



Aspects in curating lexical resources for Enggano

Gede Primahadi Wijaya Rajeg

 <https://orcid.org/0000-0002-2047-8621>

Enggano Research Seminar at the Faculty of Humanities, Udayana University (18 March 2024)

Outlines

1. What are “lexical resources”?
2. What are the existing linguistic materials for Enggano?
3. How do we curate these materials into lexical resources?

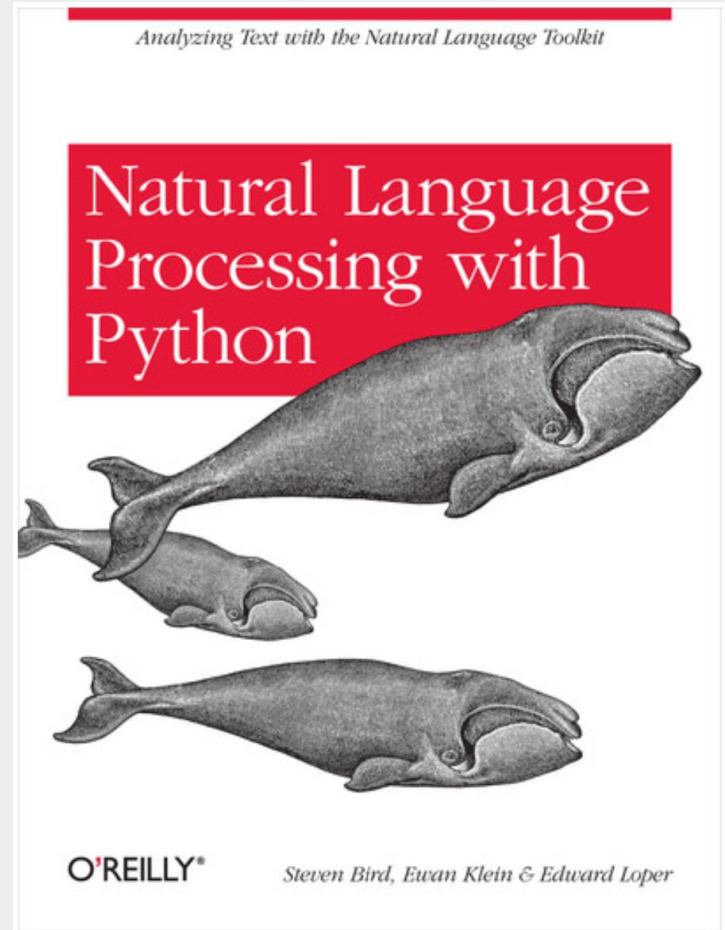
Outlines

1. What are “lexical resources”?
2. What are the existing linguistic materials for Enggano?
3. How do we curate these materials into lexical resources?

Lexicon/Lexical Resource

“a collection of words and/or phrases along with associated information, such as part-of-speech and sense definitions.”

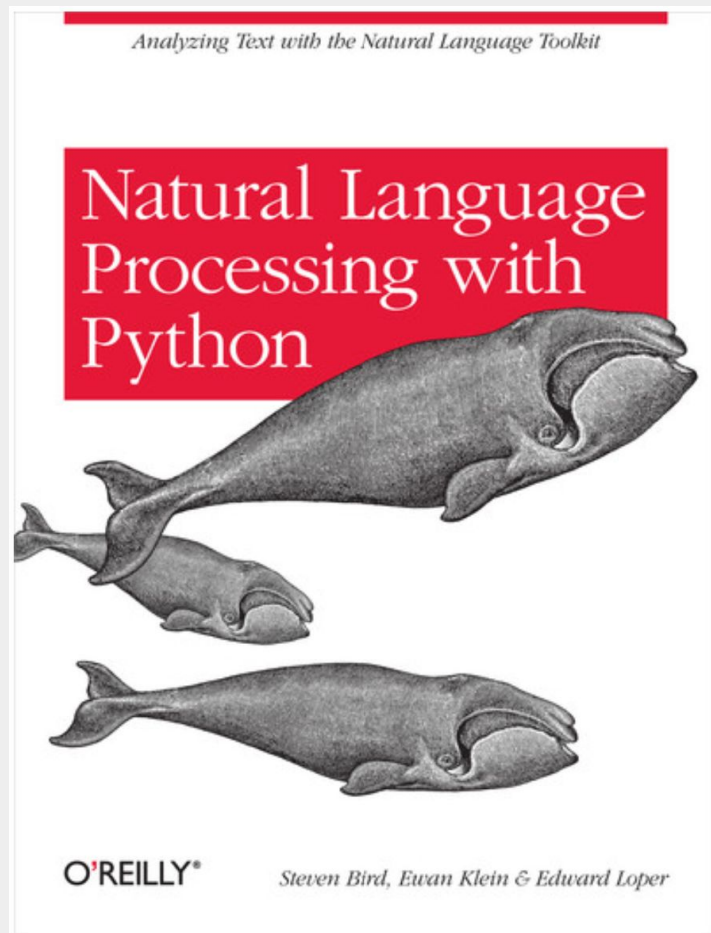
Bird et al. (2009: 59)



