



Angus McIntosh Centre
for Historical Linguistics

Comparative
Historical Corpus
of the Southern Cone



Texts, typology and history

Digital methods and the languages of the Southern Cone

Benjamin Molineaux
benjamin.molineaux@ed.ac.uk

03 February, 2026



THE UNIVERSITY
of EDINBURGH

Plan

- Context of our work (4-year project)
- Colonial histories in the Southern Cone
- The linguistic situation today
- Changing narratives and digitality
- Historical linguistics and human history
- Typology and our common humanity
- Historiography, heritage and communities
- Some conclusions



Highlands of the far south

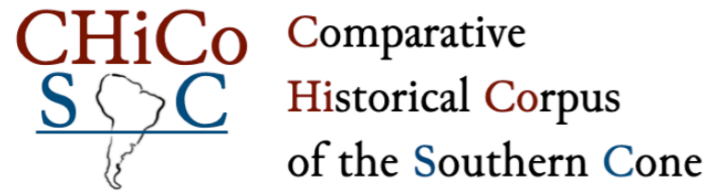


Highlands of the far north



CONTEXT (Our Project)

A four-year project



Texts, Typology & Language History In The Southern Cone:

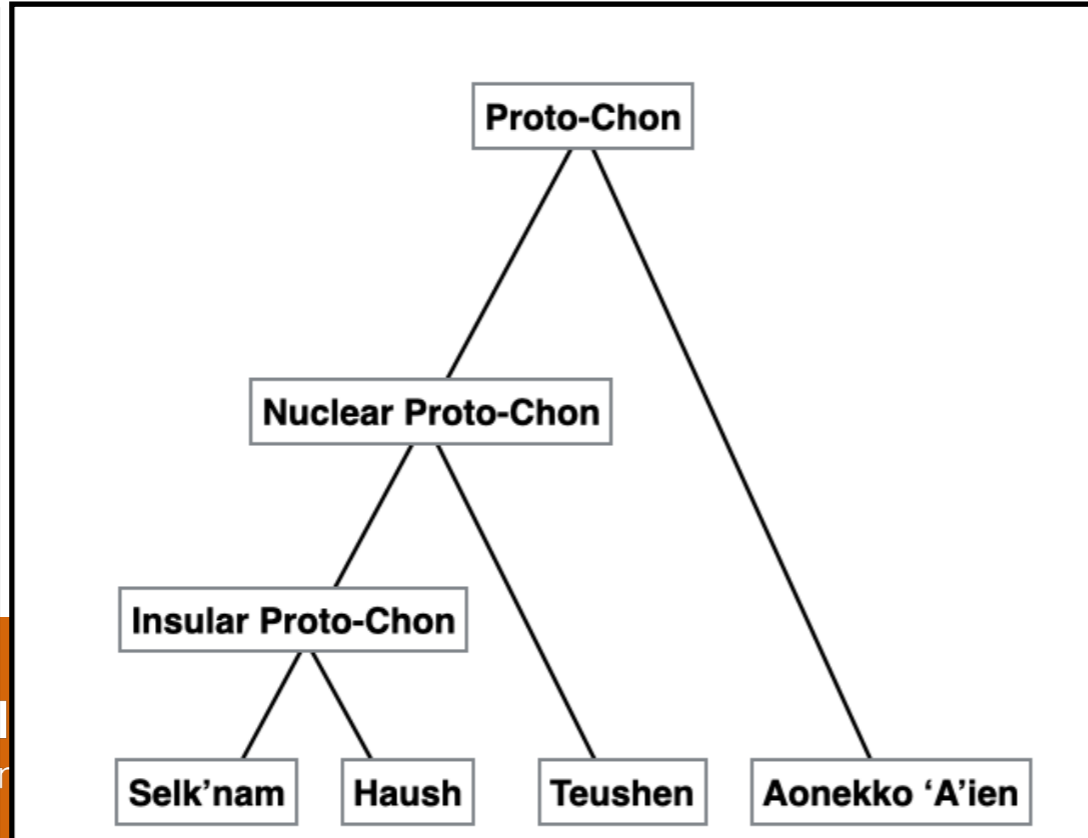
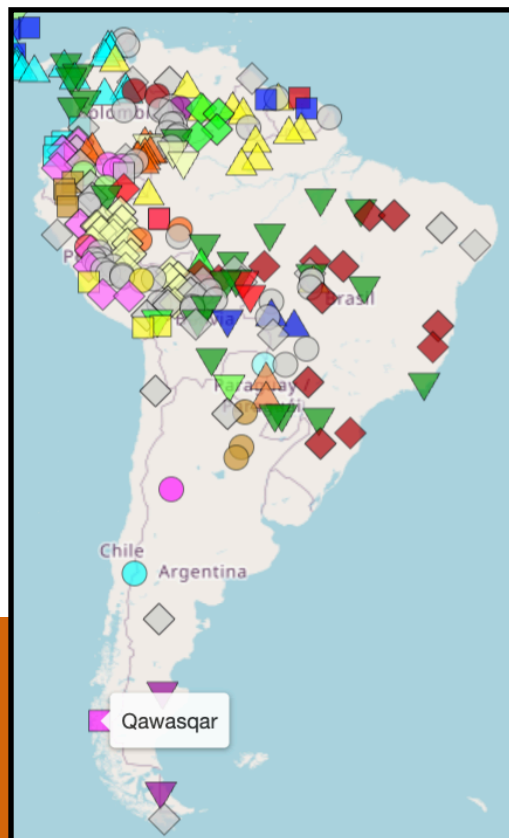
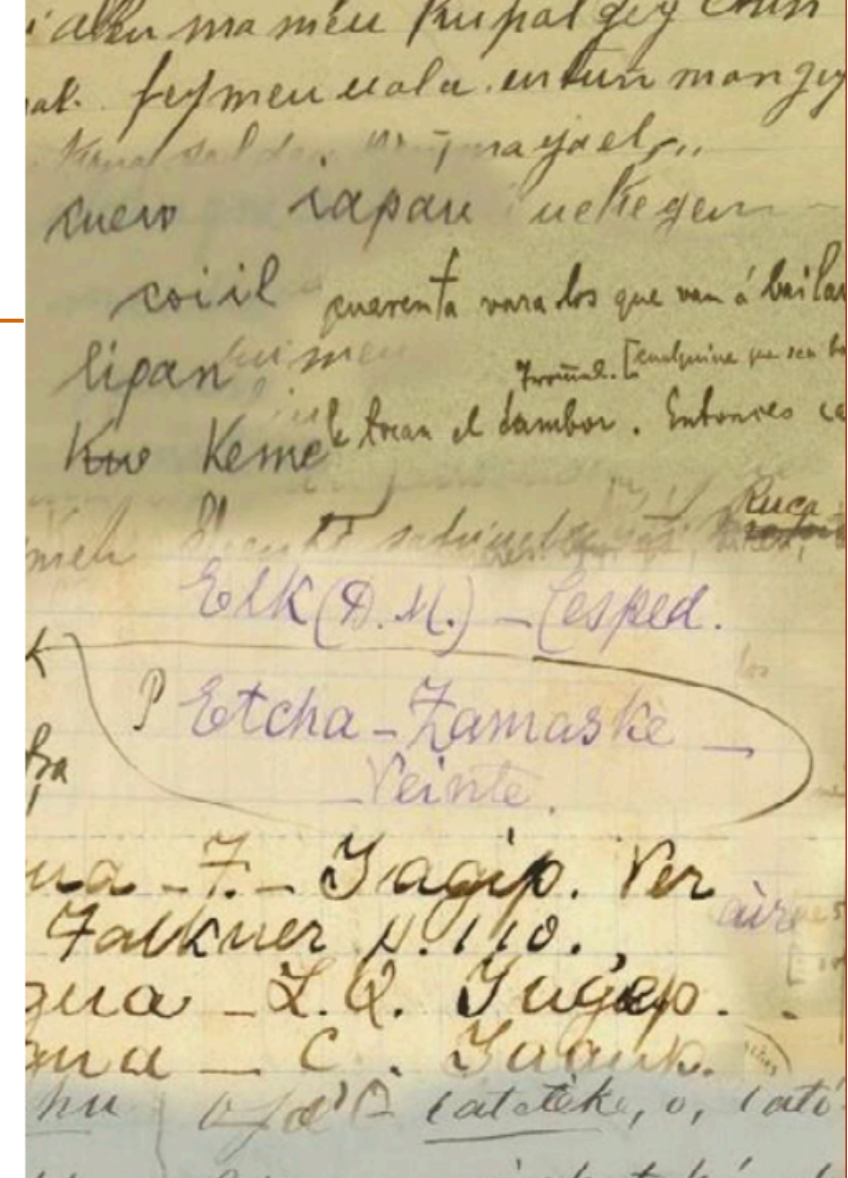
A DIGITAL FRAMEWORK

This website documents the progress of a four-year *Leverhulme Trust Research Project Grant* looking into the textual materials for languages of Patagonia and Tierra del Fuego. Its main output will be the *Comparative-Historical Corpus of the Southern Cone (CHiCo-SC)*, which will help us explore language typology and human history in the region

LEVERHULME
TRUST

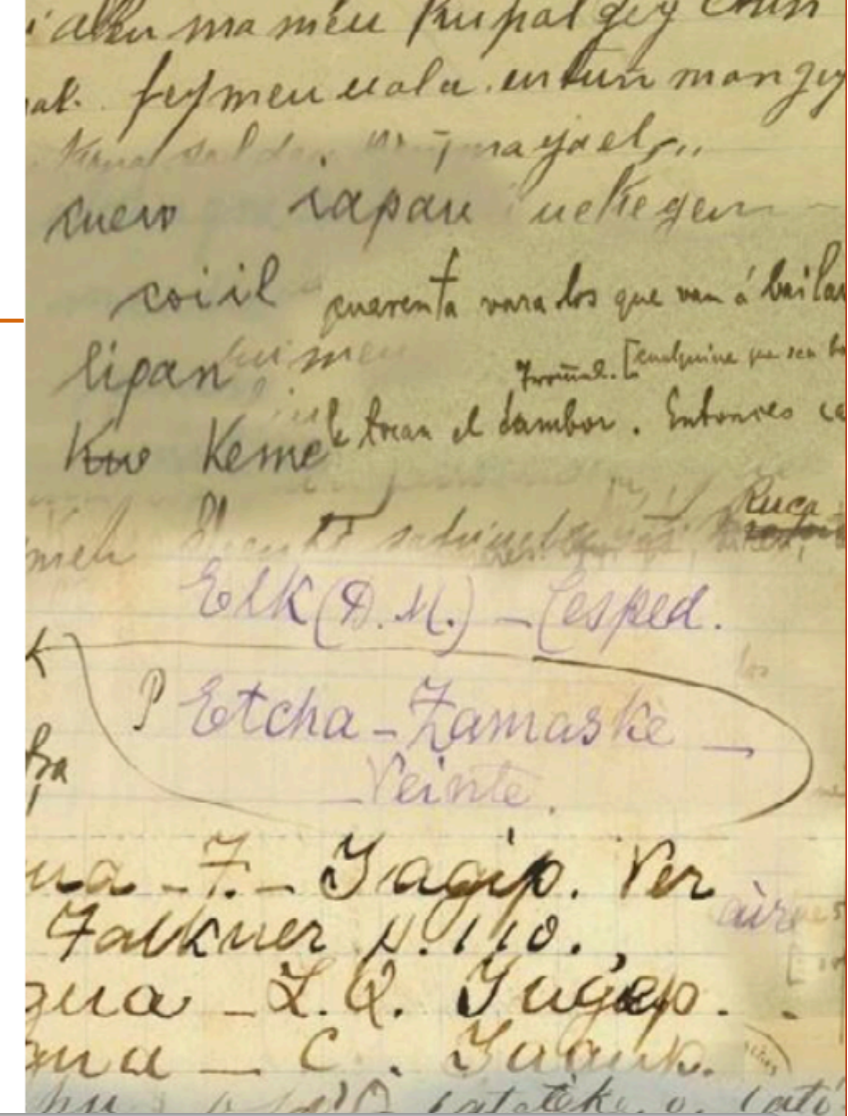
Project Goals

- Systematise and disseminate historical documentation for languages of the Southern Cone
 - Academic communities**
- Identify areal and typological linguistic features
- Reconsider genealogy of language families
- Integrate data with archaeology, anthropology and genetics, to better understand local prehistory



Project Goals

- Systematise and disseminate historical documentation for languages of the Southern Cone
 - Indigenous communities**
 - General public**
- Contribute to highlighting and revitalising these languages, creating materials for their heritage communities



1910LA-2-Relaciones

Not Secure amc-resources.lel.ed.ac.uk/CHM/data/1910LA-2-Rel...

Impresiones que el ciego José Francisco Koluñ llevó de la Misión, manifestadas en una conversación habida con sus padres, de regreso á su casa.

A. Conversación con la madre.

1.	„Kiñe akucha regaláneu Padre”, pin iñche.	„¿El prendedor que se pidió al Padre para mí?!” me dijo mi madre.
	„¿Ni ñillatuletéu chi akucha Padre meu? tañi ñuke.	LEMMMA: ñuke POS: N ENG: mother SPA: madre MORPH: ñuke mother (madre)
	„Təfa yefalelpáimeu kiñe akucha Padre” ñuke.	ni madre no sabía prenderlo.
	Iñche elufiñ; iñche ñi ñuke kimlafui rətri	conversación refirió así: (Mi madre me vió mi
2.	Kañelu nùtram fei pi: (Ñi rosario peñmáneu ñi ñuke .)	„¿Para mí no pediste al Padre un rosario?”, me dijo también mi madre.

Kimeltuwe, materiales de mapudungun

18 March 2025

Reñma. Family

Victor Carlaf | /kimeltuwe | Fierstoforo

Academic Collaborators



THE UNIVERSITY of EDINBURGH
School of Philosophy, Psychology
and Language Sciences

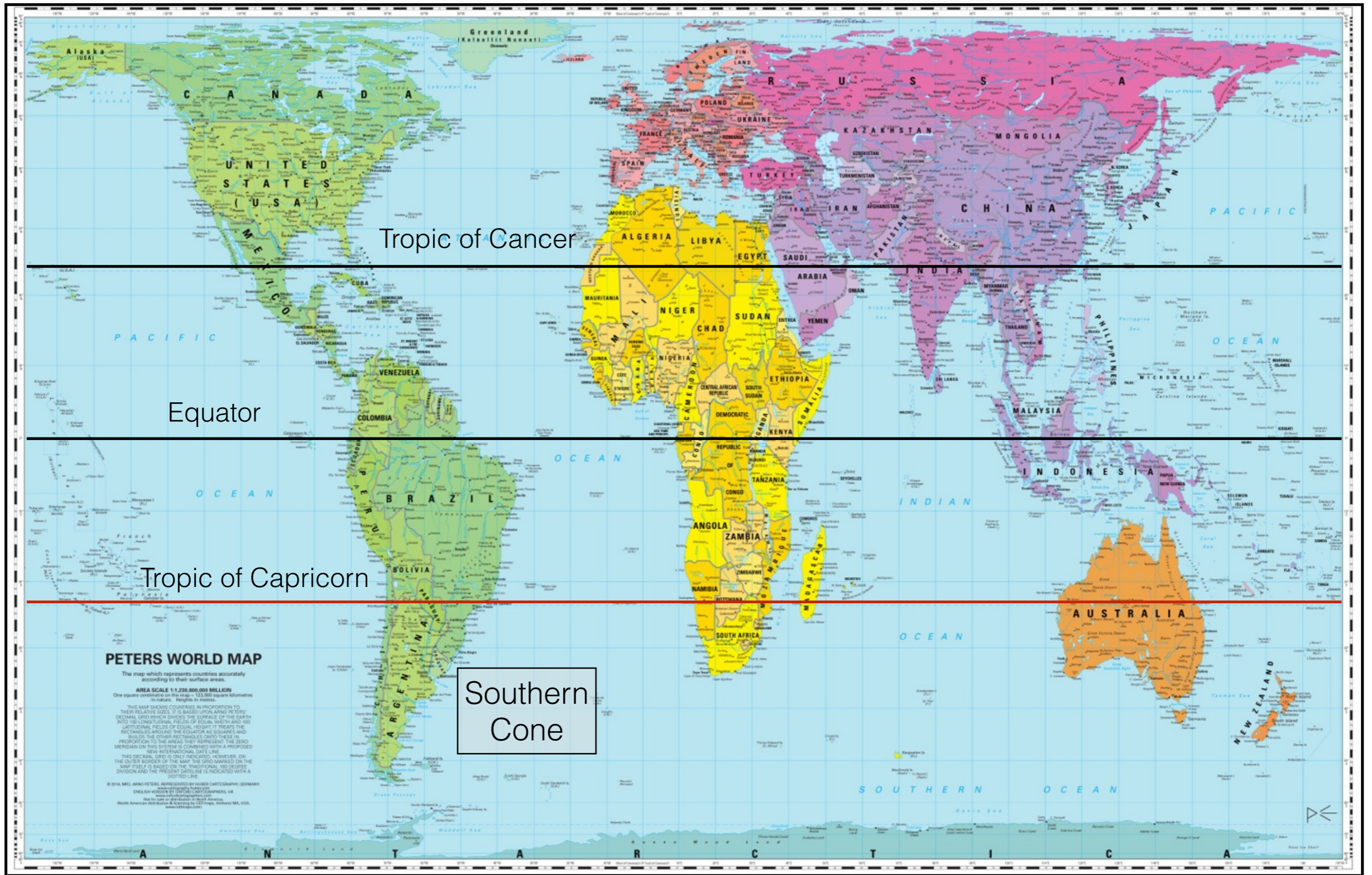


UNIVERSIDAD
DE LA FRONTERA



Universidad Nacional
de **Río Negro**





A problem...

