s</b> eph
<b>υ</b>	<b>u</b>	lu <b>u</b> tʃi 'valley'; ulap <b>u</b> 'bad'	L <b>u</b> ke, Ju <b>d</b> ea
<b>α</b>	<b>æ</b>	il <b>æ</b> 'outside'; ætʃ <b>æ</b> w <b>æ</b> i <b>æ</b> 'spirit/life'	Z <b>a</b> ch., Naz <b>a</b> reth
<b>e</b>	<b>ε</b>	t <b>ε</b> ki- 'see'; maust <b>ε</b> ka 'care for'	Jerusalem, Ma <b>r</b> y
<b>i</b>	<b>i</b>	itʃ <b>i</b> 'bear fruit'; κλsp <b>i</b> x 'spirit'	<b>I</b> srael, ci <b>t</b> y
<del><b>o</b></del>	<b>ɒ</b>	—	Jo <b>h</b> n, Prop <b>h</b> et
<b>u</b>	<b>λ</b>	λtʃi 'foreign'; λ <b>ε</b> k <b>λ</b> s 'sunlight'	Theophil <b>u</b> s
<b>α</b>	<b>α</b>	ba <b>α</b> v 'none'; ja <b>α</b> fæ 'loud'	Isra <b>e</b> l, Pa <b>s</b> sover
<b>θ</b>	<b>ɔ</b>	ɔ <b>æ</b> lai <b>æ</b> l <b>æ</b> 'barren'; m <b>ɔ</b> æ <b>l</b> æ 'day'	Go <b>d</b> , alt <b>a</b> r, la <b>w</b>
<del><b>o</b></del>	<b>ɔ</b> i	(l <b>ɔ</b> imæz 'in order')	—
<b>υ</b>	<b>au</b>	ha <b>u</b> æn 'this'; a <b>u</b> wλn 'true'	ho <b>u</b> r, to <b>w</b> er
<b>ω</b>	<b>u</b>	hæ <b>u</b> n 'also'; utʃi <b>u</b> kænæ 'urge'	bo <b>o</b> k
<del><b>ε</b></del>	<b>ə</b>	—	chapt <b>e</b> r, sold <b>i</b> er

# Re-analysis of Bridges' vowels



$\acute{\imath}$  = [ha-ɪ] = 1S-NOM

$\acute{\sigma}a$  = [ha-uæ] = 1S-POSS

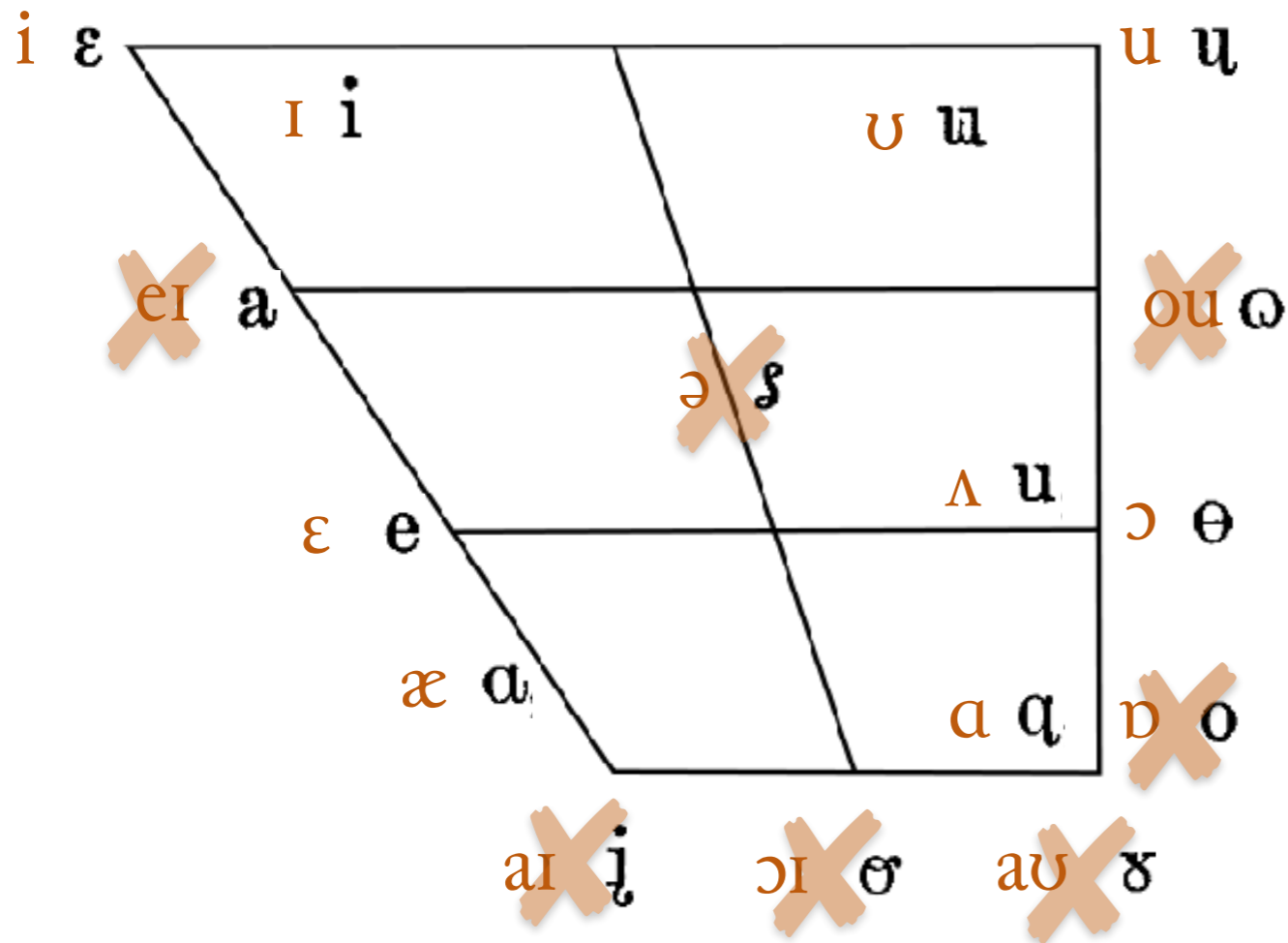
$\grave{\epsilon}c\grave{\imath}a$  = [jika-ɪæ] 'small-3S'