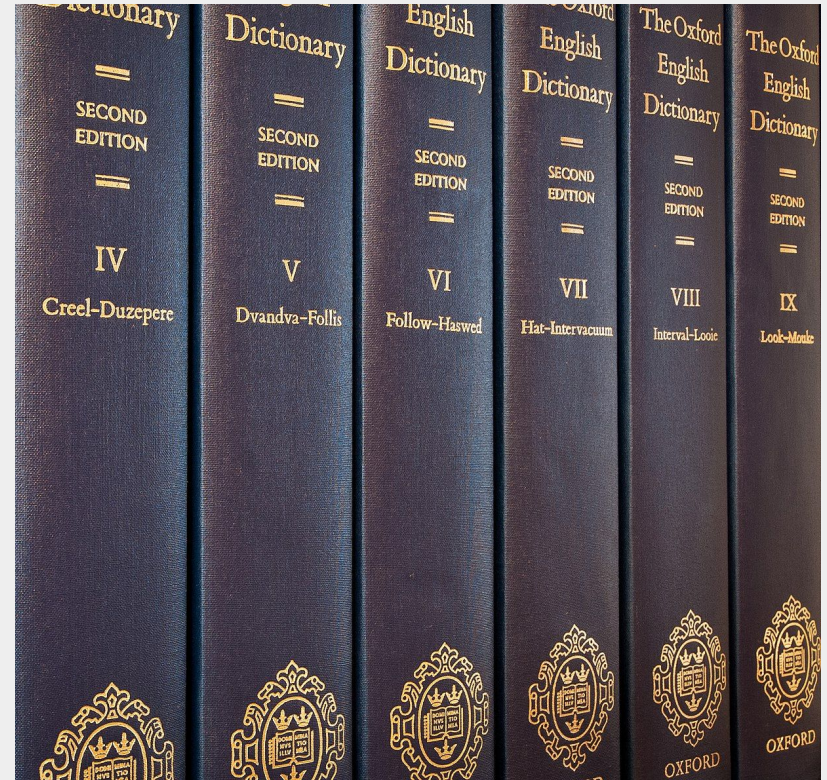
Lexicon/Lexical Resource

“a collection of words and/or phrases along with associated information, such as part-of-speech and sense definitions.”

Bird et al. (2009: 59)



dictionary...



word list

A simple lexical resource

(cf. Bird et al. 2009: 59–60)

In hunne taal tellen de Engano eilanders als volgt:

1 kahai,

2 adoea,

3 akoloe,

4 ampapa,

5 aliema,

6 akie akiena,

7 aliema adoea,

8 apa joepa,

10 apa apa adoea,

100 adoea taka tapoeloe,

200 aliema taka,

300 aliema adoea,

Boewang (1854: 391)

22 items

Boewang, Djoeragan. 1854. Verslag omtrent het Eiland Enggano. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 2. 379-393.