WIKIPEDIA 25 years of the free encyclopedia

Search Wikipedia Search

Languages of South America

8 languages

Contents hide

(Top)

- Main languages
- Indigenous languages
 - Classification
- Other non-indigenous languages
- See also
- References
- External links

Article Talk

Read Edit View history Tools

From Wikipedia, the free encyclopedia

The languages of **South America** can be divided into three broad groups:

- the languages of the (in most cases, former) **colonial powers**, primarily **Spanish** and **Portuguese**;
- many **indigenous languages**, some of which are co-official alongside the colonial languages;
- and various pockets of other languages spoken by immigrant populations.

Main languages [edit]

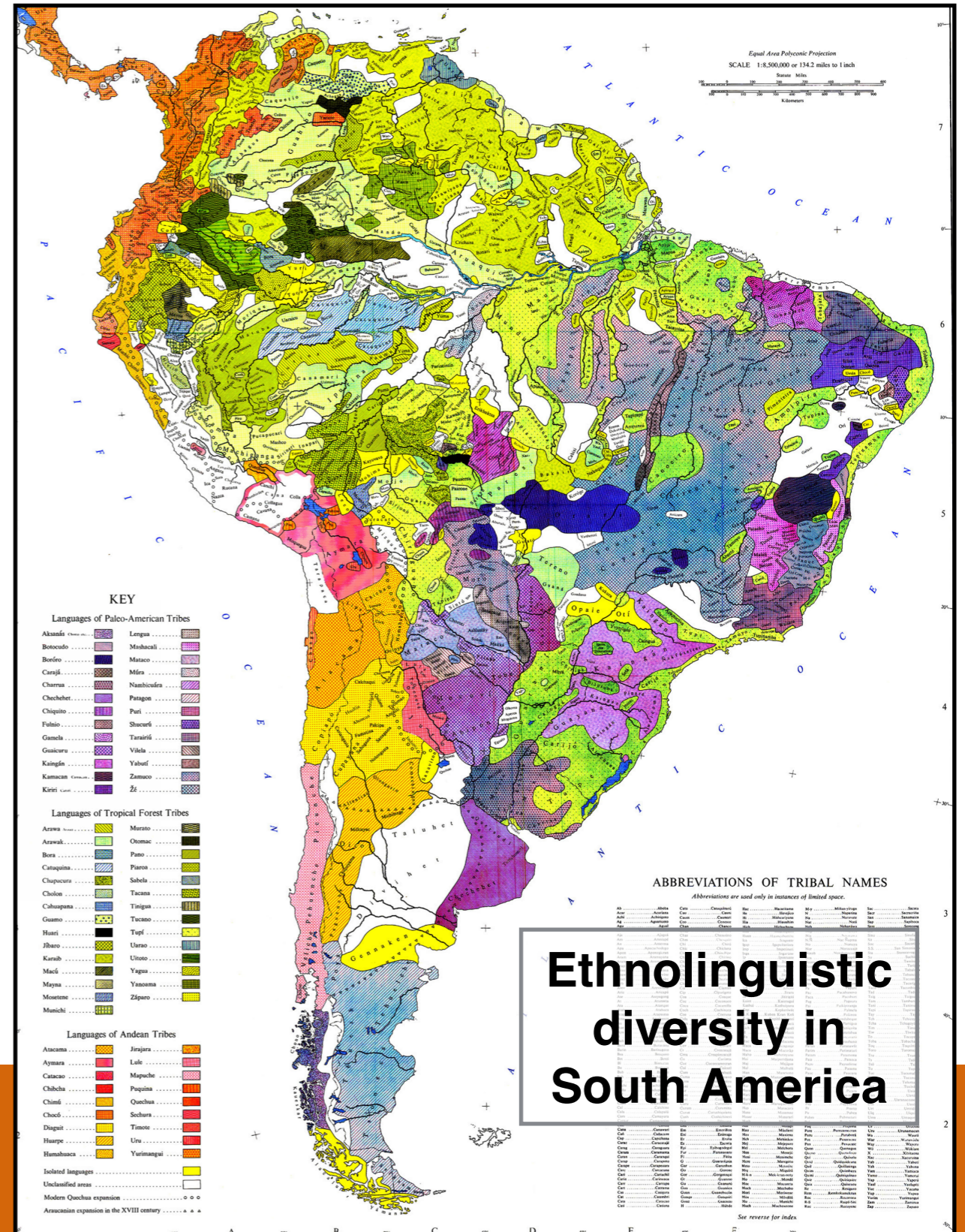
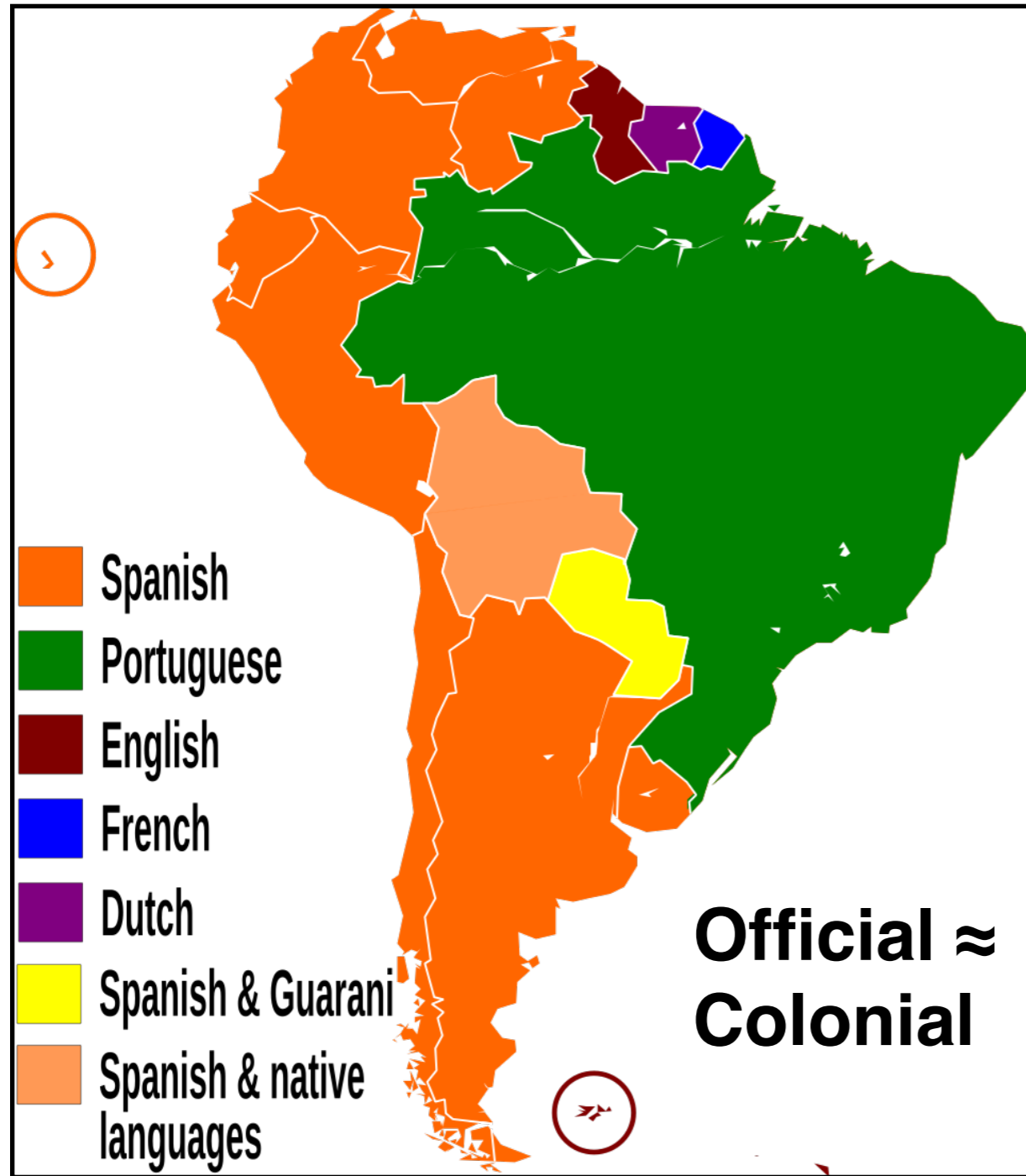
Spanish is the most spoken language in the **Americas**, but Portuguese is the most spoken language in the continent of South America, with Spanish as a close second in South America.

Other official languages with substantial number of speakers are:

- Aymara** in **Bolivia** and **Peru**
- Guaraní** in **Bolivia** and **Paraguay**
- Quechua** in **Bolivia**, **Ecuador**, and **Peru**

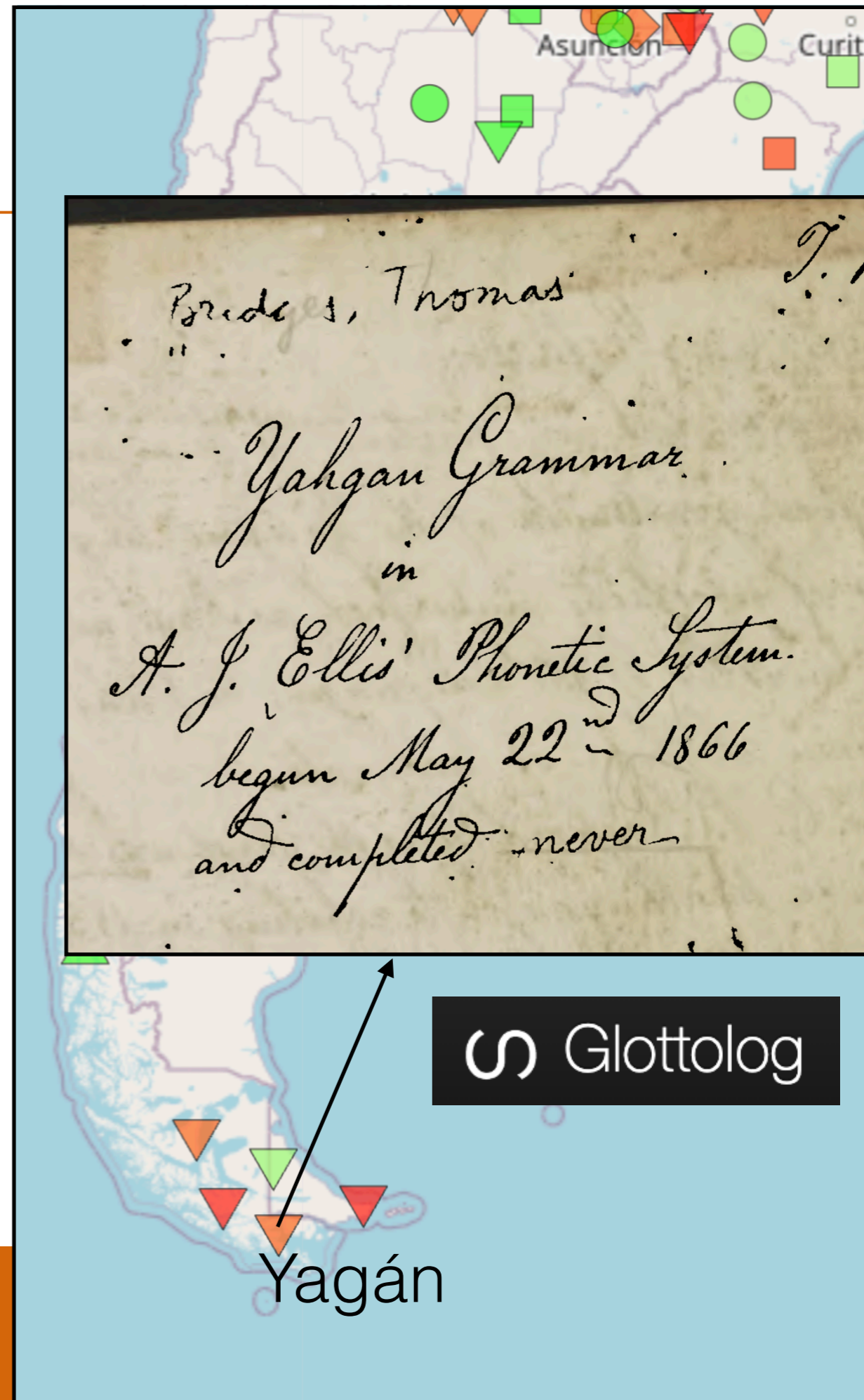
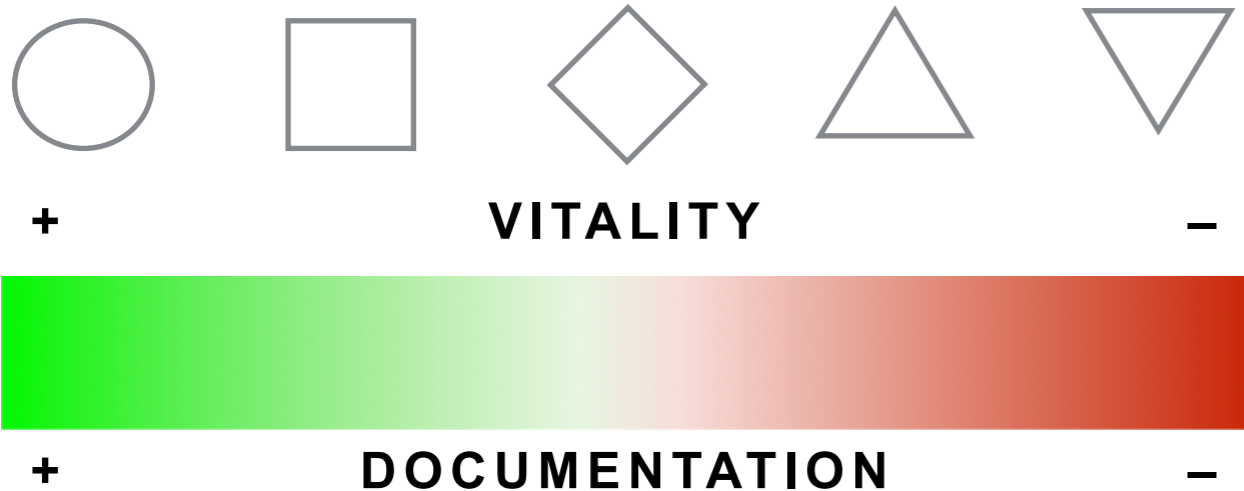
Official languages in South America