- Diphthongs can have elements belonging to separate morphemes
- As such, the remaining diphthongs must be considered composite elements

**Poblete & Salas 1997: 6**

In careful or deliberate pronunciation, each vowel is a syllable nucleus. Two or more contiguous vowels form separate syllables, as in [sæ̃.pa.im] 'coypu' or [i.la.o.é.a] 'vulture'.

# Re-analysis of Bridges' vowels



**i = [ɪ]** rarely appears in initial syllables

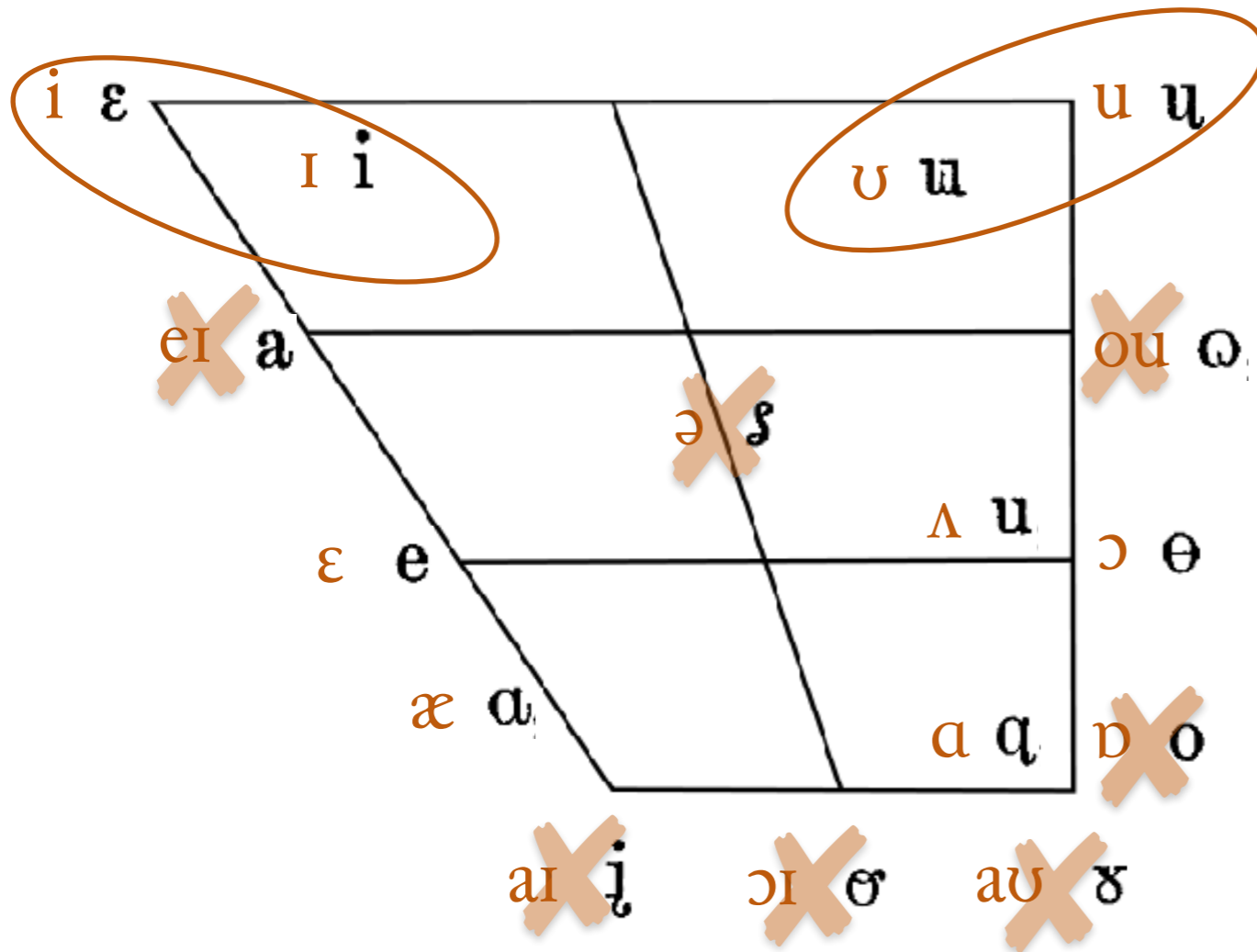
22 eunjin mənqtsieunjin eumunila eutq̄æ eundjananima, eundjan eunisanuda eunjinæ ueq̄n matstu-jealenata, usq̄tupan dagia eueq̄ḡila eundjananima.

**ε = [i]** is mostly initial

63 ej̄ida εcamana tεcamandeca. εεcamandecuda,

*The accent is placed often on the first syllable*

# Re-analysis of Bridges' vowels



ʊ = [ʊ] rarely appears in initial syllables

35 *šacandylam cajquda*, 'šan ša ma<sup>u</sup> ma<sup>u</sup>rua,  
san šla ma<sup>u</sup>fatəa cunjima.