Austronesian Basic Vocabulary Database

Word: black

Entries for "black":

ID	Language	Item	Annotation	Cognacy	Classification
183499.	Yogad	ngisít		17	A:M:P:N:N:C:Ibanagic
187233.	Dupaningan Agta	mangiset		17	A:M:P:N:N:N:Northern
187234.	Dupaningan Agta	mangitet		17	A:M:P:N:N:N:Northern
251250.	Itawis	ngisít		17	A:M:P:N:N:C:Ibanagic
299298.	Ayta Mag-antsi	nangisit		17	A:M:P:C:Sambalic
299892.	Ayta Abellen	ɲitit		17	A:M:P:C:Sambalic
300158.	Ayta Mag-anchi	ɲisit		17	A:M:P:C:Sambalic
300298.	Bolinao	ɲisit	from Ilokano	17	A:M:P:C:Sambalic
300438.	Botolan Sambal	ɲitit		17	A:M:P:C:Sambalic
300718.	Sambal	ɲitit		17	A:M:P:C:Sambalic
301934.	Malaweg	ɲísít		17	A:M:P:N:N:C:Ibanagic
52211.	Mekeo	umu		18	A:M:C:E:O:W:P:P:C:W:Nuclear
52342.	Roro	umu [ʔúmu]		18	A:M:C:E:O:W:P:P:C:W:Nuclear
106582.	Lala	umauma		18	A:M:C:E:O:W:P:P:C:W:Nuclear
328061.	Mekeo, East	umu		18	A:M:C:E:O:W:P:P:C:W:Nuclear
329989.	Lala	umauma		18	A:M:C:E:O:W:P:P:C:W:Nuclear
331461.	Mekeo, North-West	umu		18	A:M:C:E:O:W:P:P:C:W:Nuclear
332307.	Roro	umu		18	A:M:C:E:O:W:P:P:C:W:Nuclear
52303.	Punan Kelai	maèdang		19	A:M:N:N:K:K:Modang
52354.	Modang	mældung		19	A:M:N:N:K:K:Modang
52279.	Itbayaten	mavaing		20	A:M:P:B:Ivatan
121485.	Iraralay	vaheng	black	20	A:M:P:B:Ivatan
121892.	Itbayat	vaeng	Note: vaeng/mavaeng and vaweng/mavaweng is a dialectal difference	20	A:M:P:B:Ivatan
121893.	Itbayat	mavaeng	Note: vaeng/mavaeng and vaweng/mavaweng is a dialectal	20	A:M:P:B:Ivatan

Comparative
electronic
lexical database
with cognates
grouping

Greenhill, Simon J., Robert Blust & Russell D. Gray.
2008. The Austronesian Basic Vocabulary
Database: From Bioinformatics to Lexomics.
Evolutionary Bioinformatics. SAGE Publications
Ltd STM 4. EBO.S893.
<https://doi.org/10.4137/EBO.S893>.

Comparative electronic lexical database with cognates grouping (for [some] words meaning 'black' in Austronesian lang.)

299892.	Ayta Abellen	ɲitit		17	A:M:P:C:Sambalic
300158.	Ayta Mag-anchi	ɲitsit		17	A:M:P:C:Sambalic
300298.	Bolinao	ɲisit	from Ilokano	17	A:M:P:C:Sambalic
300438.	Botolan Sambal	ɲitit	Cognate set (ID 17)	17	A:M:P:C:Sambalic
300718.	Sambal	ɲitit		17	A:M:P:C:Sambalic
301934.	Malaweg	ɲísit		17	A:M:P:N:N:C:Ibanagic
52211.	Mekeo	umu		18	A:M:C:E:O:W:P:P:C:W:Nuclear
52342.	Roro	umu [ʔúmu]		18	A:M:C:E:O:W:P:P:C:W:Nuclear
106582.	Lala	umauma		18	A:M:C:E:O:W:P:P:C:W:Nuclear
328061.	Mekeo, East	umu	Cognate set (ID 18)	18	A:M:C:E:O:W:P:P:C:W:Nuclear
329989.	Lala	umauma		18	A:M:C:E:O:W:P:P:C:W:Nuclear
331461.	Mekeo, North-West	umu		18	A:M:C:E:O:W:P:P:C:W:Nuclear
332307.	Roro	umu		18	A:M:C:E:O:W:P:P:C:W:Nuclear
52303.	Punan Kelai	maèdang		19	A:M:N:N:K:K:Modang
52354.	Modang	mældung		19	A:M:N:N:K:K:Modang
52279.	Itbayaten	mavaiŋ		20	A:M:P:B:Ivatan
121485.	Iraralay	vaheng	black	20	A:M:P:B:Ivatan

Comparative electronic lexical database with cognates grouping
(for [some] words meaning 'black' in Austronesian lang.)