Official v. Indigenous languages



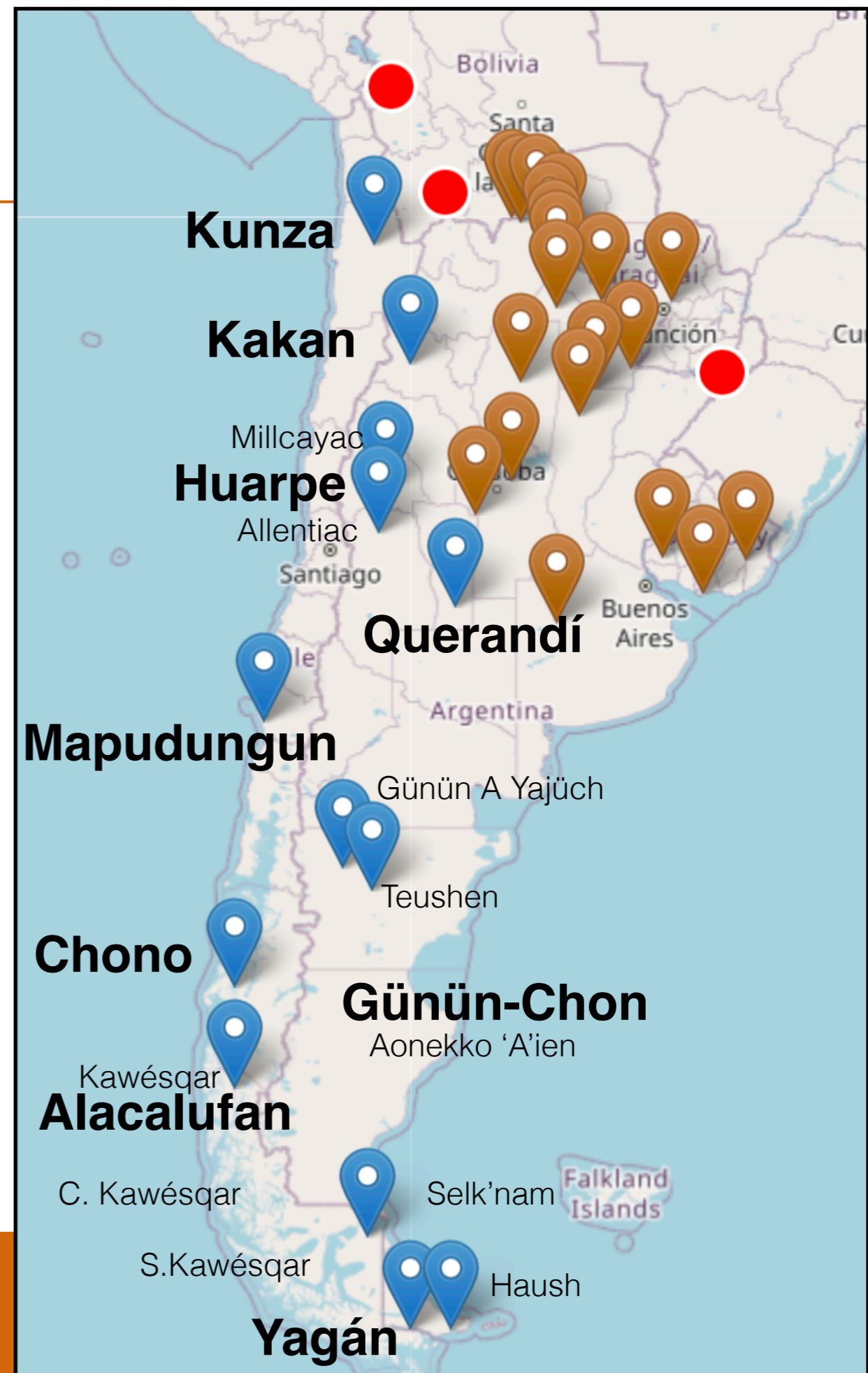
Southern Cone languages

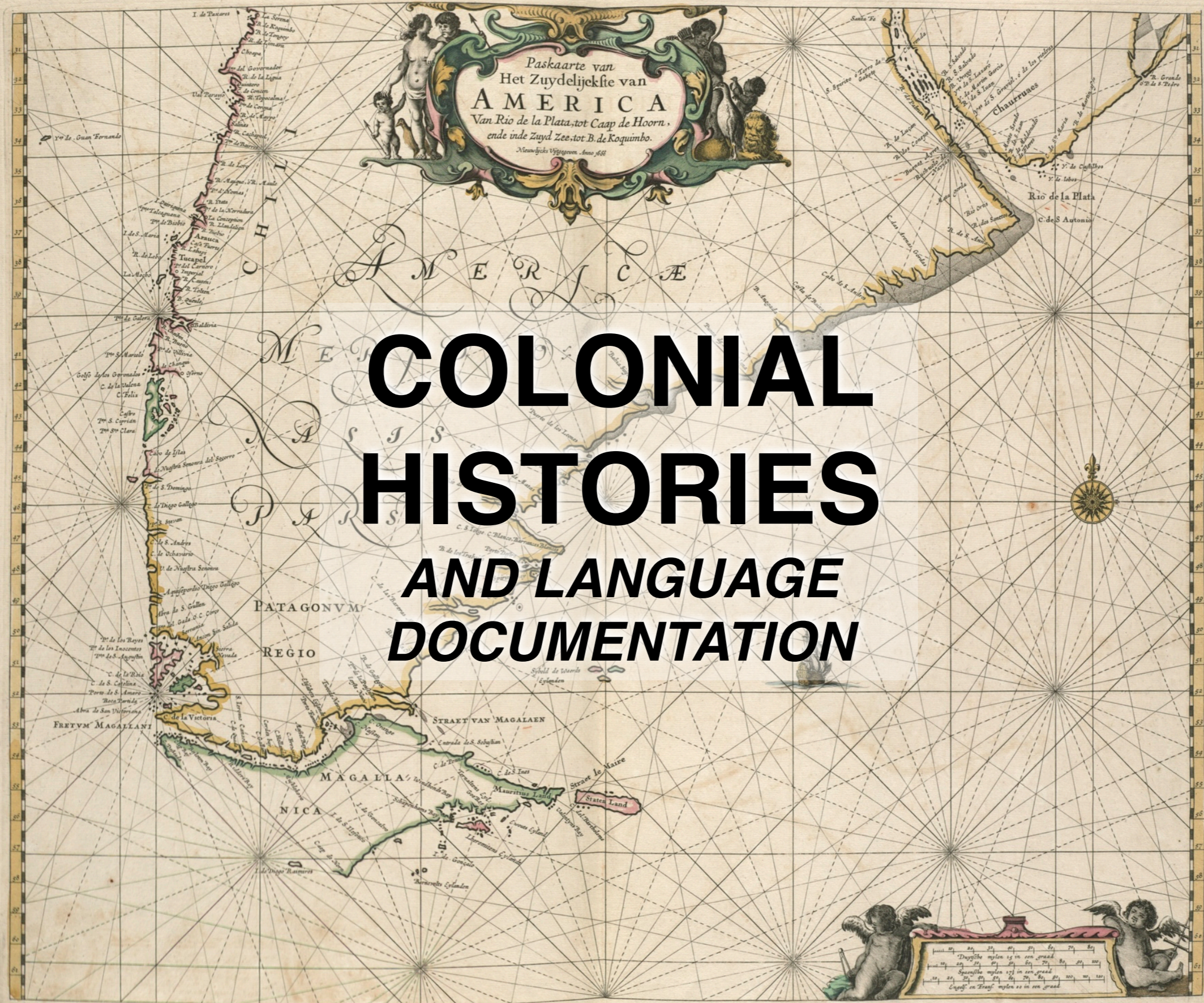
- Traditionally, an area of lower population and language density
- Today, native languages of the area have lost ground to colonial languages — they are mostly *dormant*
- Available documentation is imbalanced
 - And yet...
 - Heritage communities remain
 - Much documentation is unexplored



Target Languages

- **Alacalufan** (Kawésqar; Central Kawésqar; Southern Kawésqar)
- **Chono**
- **Gүнүн-Chonan** (Aonekko 'A'ien; Selk'nam; Haush; Teushen; Gүнүн A Yajüch)
- **Huarpe** (Allentiac; Millcayac)
- **Kakan**
- **Kunza**
- **Mapudungun**
- **Querandí**
- **Yagán**
 - **Chaco/buffer zone languages**
 - **Quechua, Aymara, Guaraní (better supported)**





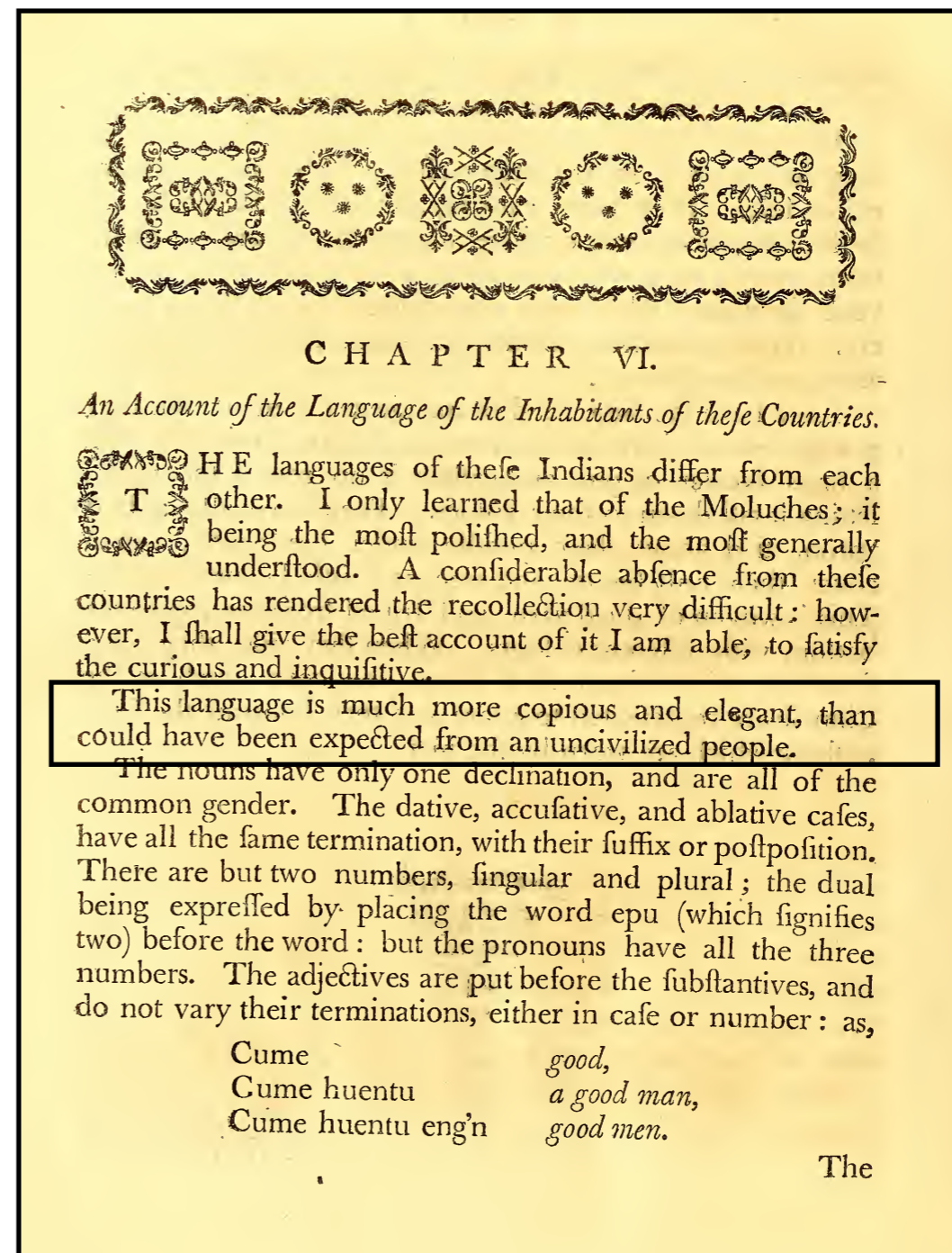
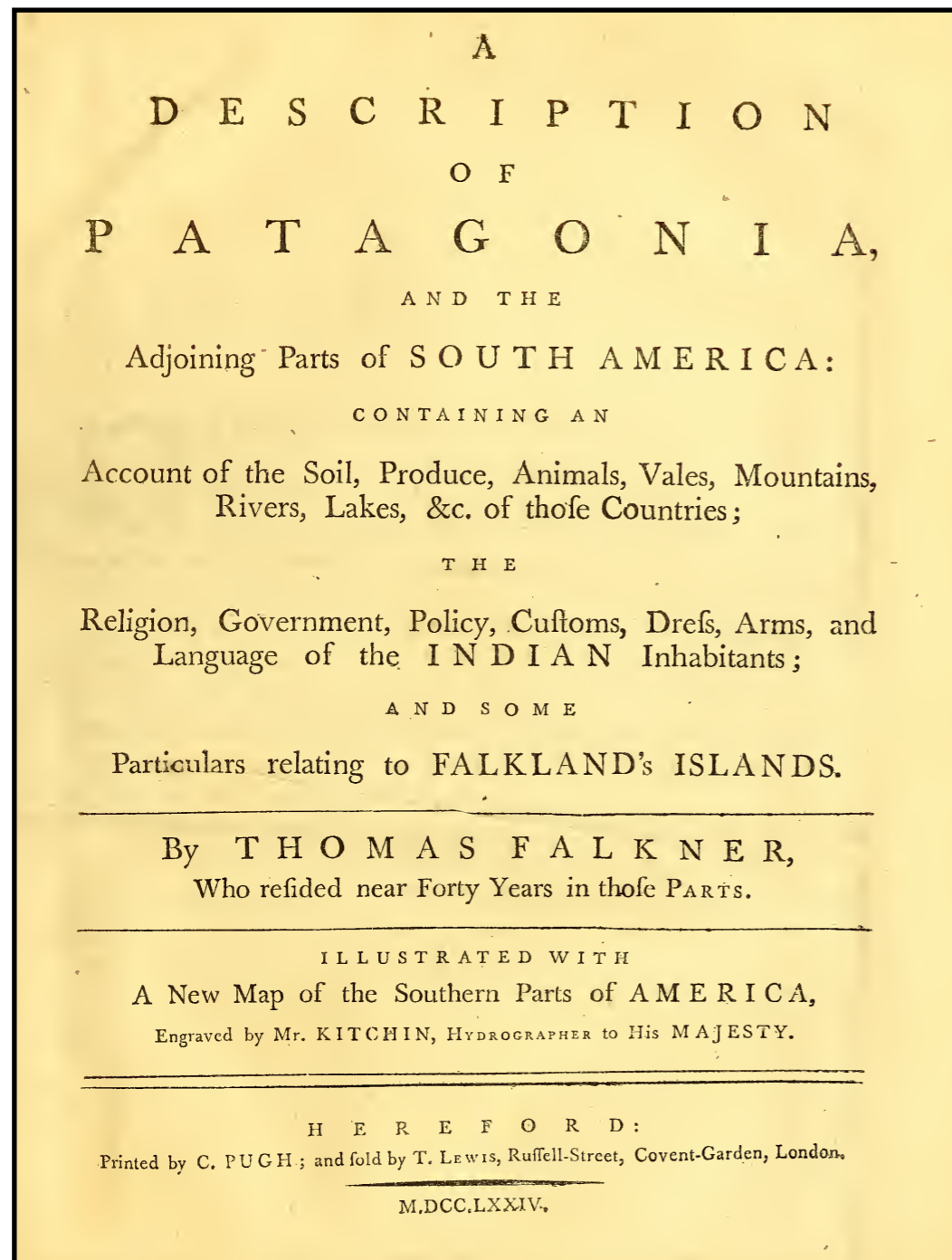
COLONIAL HISTORIES AND LANGUAGE DOCUMENTATION

How we got here...



- A familiar story
- European expansionism
- Indigenous peoples of the Southern Cone were not centralised
- Protracted guerrilla warfare
- Limited commercial interest for the Crown
- Evangelisation, acculturation as official strategies
- De-facto independence

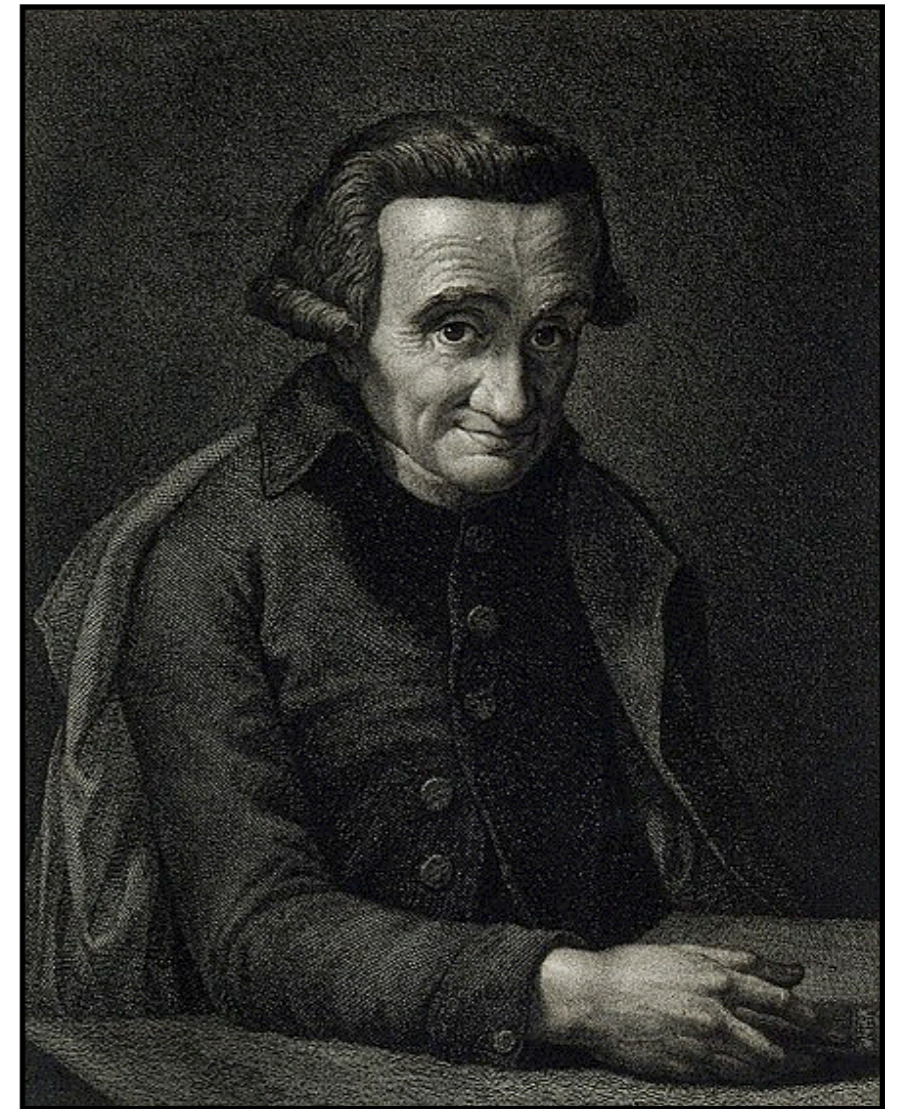
Missionary linguistics: *Admiration and othering*



A narrative of decline

On examining the harmony and richness of their language, we are naturally led to conclude that the Chilians must have, in former times, possessed a much greater degree of civilization than at present; or, at least, that they are the remains of a great and illustrious nation, ruined by some of those physical or moral revolutions so common to our globe.