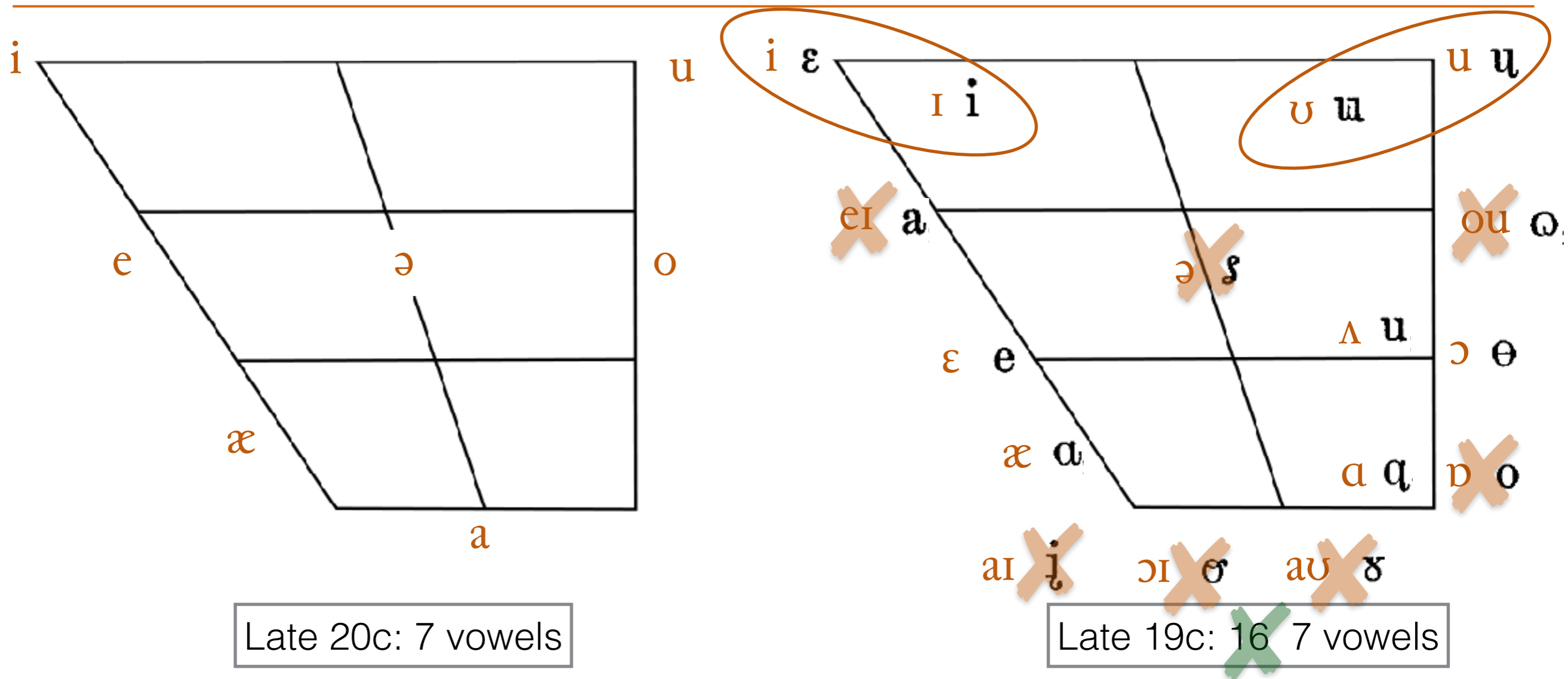
u = [u] is mostly initial

'cunšan <sup>u</sup>lapaej <sup>u</sup>taŋgjatacun?

- Bridge's four high vowels appear to reflect allophonic distinctions
- Alternations appear to be governed by stress position
- Low stress laxing; unstressed vowel reduction; etc.