Ayta Abellen	ŋitit	17	A:M:P:C:Sambalic
Ayta Mag-anchi	ŋitsit	17	A:M:P:C:Sambalic
Bolinao	ŋisit	17	A:M:P:C:Sambalic
Botolan Sambal	ŋitit	17	A:M:P:C:Sambalic
Sambal	ŋitit	17	A:M:P:C:Sambalic
Malaweg	ŋísít	17	A:M:P:N:N:C:Ibanagic

Cognate set (ID 17)

Comparative electronic lexical database with cognates grouping (for [some] words meaning 'black' in Austronesian lang.)

<u>Mekeo</u>	umu	18	A:M:C:E:O:W:P:P:C:W:Nuclear
<u>Roro</u>	umu [ʔúmu]	18	A:M:C:E:O:W:P:P:C:W:Nuclear
<u>Lala</u>	umauma	18	A:M:C:E:O:W:P:P:C:W:Nuclear
<u>Mekeo, East</u>	umu	18	A:M:C:E:O:W:P:P:C:W:Nuclear
<u>Lala</u>	umauma	18	A:M:C:E:O:W:P:P:C:W:Nuclear
<u>Mekeo, North-West</u>	umu	18	A:M:C:E:O:W:P:P:C:W:Nuclear
<u>Roro</u>	umu	18	A:M:C:E:O:W:P:P:C:W:Nuclear
<u>Punan Kelai</u>	maèdang	19	A:M:N:N:K:K:Modang
<u>Modang</u>	mældung	19	A:M:N:N:K:K:Modang
<u>Itbayaten</u>	mavaiŋ	20	A:M:P:B:Ivatan
<u>Iraralay</u>	vaheng	20	A:M:P:B:Ivatan
<u>Itbayat</u>	vaeng	20	A:M:P:B:Ivatan

Comparative electronic lexical database with cognates grouping & alignment of sound segments

Form	Language	Concept	Alignment
akani	Abui, Fuimelang	black	- a k a n i -
a.ka.'ni	Abui, Petleng	black	- a k a n i -
ä.'qän	Abui, Takalelang	black	- a q a n - -
akani	Abui, Atimelang	black	- a k a n i -
akan	Kafoa	black	- a k a n - -
aqana	proto-AP	black	- a q a n a -
ani	Reta, Ternate	black	- a - - n i -
akeŋ	Wersing, Taramana	black	- a k e ŋ - -
kaʔana	Blagar, Manatang	black	k a ʔ a n a -

Kaiping, Gereon A. & Marian Klamer. 2018. LexiRumah: An online lexical database of the Lesser Sunda Islands. *PLOS ONE*. Public Library of Science 13(10). e0205250.

<https://doi.org/10.1371/journal.pone.0205250>.

Other well-known lexical resources

- **WordNet** (lexical database for English for synonym groups)
- **FrameNet** (lexical database for English implementing the theories of Frame Semantics and Construction Grammar)
- **Lexibank** (large-scale cross-linguistic lexical data)

Goals of the *Lexical
Resources* research
grant for Enggano

#1

Lexical database
from different time
slices of Enggano


#2

Digital contemporary
Enggano dictionary
(online and in mobile-app)

#3

Learner's dictionary
(coordinated with the textbooks)

Outlines

1. What are “lexical resources”? 
2. What are the existing linguistic materials for Enggano?
3. How do we curate these materials into lexical resources?

- **word lists** (from mid-19th century [*cf. Krauße, under review, for the detailed discussion of these word lists*])
- **texts**
- **grammar sketches**
- **dictionary**

- **word lists** (from mid-19th century [*cf. Krauße, under review, for the detailed discussion of these word lists*])
- texts
- grammar sketches
- **dictionary**

Brouwer (c. early 1850)

(Thanks to Daniel Krauß for deciphering this cryptic yet valuable hand-writing)

Woordenlijst van het weinige
door de Inlanders op Engano, te Barhao
ons medegedeelde der Engaanesche taal
Siedelste door den Chinese Tji-bong
van den Luikje D. J. Brouwer

Soewel mogelijk door de Soera, de juiste uitpraak
aangegeven.