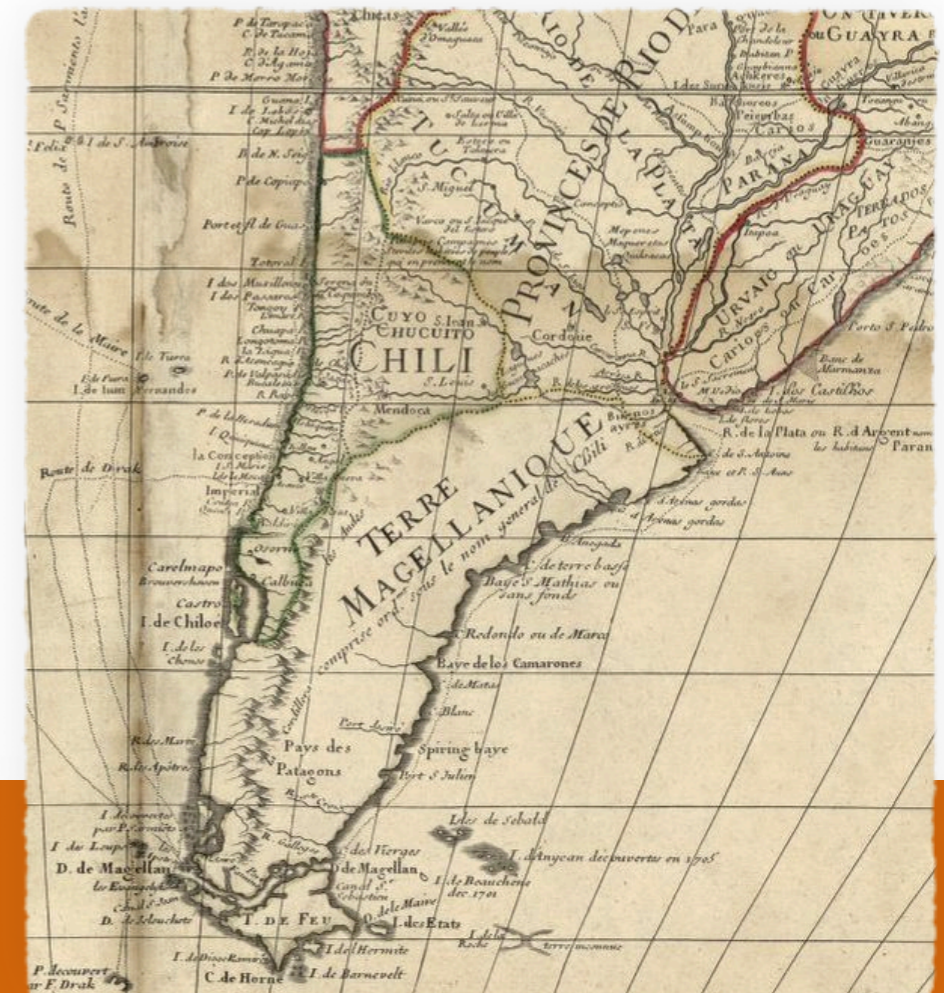
(Molina 1808[1776])



Independence! (for some)

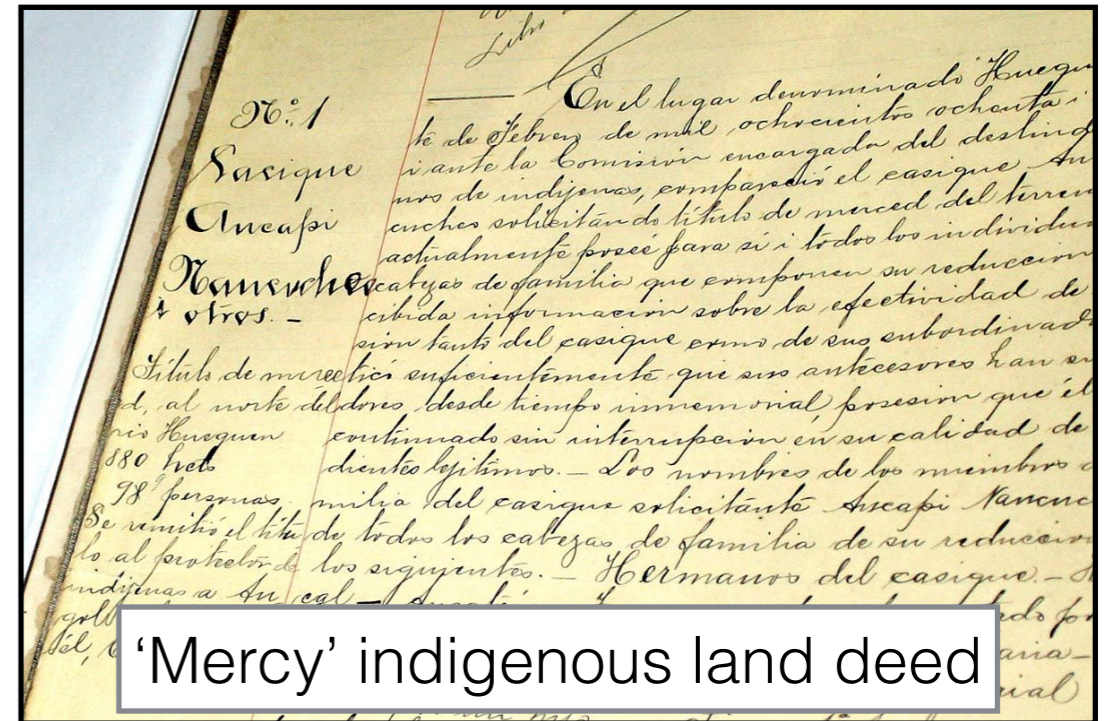


- 19c independence movements
- This was led by American-born descendants of Europeans
- Their rise to power led to the consolidation of territories



Reduction

- Various 19c territorial campaigns
 - ‘Conquest of the desert’ (Argentina)
 - ‘Pacifying of Araucanía’ (Chile)
 - ‘Clearing of Tierra del Fuego’ (Chile/Argentina/UK)
- Forced relocation, territorial usurpation, erasure, genocide



‘Mercy’ indigenous land deed



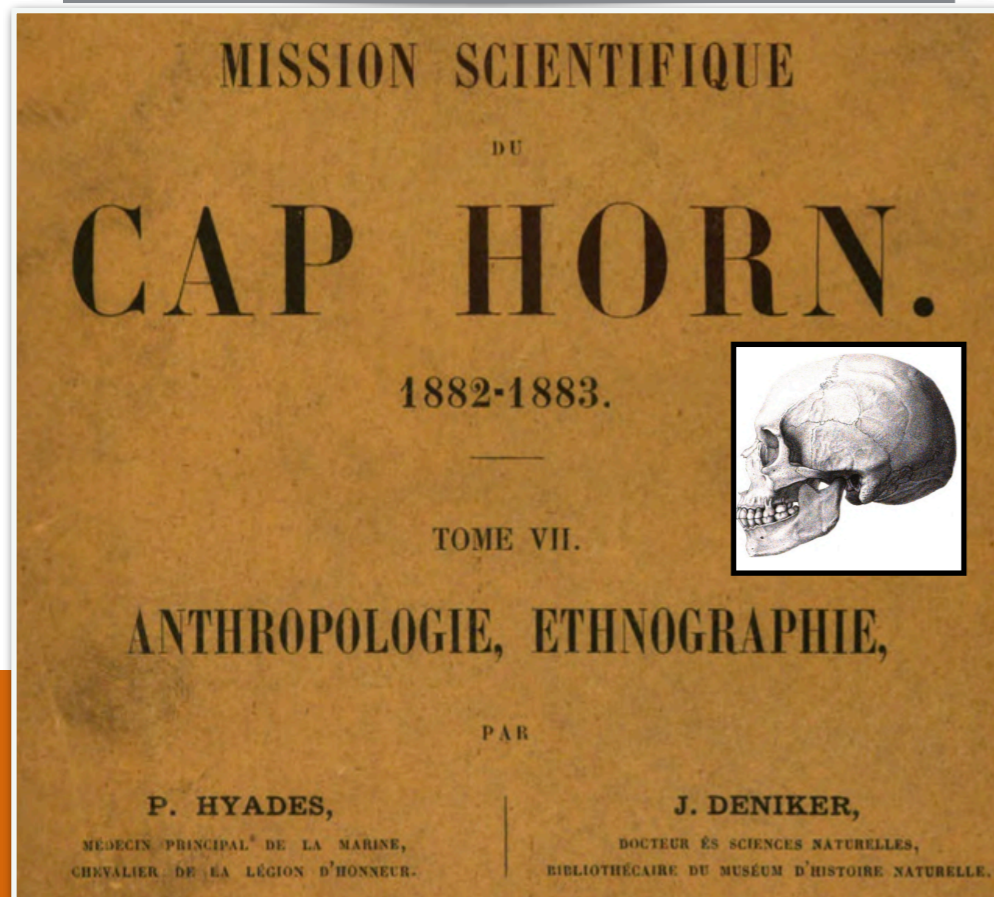
Indigenous concentration camp, 1879, Río Negro, Argentina | Selk'nam 'hunters', c.1900, TdF

Anthropologists, entrepreneurs and more missionaries



Fr. De Agostini and Pacec (Selk'nam)

- Much of the 19th and 20th century documentation is tainted with 'exoticism'
- A 'narrative of extinction' becomes established



Anthropologists, entrepreneurs and more missionaries



Fr. Th Bridges, with Yahgan kids

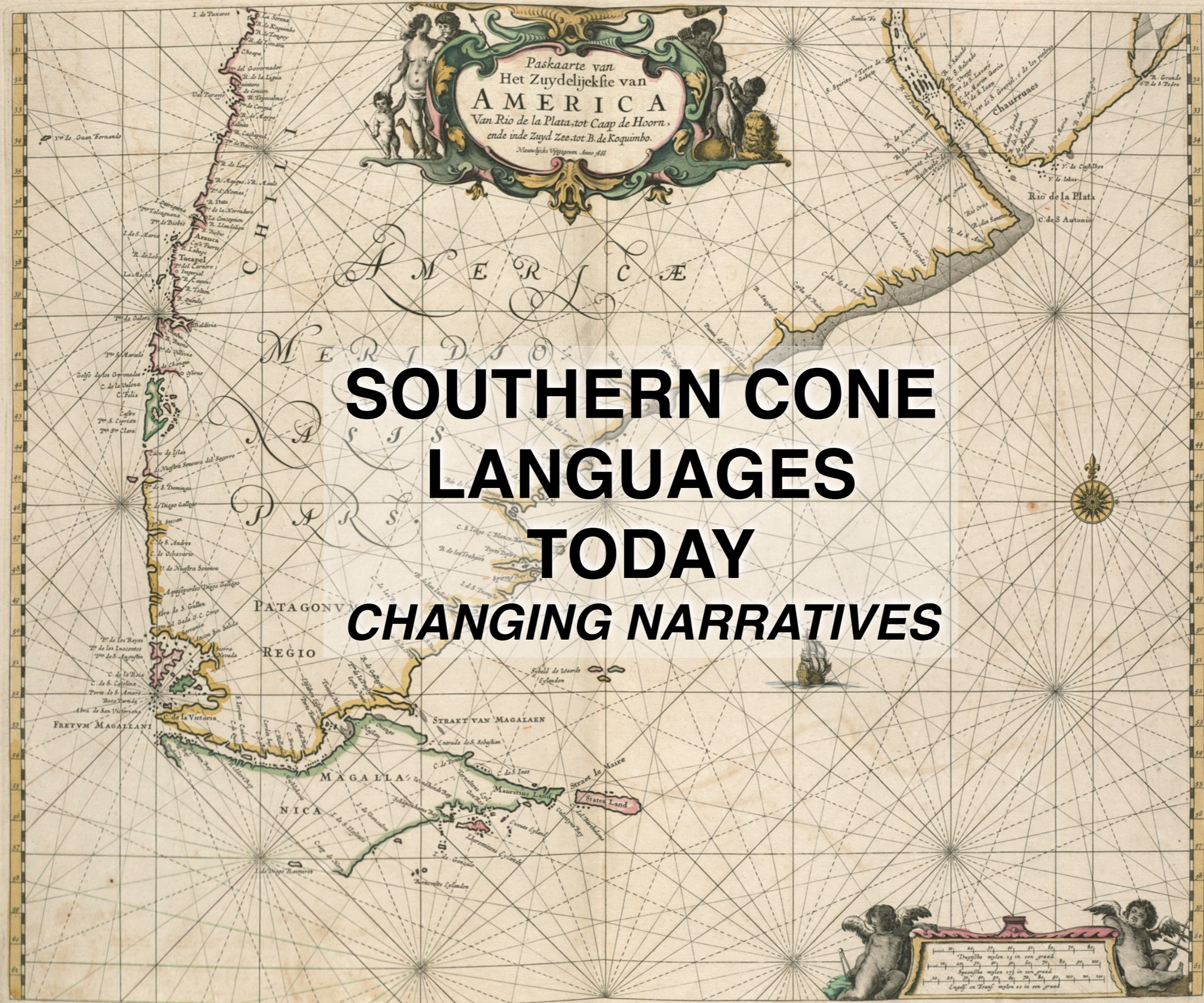
- Anglican and various Catholic orders establish 'missions'
- Infectious diseases ravaged these communities
- Indigenous peoples were forced to abandon traditional lifestyles and join the urban and rural proletariat



Selk'nam girls, with Salesian nuns



Mapuche family in Buenos Aires



**SOUTHERN CONE
LANGUAGES
TODAY
CHANGING NARRATIVES**

The story does not end here!



The territorial battles continue

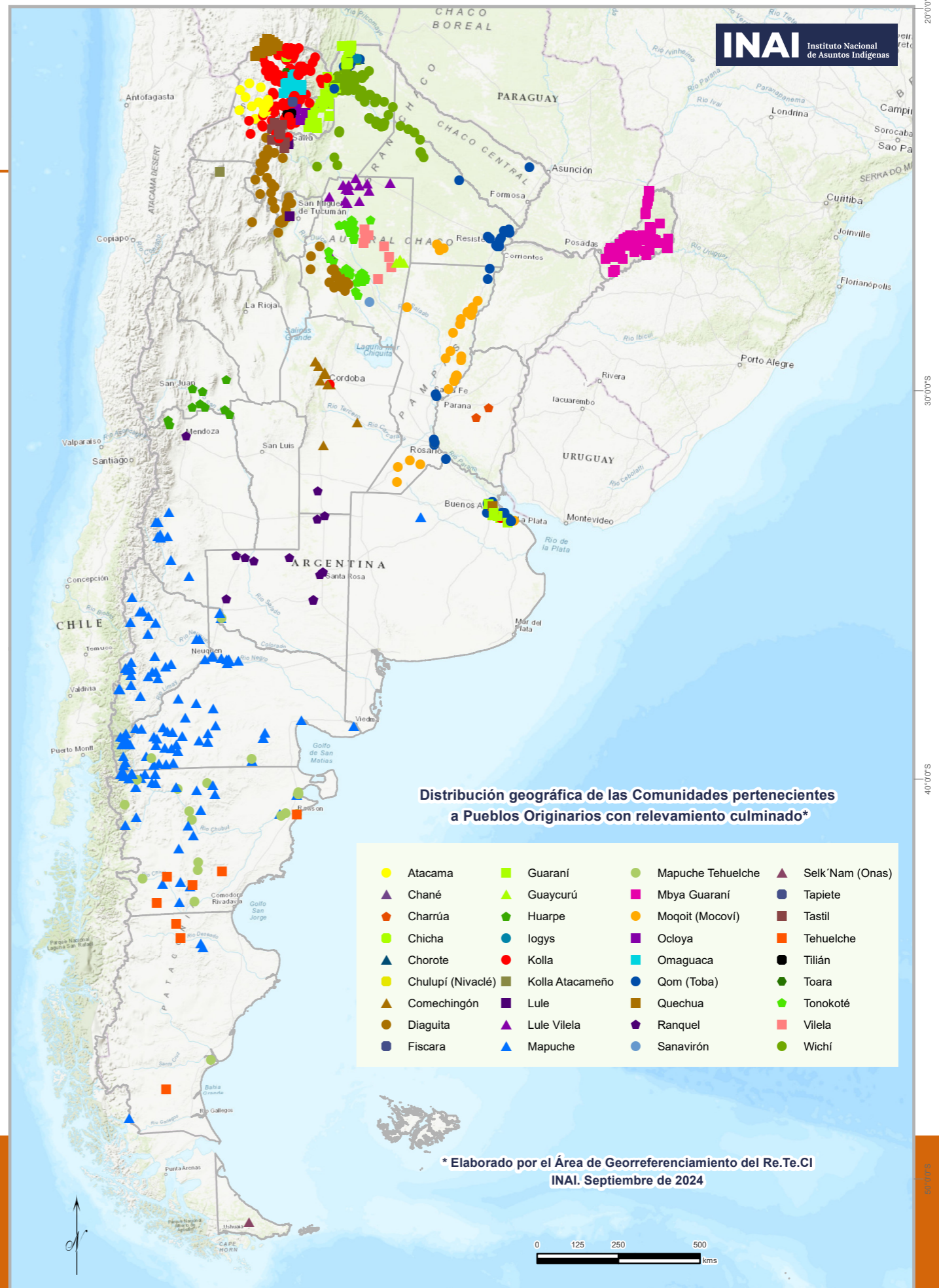
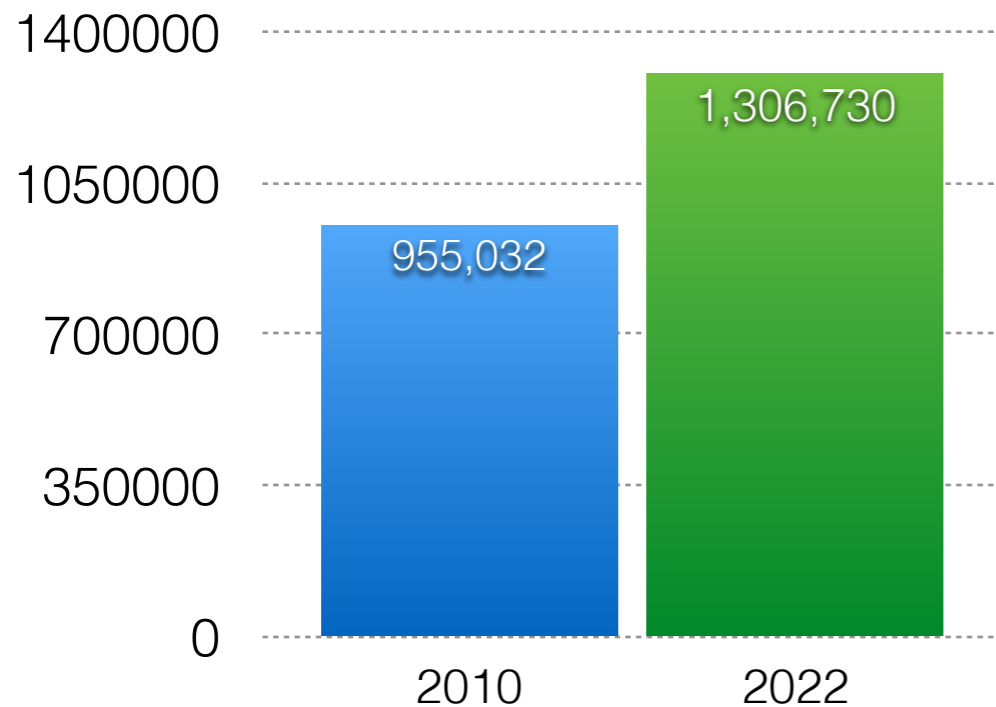