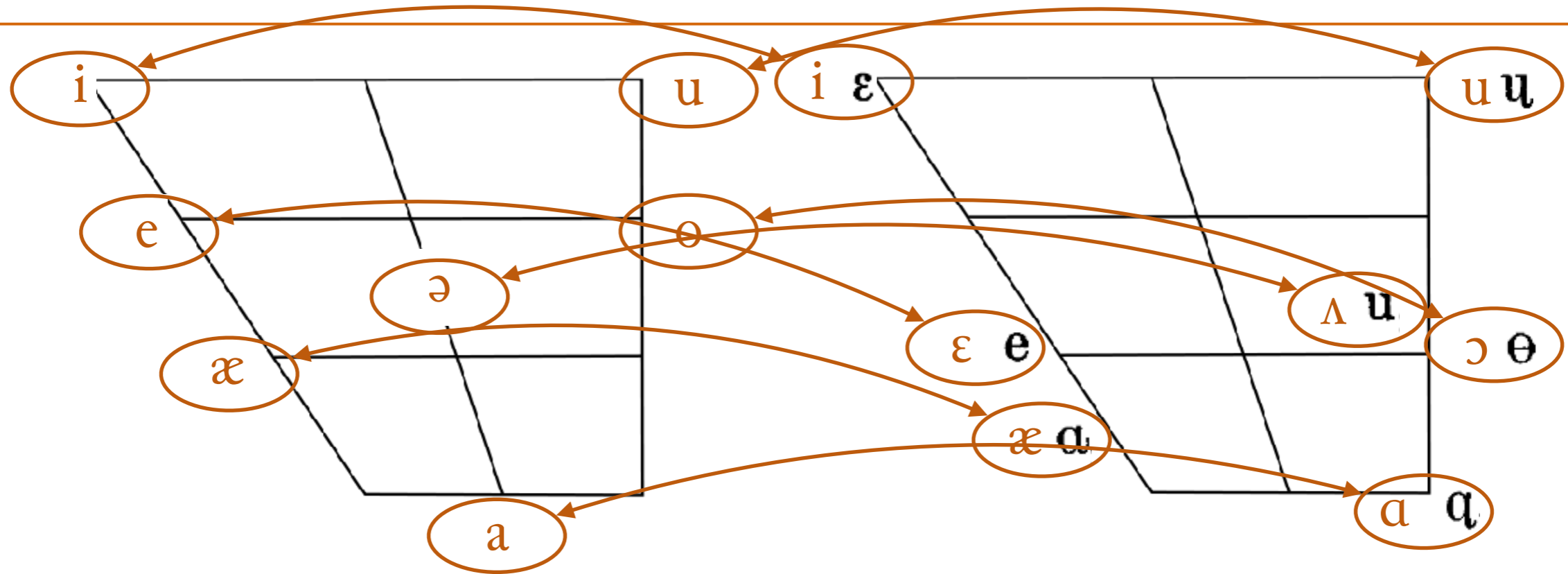
[-STRESS]  
|  
/u,i/ → [ʊ,ɪ]/\_\_\_

# Descriptive differences:



- If we look only at the vowels in native words; treat diphthongs as strings of vowels; and assume stress-based allophony, Bridge's vowel system begins to look a lot like the late 20c data