- Indigenous matters are a major political issue in Chile & Argentina
- Territorial claims have intensified and militarisation has increased

Indigenous identity

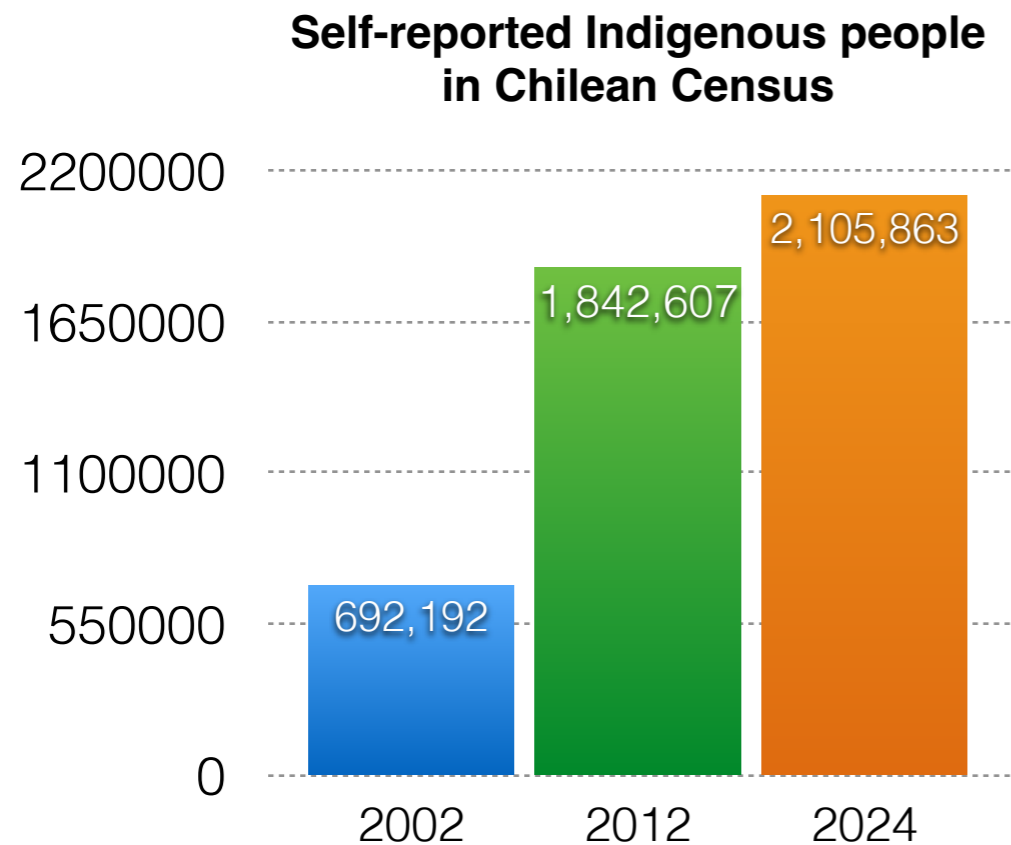
- Argentina:
- n° of legally recognised native settlements has grown
- self-reported indigenous group membership is up

Self-reported Indigenous people in Argentinian Census



Indigenous identities today

- **Chile:**
- Census numbers of self-identified indigenous Chileans almost tripled between the 2002 and 2012 censuses
- This reflects a change in the perception of indigenous identity

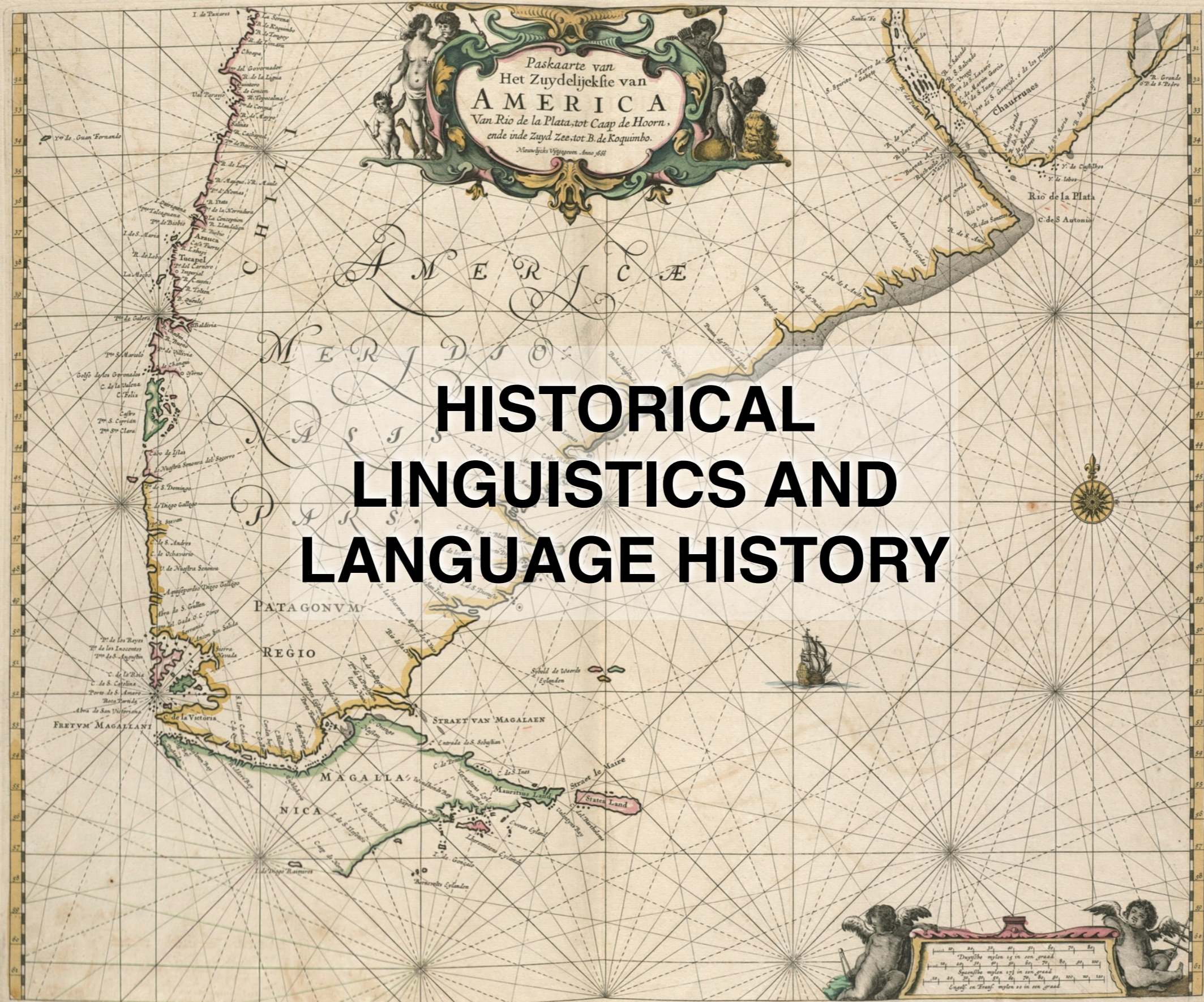


Language reclamation

- For most languages:
 - possibilities of obtaining new data are limited
 - we do have under-explored heritage materials, but...
 - they are difficult to access
 - of variable quality
 - difficult to interpret
 - extremely valuable for indigenous heritage and identity
 - both indigenous groups and academics have begun the task of bringing this material to light



3 Cetupan cuureajtaguda usipi, acan bqy aid ujud-
guda èca.



HISTORICAL LINGUISTICS AND LANGUAGE HISTORY



The story does not begin here!



Angus McIntosh Centre
for Historical Linguistics



THE UNIVERSITY of EDINBURGH

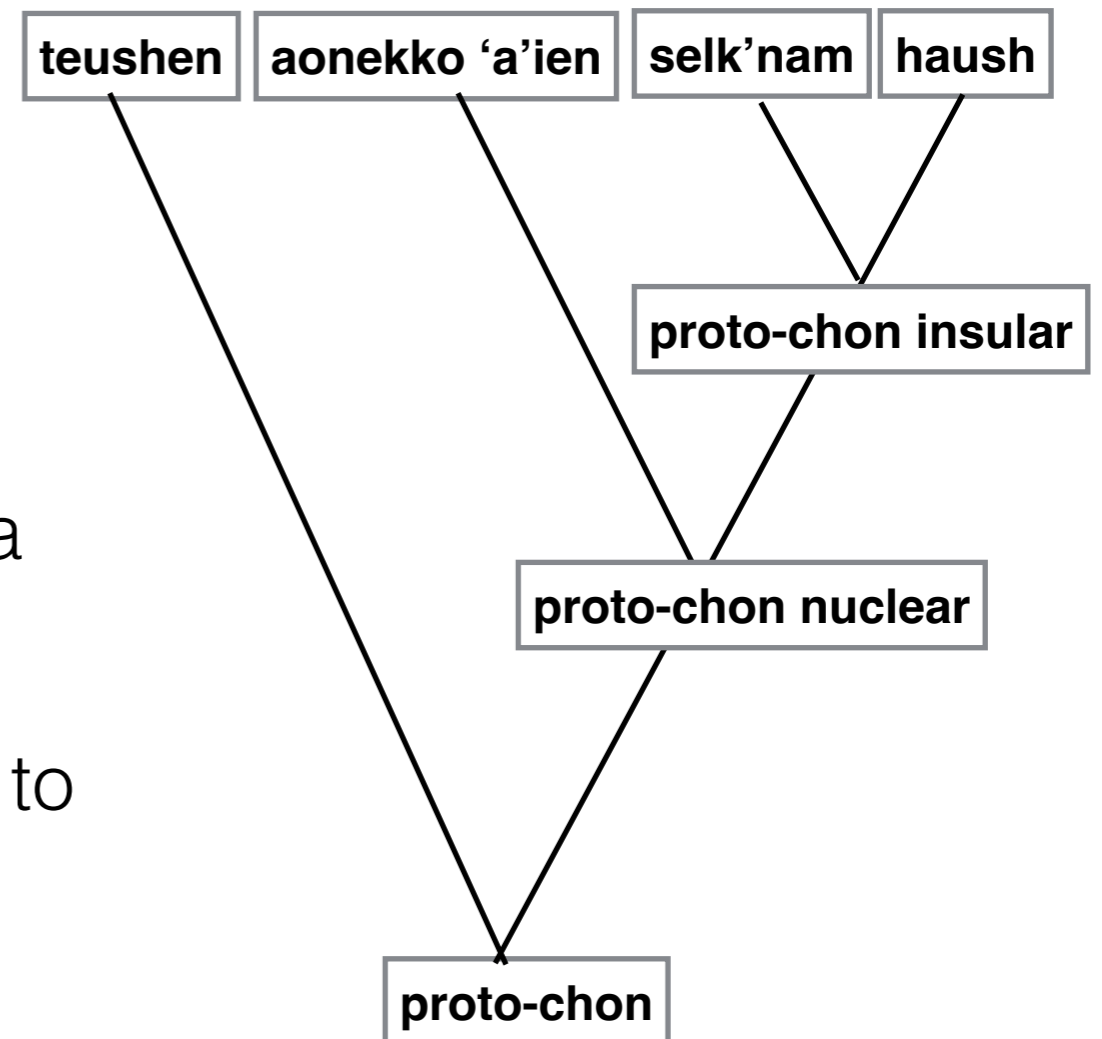
Quick, to the DeLorean!



Historical linguistics and human history

- **Comparative philology**

- Compares languages to establish genealogical relations
- Reconstruct **proto-languages**
- Requires a fairly robust amount of data
- Family relations are uncertain for now
- Most Southern Cone languages seem to be **isolates** or very **small families**



(Viegas Barros 2023)

Historical linguistics and human history

- **Comparative philology**

- Compares languages to establish genealogical relations
- Reconstruct **proto-languages**
- Requires a fairly robust amount of data
- Family relations are uncertain for new

We need more data!!!

- Maps & distribution of some languages seem to be **isolates** or very **small families**
- Furthermore, **no clear connections seem to exist with the larger language families of South America**



Historical linguistics and human history

○ Narrative Historical Linguistics

- Establish connections between language change, population dynamics and cultural change



Incaica expansion c. AD1470

Mapudungun	Quechua	Glosa
<i>apo/apu</i>	<i>apu</i>	'lord, chief'
<i>yanakon'a</i>	<i>yana</i>	'servant'
<i>chillka</i>	<i>qellqa</i>	'drawing, writing'
<i>mishki</i>	<i>misk'i</i>	'sweet money'

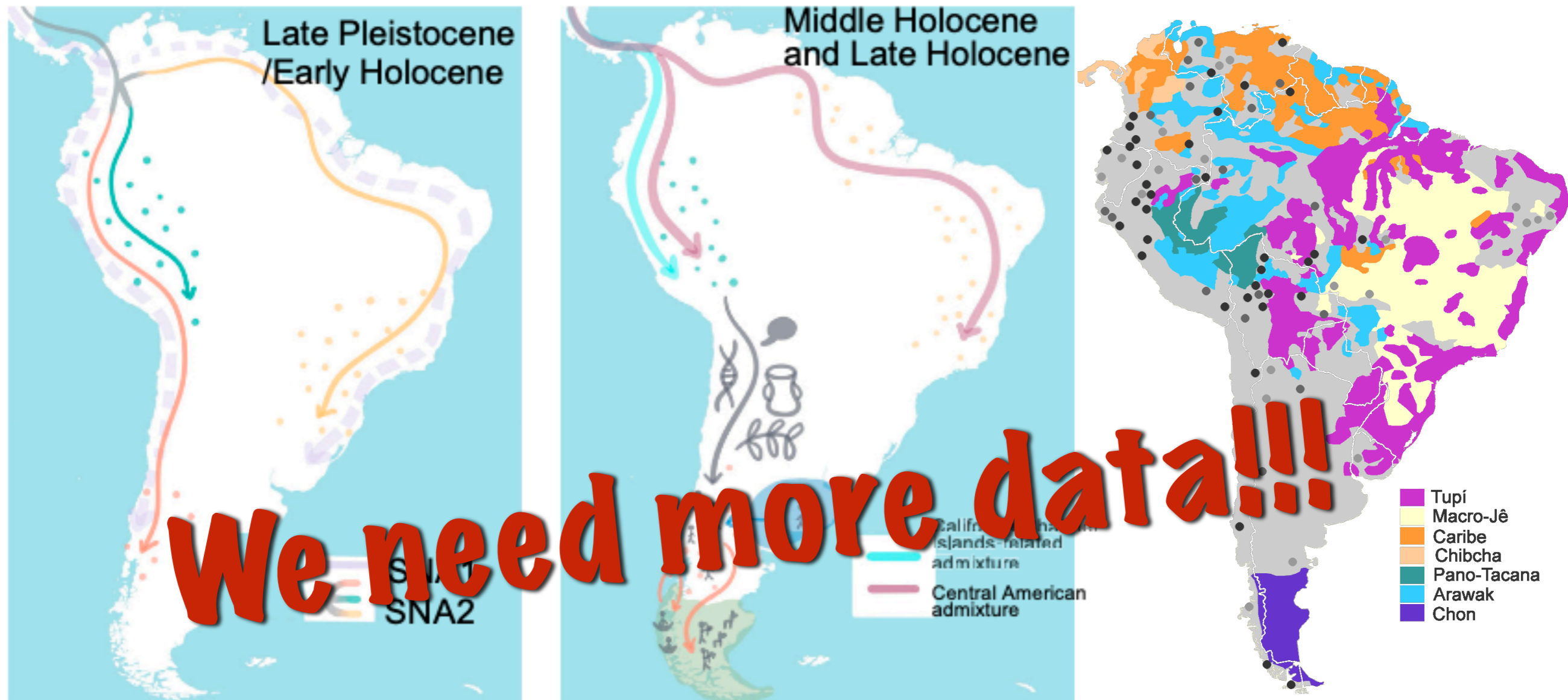
We need more data!!!

○ Mapudungun correlate with (Sánchez 2020)

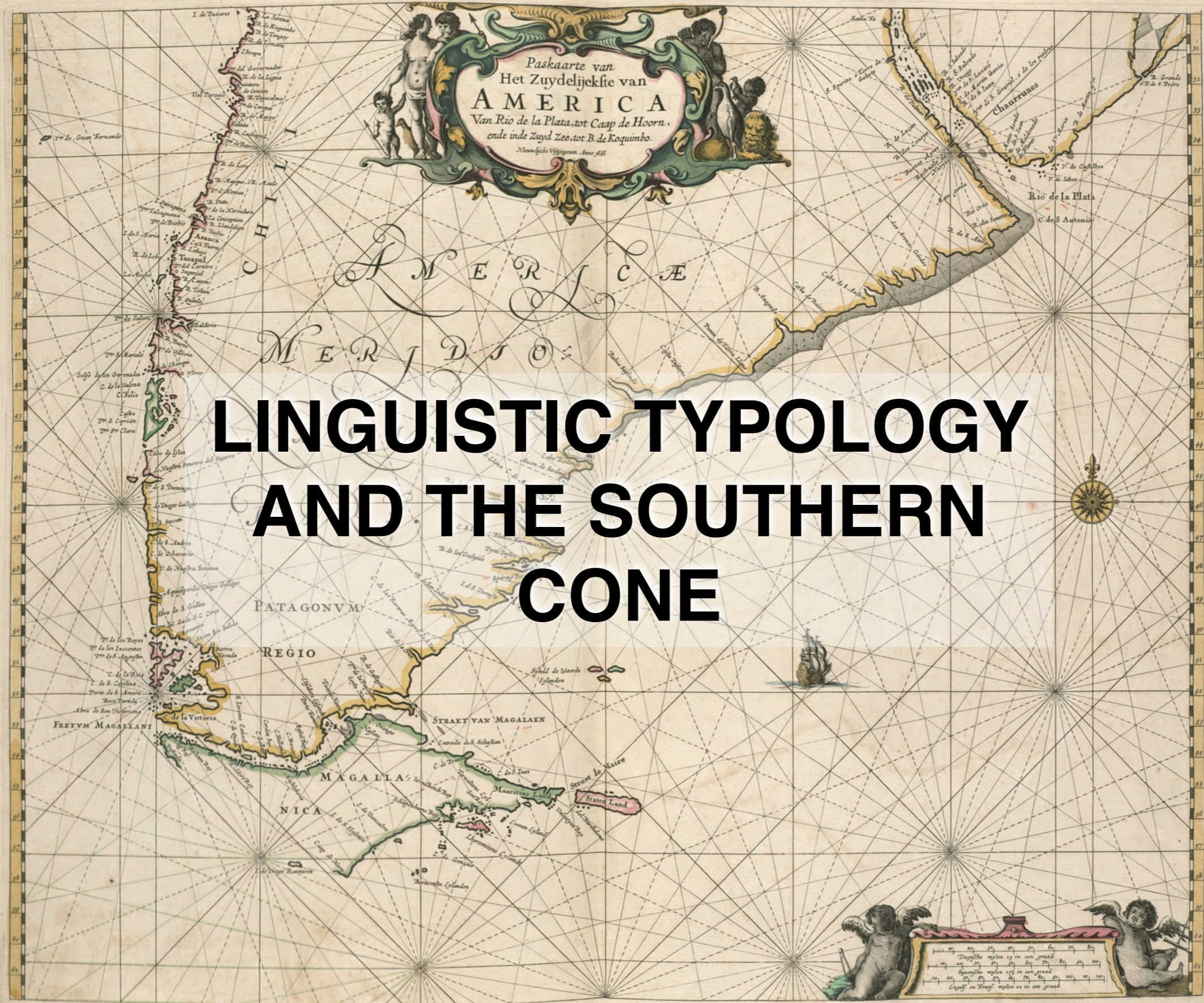
- Mapudungun correlate with
- Oral histories / shared customs
- Material culture (archeology)
- Genetic evidence

Historical linguistics and human history

- Genetic isolation of southern Andean populations



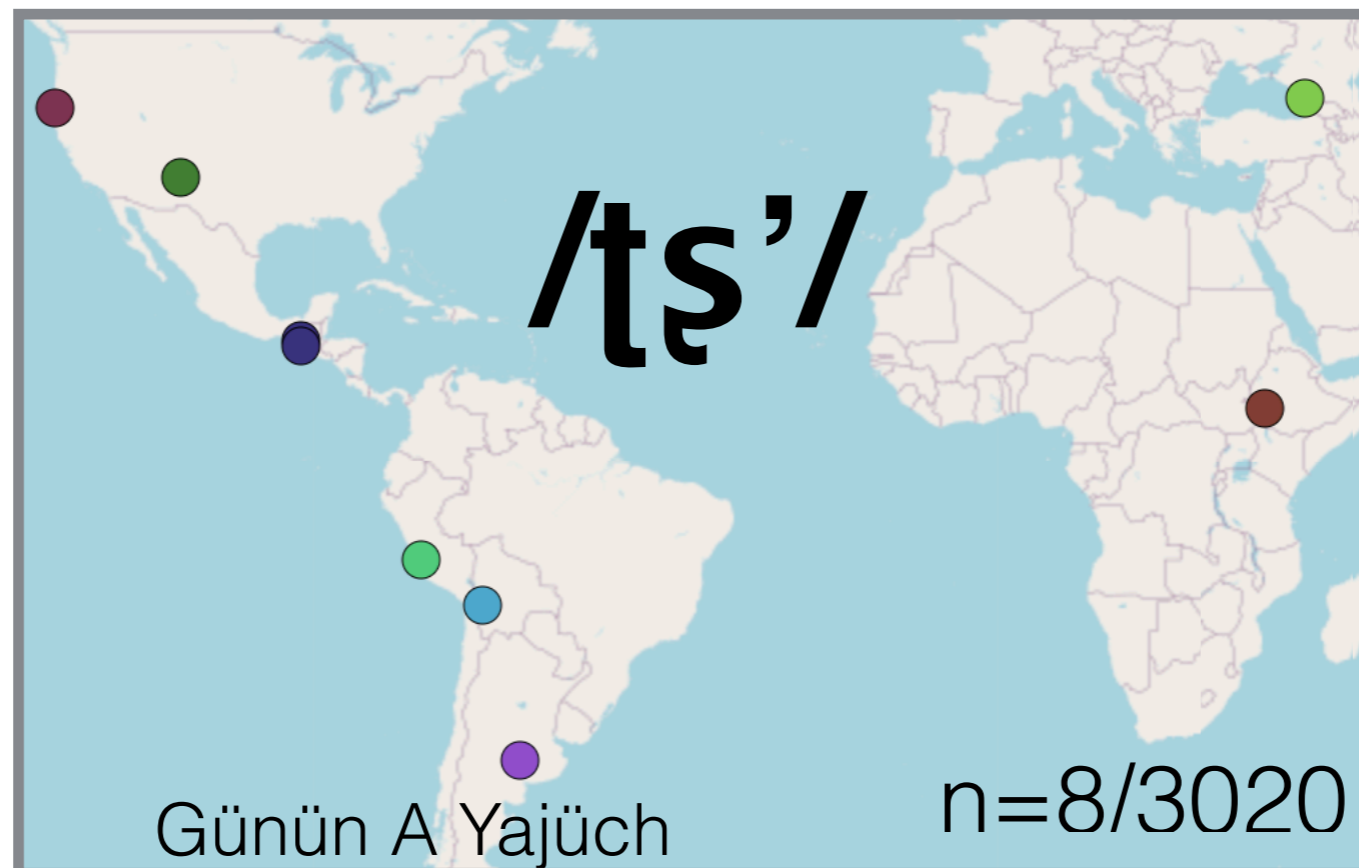
Arango-Isaza et al., 2023. The genetic history of the Southern Andes from present-day Mapuche ancestry, *Current Biology*.



LINGUISTIC TYPOLOGY AND THE SOUTHERN CONE

Linguistic typology

- Explores the distribution of linguistic features across the world, establishing **'what's where when?'** (Bickel 2007:248)
- Seeks to explore the diversity of languages in the world to establish its limits (discovering universal patterns in language)
- The threat to linguistic diversity worldwide makes this an urgent challenge









PHOIBLE
[**'fɔɪ.bʈ**]

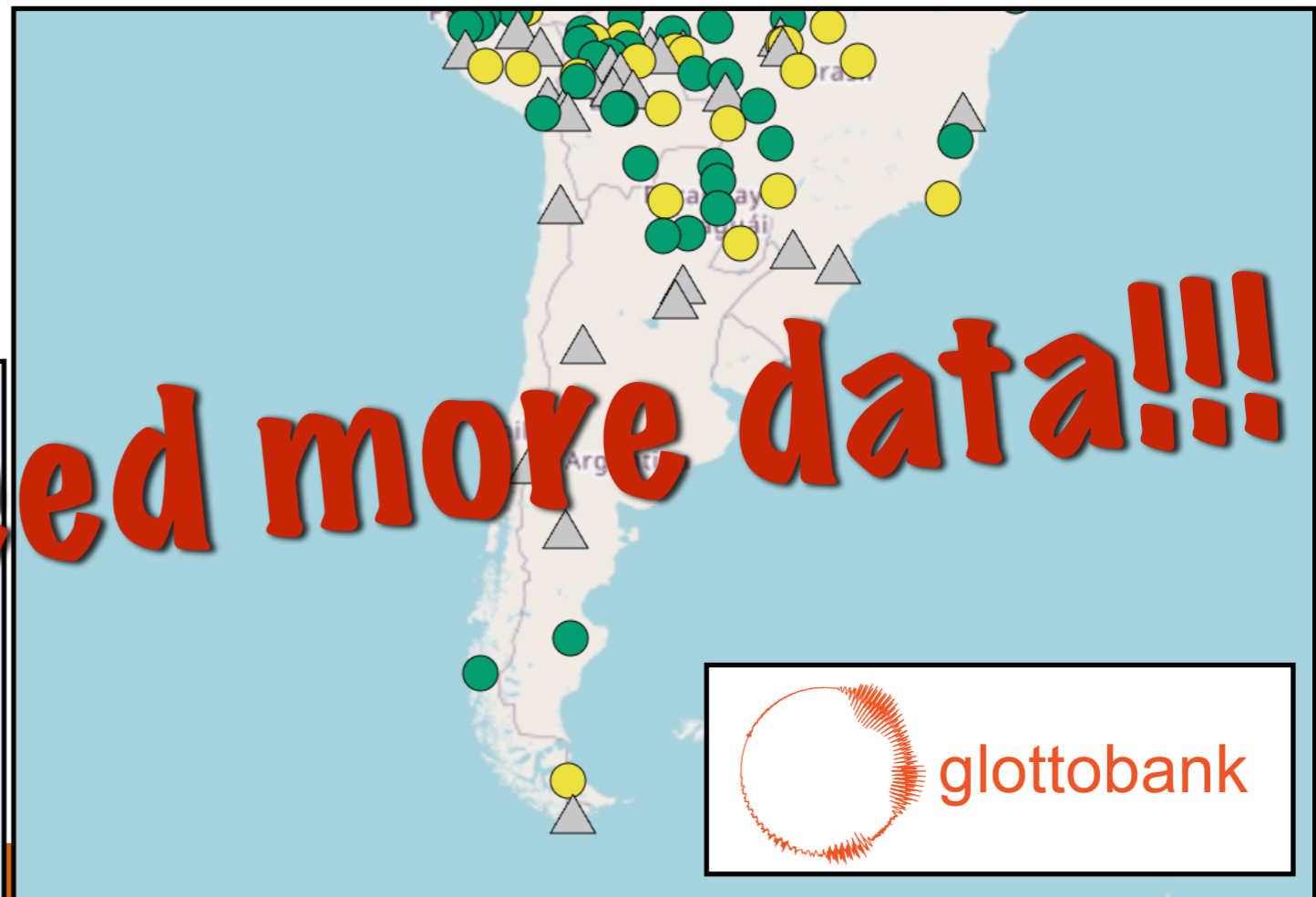
Linguistic typology

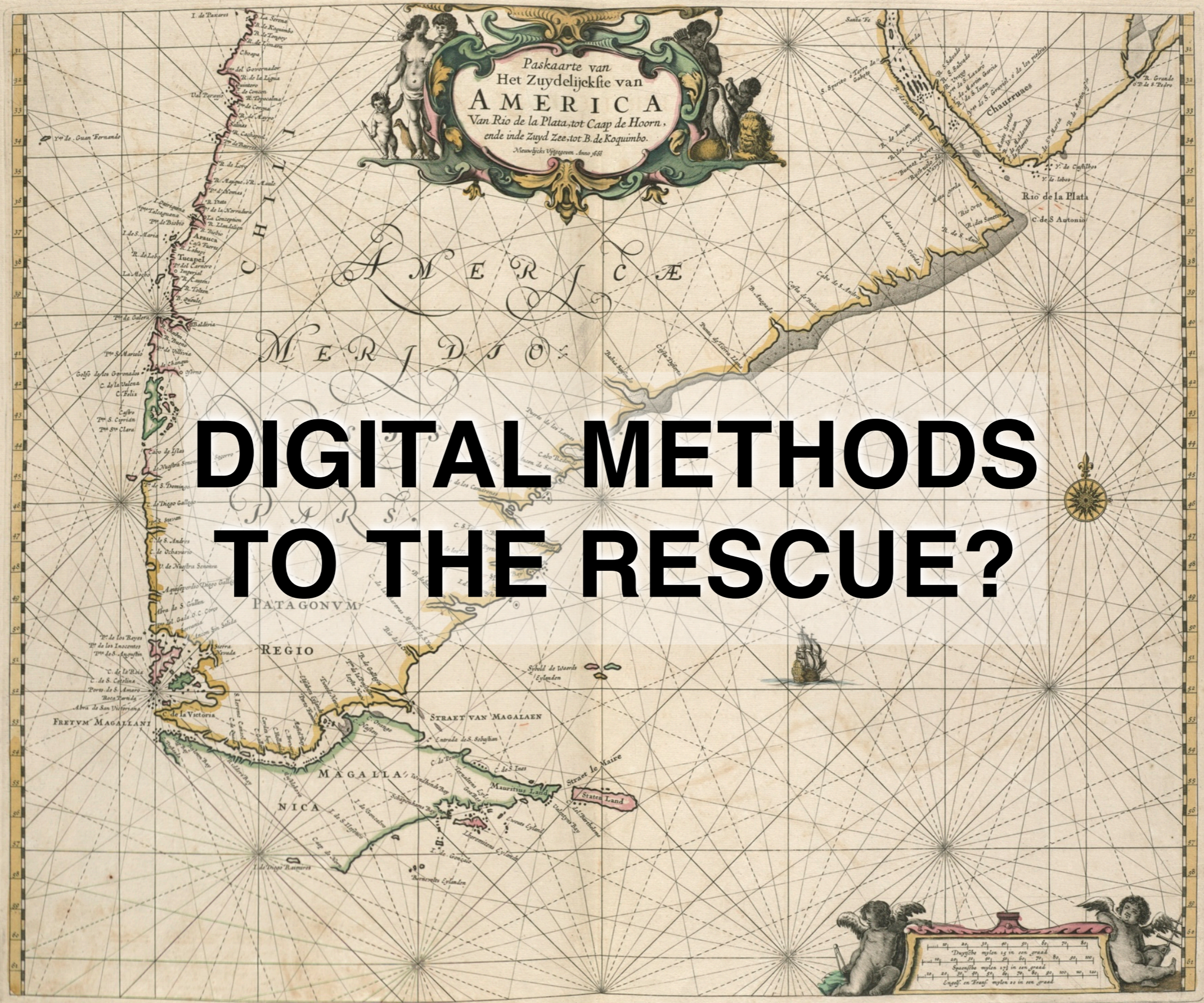
- Explores the distribution of linguistic features across the world, establishing **‘what’s where when?’** (Bickel 2007:248)
- Seeks to explore the diversity of languages in the world to establish its limits (discovering universal patterns in language)
- The threat to linguistic diversity worldwide makes this an urgent challenge

Feature GB188: Is there any productive augmentative marking on the noun

		0 absent	1087
		1 present	244
		? Not known	852

We need more data!!!