# Lexical correspondences:



Gloss	Vowel	Salas & Valencia	Bridges
'woman'	<i>i/i</i> ( <i>ε</i> )	[k <i>i</i> pa]	Сера,
'see'	<i>e/ε</i> ( <i>e</i> )	[t <i>e</i> ki]	tec <i>i</i> ,
'river'	<i>æ/æ</i> ( <i>a</i> )	[w <i>æ</i> en]	<i>Wæu.</i>
'arm'	<i>a/a</i> ( <i>a</i> )	[k <i>a</i> mein]	<i>Съмъи.</i>
'one'	<i>o/ɔ</i> ( <i>θ</i> )	[ok <i>o</i> le]	u <i>θ</i> ali
'big'	<i>u/u</i> ( <i>u</i> )	[h <i>u</i> lua]	ú <i>u</i>
'louse'	<i>ə/ʌ</i> ( <i>u</i> )	[w <i>ə</i> n]	<i>Wum.</i>

# Explaining the differences

---

- Descriptive shortcomings
  - Bridges started acquiring the Yahgan quite young (13)
  - He probably had a fairly good grasp of the phonology
  - His goal, however, was always phonetic
  - The Yahgan nicknamed him *Hucheistauwa* ‘deaf one’ due to his constant requests for repetitions during elicitation
  - The *phonotypic alphabet* created biases towards English-like vowels
  - Vowel alternations between high vowels were likely allophonic
  - The abundance of vowel symbols available allows Bridges to capture some alternations

Allophonography, yay?

# When Ellis found out...

---

When I saw Mr. Bridges on his first coming over I told him that my alphabet was not well fitted for his purpose, but a very great mistake has been made I think in altering and patching it up in this extraordinary manner.

A.J. Ellis, 'President's Report', The Philological Society, 1884:33-34

But from the language being written in an alphabet which I had a chief hand in inventing, I have cherished a kind of parental feeling towards it, and have therefore endeavoured to put Mr. Bridges' account into a suitable form. The first need was to change the alphabet, so as to avoid his new letters. The phonetic alphabet of 1846, which was the one Mr. Bridges adopted, was founded on the English vowel analogies, and though, as was proved by much printing and teaching, exceedingly well adapted for the English language, was not suitable for missionary purposes.

A.J. Ellis, 'President's Report', The Philological Society, 1884:33-34

# Some conclusions

---

- The documentary materials for 19th century Yahgan are a rich treasure-trove of linguistic data
  - The writing system, however, reflects Jespersen's idea of an **articulation basis**, in this case, in favour of English
- Despite this, comparison with contemporary data for the vocalic system of Yahgan suggest the language underwent fairly minimal changes in the last 150 years
- The English Phonotypic Alphabet, on the one hand, is not a typical alphabet, as it lack the redundancy of typical *litteral* sets
- It is firmly encouched in English phonology and phonetics
- We can only observe the allophony – and even potential phonemic categories – insofar as they are part of the set of English

# Allophonography

---

- Native writing systems, both standard and non-standard, tend towards redundancy and are often on a cline towards morphography or logography as spelling systems gain 'depth'
- For native writing, sub-phonemic distinction (allophonography) is rare
- The opposite is not true, with some phonemic contrasts lacking direct expression,
- This is in line with the mnemonic function of native writing
- Littera-based, 'proper' alphabets, don't seem to disregard phonetic detail, on the contrary, they represent language-specific articulatory targets
- When used to script a new language, a proper alphabets does not rely on 'graphemes' as parallels of contrastive phonemes, but on the concrete settings of the source language (the potestates of each littera)
- Allophonography emerges mostly in non-native contexts, where an alphabetic set is applied to a new language without the mediation of the target language phonology





**Angus McIntosh Centre**  
for Historical Linguistics

Thanks!



THE UNIVERSITY  
of EDINBURGH