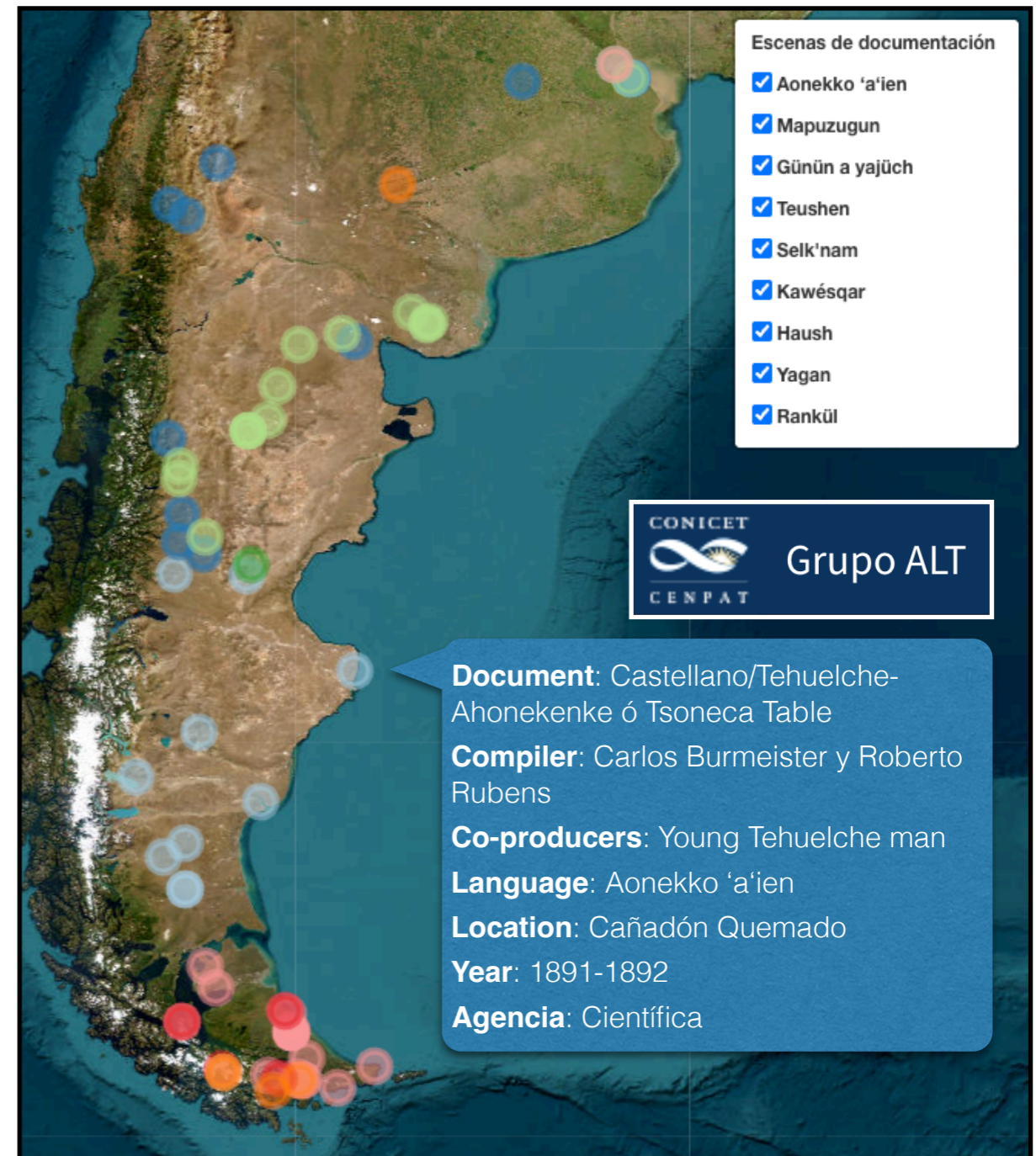




DIGITAL METHODS TO THE RESCUE?

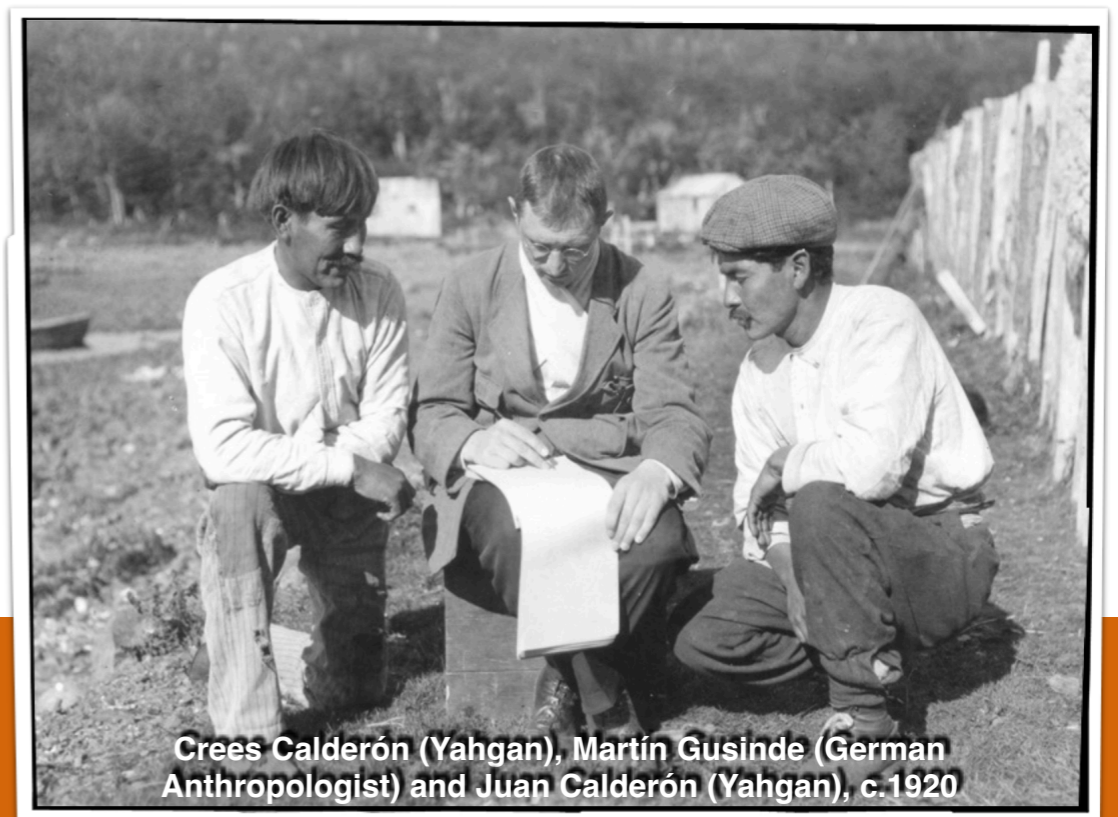
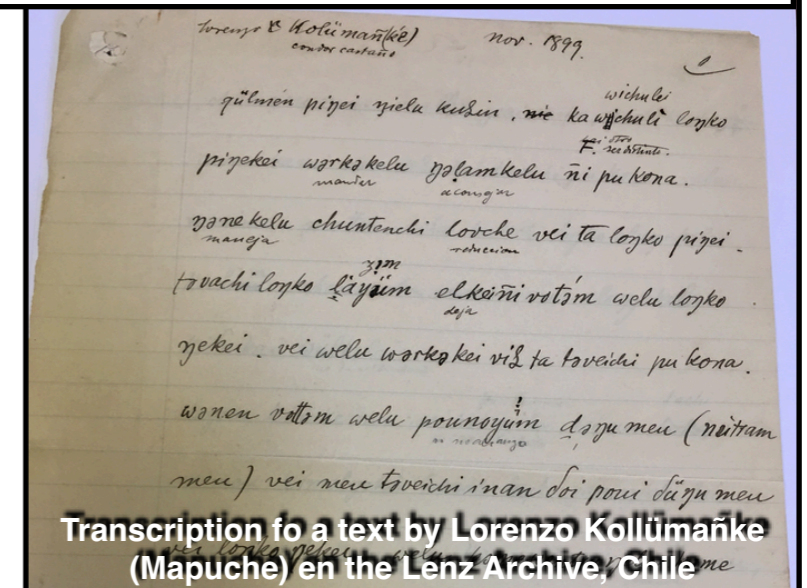
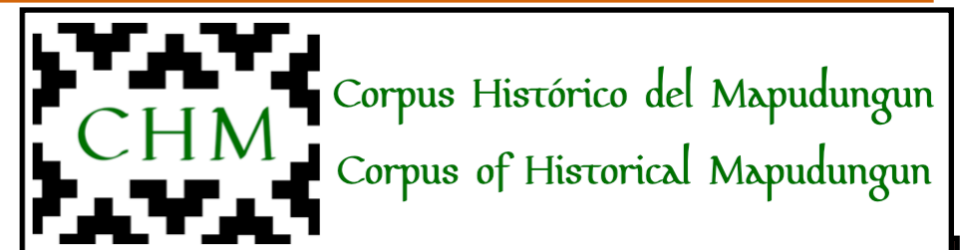
Problems that digital methods can mitigate

- Geographical dispersion of texts
- Some difficulties in interpreting materials
- The multilingual nature of the texts
- The asymmetry in descriptive frameworks
- Dialectal and proficiency differences of consultants / co-producers
- Orthographic/palaeographic challenges



First steps: A master-list of texts

- Identifying target texts
 - Already digitised/transcribed
 - Readily available in analogue form
 - Held in local/national archives or personal collections
- Obtaining copyright / ethical permissions
- Describing the materials
 - Languages/text types
 - Location/period of documentation
 - Compilers and co-producers
 - Context of documentation



Digitisation and OCR

- **Digitisation**

- Archival visits
- Full digitisation, where possible, otherwise, direct transcription
- Checking for additional materials 'lost' in archives

- **Optical Character Recognition**

- Longer texts to be transcribed using specialised software and developing bespoke models



29 scja acimus gesa satumuacua ácupi ácwan; sina
eloc sijnatamus cutupi sa bqv meàmœa sina cot
ácwan.

Print version of Gospel of St Luke in Yahgan (Bridges 1881)

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

MS of the Yahgan dictionary by Thomas Bridges (British Library)

Digitisation and OCR

The image shows a software interface for digitising and transcribing handwritten text. The top bar displays the file name 'Yahgan / Bridges 1881 Ya... / p016.jpg' and the status 'Ground Truth'. Below this, a second bar shows 'Yahgan / Yahgan Bridges ... / 0010.JPG' and 'In Progress'. The main area is divided into three sections: a thumbnail gallery on the left, a large image of the handwritten manuscript in the center, and a transcription panel on the right. The manuscript contains entries for various words in the Yahgan language, such as 'Alagunjna', 'Alagunzque', 'Alaguetulli', etc., with their corresponding meanings in English. The transcription panel on the right lists these words and their meanings in a numbered list, showing the output of an OCR process.

Yahgan / Bridges 1881 Ya... / p016.jpg Ground Truth Recognise 0 unsaved changes Save

Yahgan / Yahgan Bridges ... / 0010.JPG In Progress Recognise 0 unsaved changes Save

8 9 10 11 12

Alagunjna. To look about for search to look.
 Alagunzque. To look at for oneself.
 Alaguetulli. To look about, To chase with the eyes.
 Alageméini. To look at standing, sitting.
 Alaguinata. To us to get a sight of to sight.
 Mamuluk. = Cucyren. An edible shell fish.
 Alapuna. To fall into fall water.
 Alasggue. Any appearance of smoke To be uncertain. To be deceived.
 Alig. The too long feathers in herons heads.
 Alaguinata. To get a glimpse only. To recognise. To see indistinctly.
 Alonama. adj spotted with light spots said of seals skins.
 Alanguca. A stink, smell. The common steamer duck.
 Alucustal. The name of a star. A species of limpet. eatable
 Alona. a xv. intr. thinking, to shake. To blacken with charcoal. to scrub.
 Almeun. To scratch off. To take by scratching.
 Almeun. = Alca = fyun. To scratch bare to pieces.
 Almeun. To vomit. Vomit.

Region 1

- 1 To look about for search to look.
- 2 Alagunjna.
- 3 To look at for oneself.
- 4 Alagunzque.
- 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-Cugsin
- 13 To fall into fall water
- 14 Alapuna
- 15 Alasggue. An appearance of enke To be—:—: The idewein. To be 18à-

Results from a language model trained for Thomas Bridges' Yahgan alphabet (**handwriting version**)

Lematisation

see { *saw*
seeing
sean

- The goal is to provide a single tag (Lemma) for related words, eliminating grammatical and orthographic heterogeneity
 - A single orthography is adopted and a ‘dictionary form’ given
 - An English and a Spanish translation are provided
 - A part of speech tag (POS) is given to each lemma

MAPUDUNGUN		
Attested	Lemma	POS
< wera layay>	wedan ‘ser/estar mal/o/a’	V
< hueda lay>	wedan ‘ser/estar mal/o/a’	V



- Coded as XML, using TEI standards:

```
<w lemma="wedan" pos="V" corresp="be bad" sameAs="ser malo">huedalay</w>
```

Morphological tagging

- Allows us to identify and compare individual parts of words (roots and affixes)
- Each 'morpheme' is given a tag with its function or meaning
 - The process is expanded to words with similar structure

Orthography	Morphemes	Lemma	POS
<lag-ùm-a-bi-mi>	<i>morir</i> -CAUS-FUT-3DIR-IND.2.S	<i>l'angümün</i> 'kill'	V
<·La-uye-y>	<i>morir</i> -PERF-IND.3	<i>l'an</i> 'die'	V

```
<w xml:lang="arn" lemma="l'an" pos="V" corresp="die" sameAs="morir">  
  <m baseForm="l'a" type="root" corresp="die" sameAs="morir">·La</m>  
  <m baseForm="uye" type="sfx" subtype="inflectional" corresp="perf" sameAs="perf">uye</m>  
  <m baseForm="iy" type="sfx" subtype="inflectional" corresp="ind3" sameAs="ind3">y</m>  
</w>
```

Front-end search form

(Based on CHM interface)

×
Buscar en el Corpus Histórico del Mapudungun (CHM)
Español ▾

lema	<input type="text" value="weke"/>	ortografía original	<input type="text" value="e.g. duamimi"/>
categoría léxica	<input type="text" value="e.g. V"/>	período	<input type="text" value="e.g. 160f"/> – <input type="text" value="e.g. 193f"/>
traducción española	<input type="text" value="e.g. *know*"/>	locación	<input type="text" value="e.g. Panguipulli"/>
morfema	<input type="text" value="e.g. kura, imi"/>	título del texto	<input type="text" value="e.g. *sermon*"/>
tipo de morfema	<input type="text" value="e.g. root, sfx"/>	autor/compilador/transcriptor	<input type="text" value="e.g. *Lenz"/>
significado/función del morfema	<input type="text" value="e.g. stone, ind2s"/>	orador/narrador/informante	<input type="text" value="e.g. *wenuñamko"/>

Chillihueque

	<p>Ancatemú, * ca quiñe dugu tùculelavin gañi piuque meu gañi pu Ghùlmen ga , piquen ta tvá; * Eymn tamn themgen allcùMissapayaymn, Ancatemú, * roquiñyepayaymn cùme que glam, Dios ñi dugu máy; * tamn pu piñeñ huelu rezapayay, * rùcùñmamoquili ta tvá; * Marimari ca Ancatemú, * re hueque cam vemgeay gamn pu piñeñ? * quimDiosnoalu cam themaygn vè ? * inche quimueruam, pipereyu ta tva, Marimari ca Ancatemu. *</p>	<p>Otra cosa quiero meter en el corazon de mis Caciques, Ancatemu: que vosotros los grandes vandreis à oír Misa, y à llevar como de cocabì buenos consejos, èsto es, cosas de Dios, mas vuestros chiquillos vendrán à rezar, no me los mezquineis: se han de criar solo come bestias, sin conocer à Dios? yo os los enseñarè, te digo, Ancatemu.</p>
Ancatemu	Vemay veychi dugu ga, pien, Señor Payé, * re tami cortesiavin contupapeyu, Señor Payé: * venten tuquellayu ta tva, Marimari ca Señor Paye Millaleuvu.	Assì serà, Padre, lo que me dices: yo solo vine para saludarte, y assi no tengo mas que decirte.
Millaleuvu	Veyllechi, Ancatemú. *	Està bien, Ancatemu.
Ancatemu	Marimari Señor Paye. *	A Dios, Padre.
Millaleuvu	Marimari Ancatemú. *	A Dios, Ancatemu.

oveja/pequeño camélido

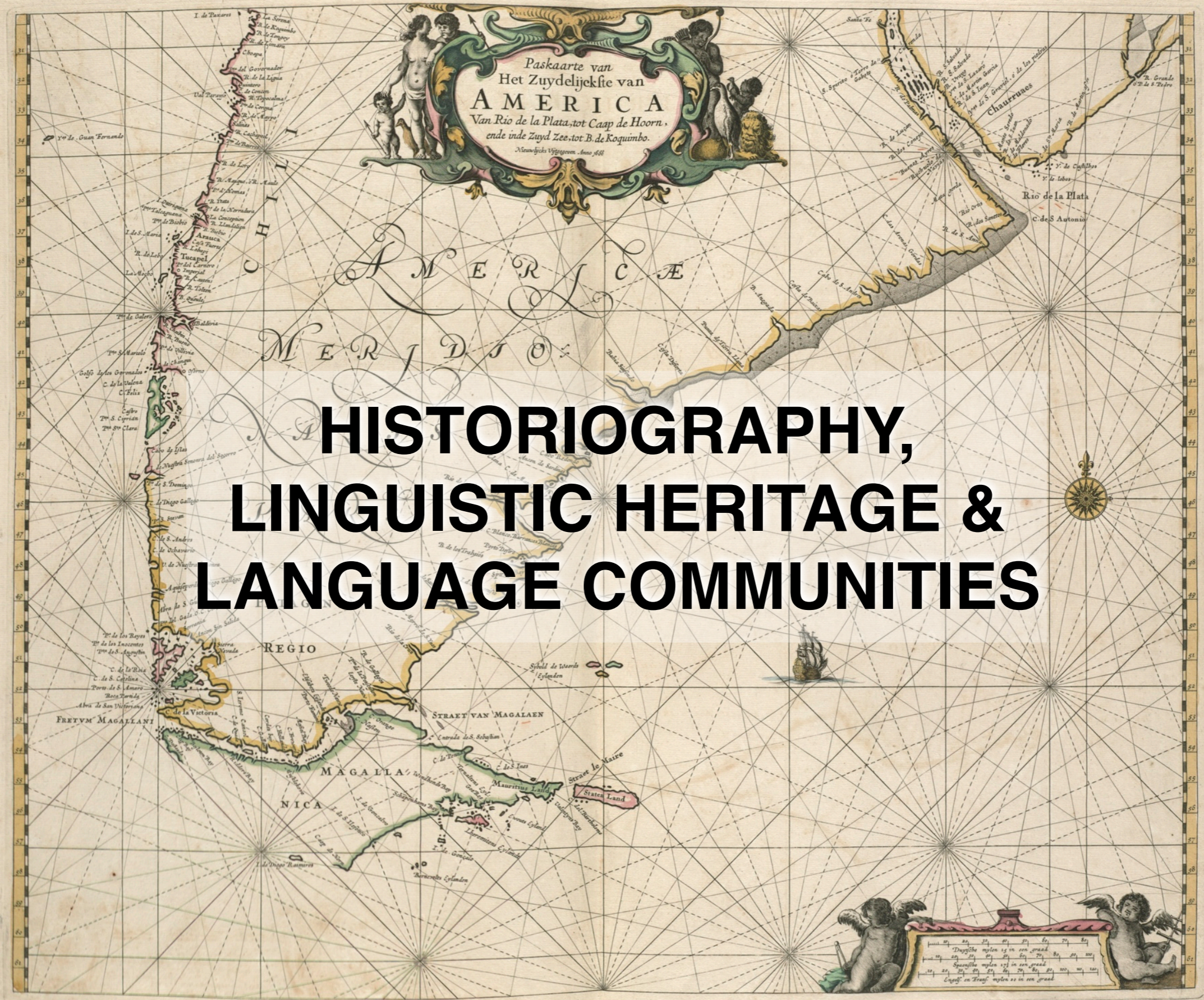
weke

weke/N	1765FEBRES-Coyaghtun	1
weke/N	1910LA-1-Costumbres	1
weke/N	1910LA-4-Canciones	1

Comparative search form

- Will allow searches across lemmas and morphemes in different languages:



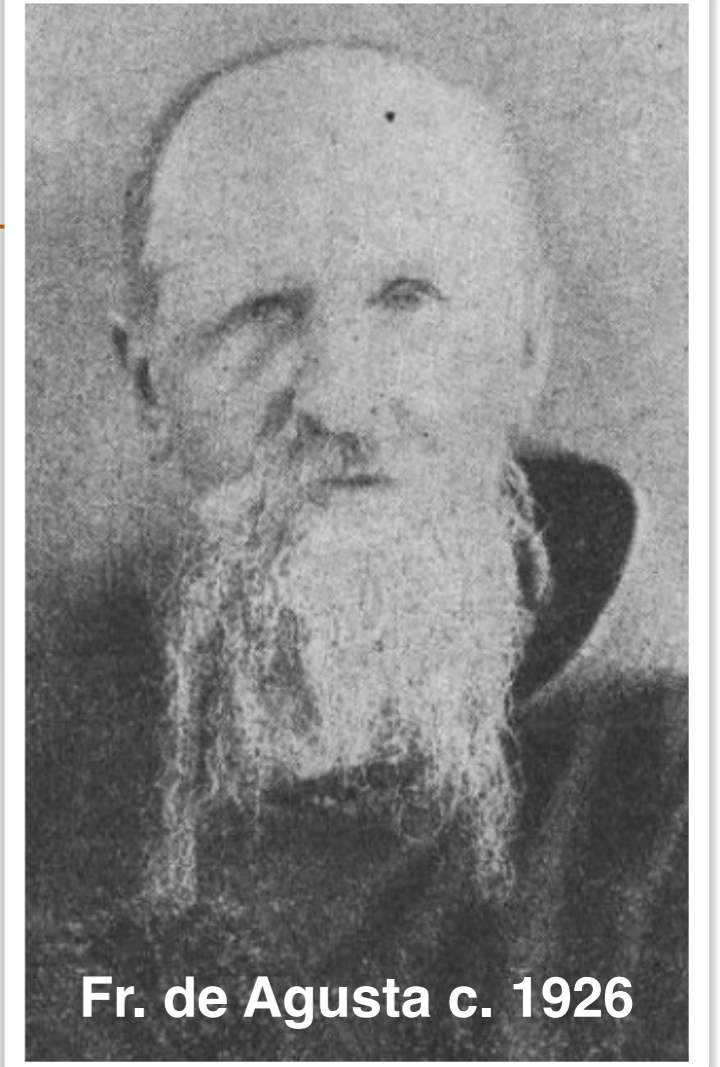


HISTORIOGRAPHY, LINGUISTIC HERITAGE & LANGUAGE COMMUNITIES

Supporting a narrative change

- Even in some of the more prolific documentarists, the ‘narrative of extinction’ is deeply entrenched.
 - Félix de Augusta (1910, p. vii):

La raza araucana pasa hoy por un período de transformación. Lo que nuestra obra refiere acerca de sus costumbres y supersticiones, dentro de poco ya no corresponderá á la realidad, ni se guardará memoria de ello; por eso hemos querido conservar estos detalles para la ciencia etnológica;



Fr. de Augusta c. 1926

‘The Araucanian race is going through a process of transformation today. What our work refers about their customs and superstitions, will shortly no longer correspond to reality, nor will there be memories of it; this is why we have wanted to preserve these details for the science of ethnography’

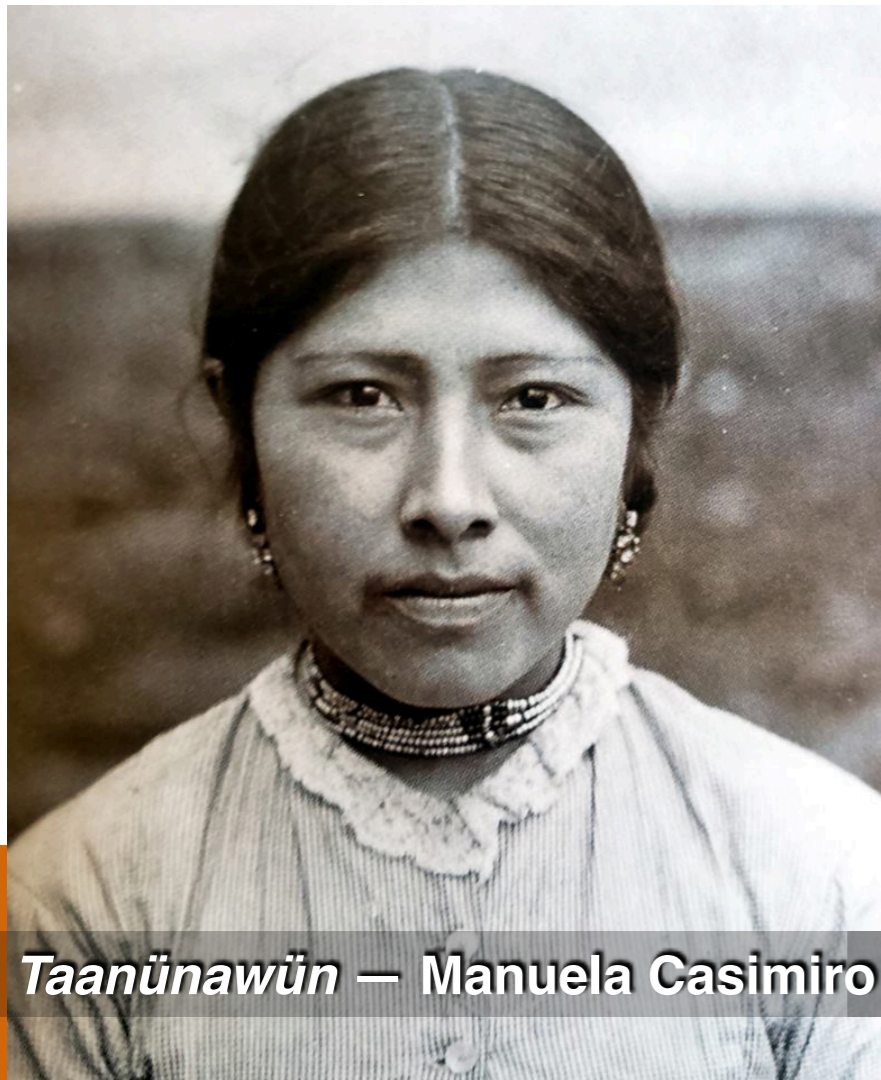
Supporting a narrative change

- Today, indigenous organisations and communities are eager to engage with many of these heritage materials



Historiography and community integration

- **Our project seeks to support these processes:**
 - Highlighting the agency of the speakers who were part of the process of documentation
 - Gathering testimonials from present-day communities in relation to the texts (oral memoirs, readings, etc.)



Taanünawün — Manuela Casimiro



Kalvün — Segundo Jara



Kimelfe Fresia Loncon Antileo

Historiography and community integration

- **Our project seeks to support these processes:**
 - Providing a non-academic interface to access texts and metadata
 - Producing reference/pedagogical materials, following community needs and team availability (see Mapudungun ‘advanced reader’)



Füta Amulel Chillkatuwe Mapudungun Mew

(Lecturas Avanzadas en Mapudungun 🦊 Advanced Mapudungun Reader)

👋 Mari mari! Iñchiw **Benjamín Molineaux** ka **Fresia Loncon Antileo**

Küme elkünufiyu tüfachi trokiñ epew, nütram, ka koyagtun.

Chilkatunge iñchu engu!

Kom tüfachi chillka nieyngün *notas* wingkadungun mew. Müley ka kiñe Trürn'emülwe.

Kom chillka Reñma mongen L'awen'tun Iyagel Palin Epew Kawiñ Ül



Kiñe chillka (Wenchukoy 1910)

Esta es una de las cartas contenidas en las *Lecturas Araucanas* compiladas por el padre capuchino Félix de



Lladkün (Nekul 1887)

Este poema, sobre el duelo de un hombre *williche*, pertenece a la colección de llamados *kollag*, escritos por Nekul o Juan Elías



Ñarki Machi (Naupa 1981)

Este texto pertenece a un epew relatado por el kimche pehuenche Lorenzo Naupa Epuñán, al investigador Gilberto

Conclusions

- The main deliverable for the project is the **CHiCo: Southern Cone**, however, we also plan to produce
- **Academic articles** on the typology, historical linguistics and historiography of the target languages
- **Contributions to typological databases** like Grambank, WALS and Phoible
- **Community materials** for the promotion of these languages (via our non-academic web presence)
- Collaborate with **young researchers and community members** in Chile, Argentina and the world at large, with a focus on Native American languages
- *Bring in more eyes, more voices to what we know about the languages of the world, and what we can say about who we are, who we've been and who we'd like to one day be*



Angus McIntosh Centre
for Historical Linguistics

CHiCo
C S

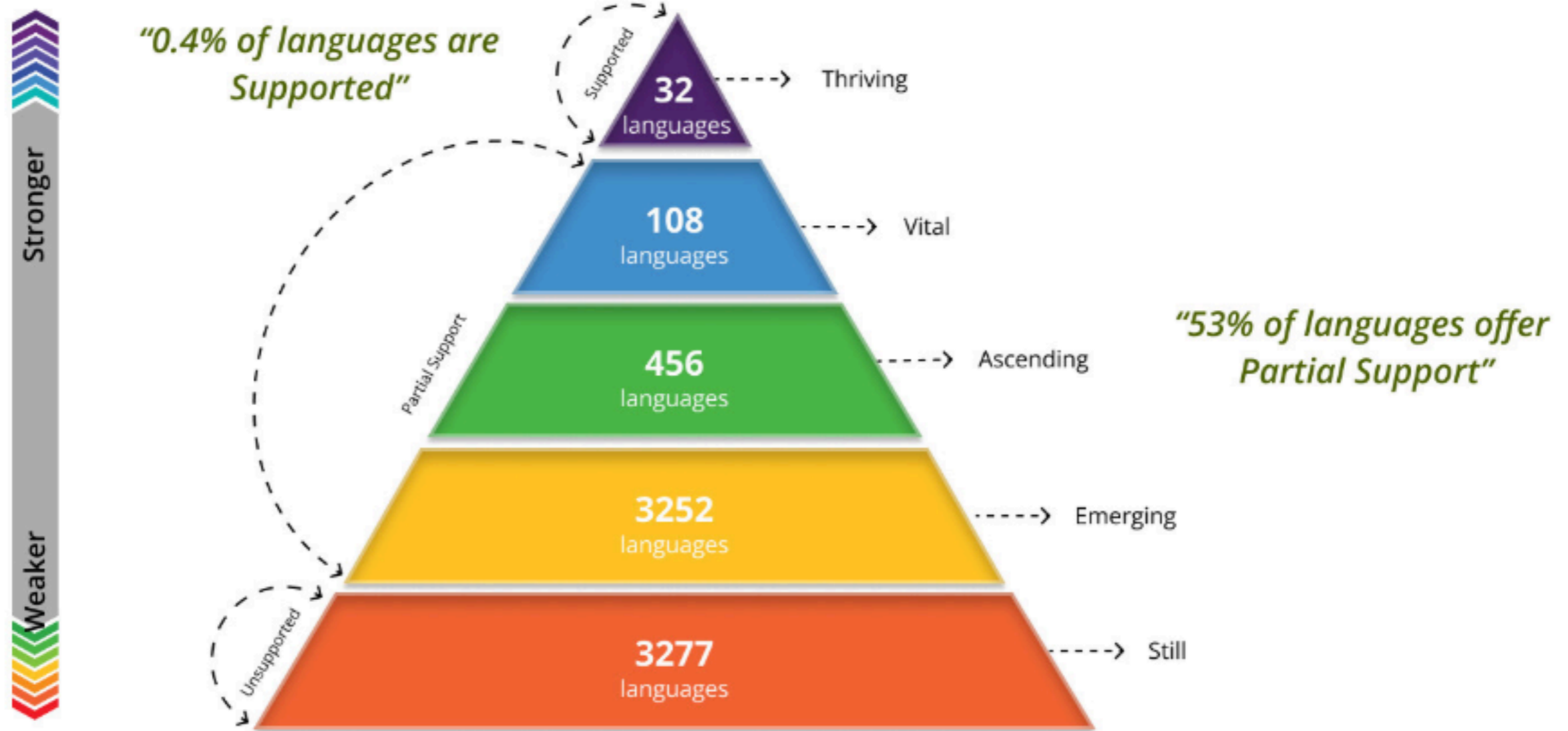

Thanks!



THE UNIVERSITY
of EDINBURGH

Analysis of languages within Digital Language Support categories

* 43 languages (0.6%) are Unclassified



"46% of languages are Unsupported"



Ethnologue