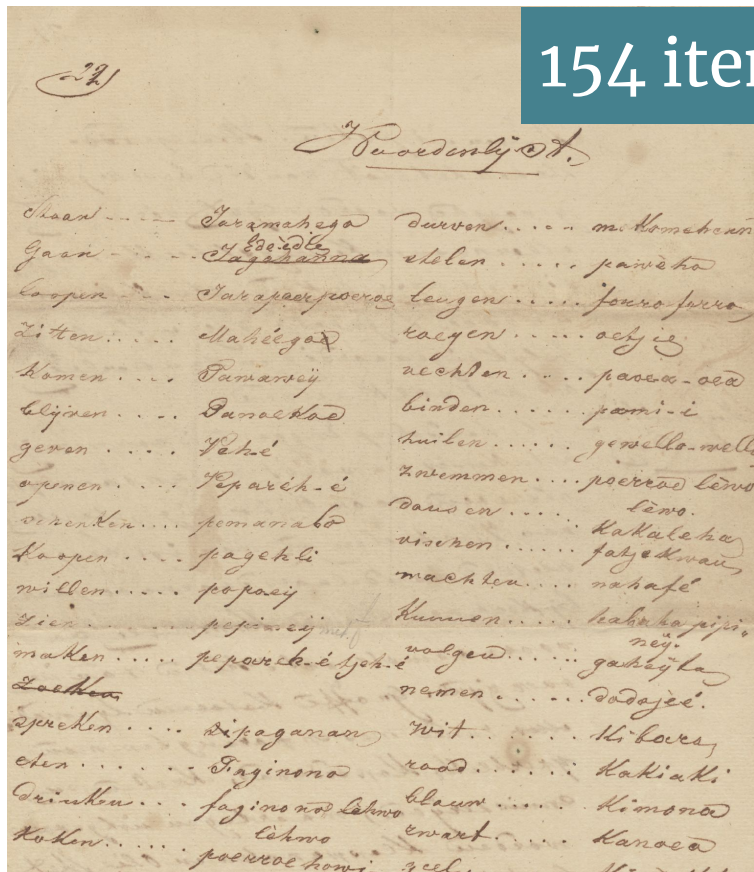
Weg	Tob	Luiven	Lechoe
Maan	Mona	Papagaa	Kata
nacht	Kohob-kojab	Sarkas	abi-abi
Zon	Kahad	Kasse	wafob
Herren	apulhoa	muis	ko-ab
Egen	Suwob	musket	Taka
Strom	Kioe	man	amama
Wander	Kahoe	Wrau	amini
Alleen	Kahoko-Mona	Koro	awoboo
Waar	Kopob	Tader	amama
Water	Koho-labo	Lustoe	Kahomun
Heer	abi-abi	Wodoe	Johé-awoboo
Heer	Biloe		Konni
Berg	Kahon	Shuru	Papohida
Eiland	Kohob	Koste	Sela

Schip	Maeta-bon-Sera
Leem	oedje
Heer	Praauw
Kanjoeng	Jeeba
Kas	Kahoe
Frucht	Koero-Kahoe
Heeren	Kahoko
Kost	Panwoboo-awoboo
Kanon	Kota-Kahoe
Leem	Koi-Kowé
Frucht	Maeta-awoboo
Wander	oko
Frucht	Frucht
Leem	Leem
Kost	awoboo
Kost	Kahoe
Leem	oko
Leem	Kota
Leem	Kahoe
Leem	awoboo-awoboo
Kost	apaj-afab
Leem	awoboo-afab
Leem	Kahoe

104 items

Von Rosenberg (1855)

154 items



Source gallica.bnf.fr / Bibliothèque nationale de France

Karl Benjamin Herman von Rosenberg
([source](#) with cco license and in [Bibliothèque National](#))

Von Rosenberg (1855)

154 items

22

Woordenlijst.

Staan	Jaramatrego	duiven	mi-komabeun
Gaan	^{Edie} Edie	stelen	paiveta
Loopen	Jarapoerpoera	leugen	pauro-puro
Zitten	ellaké-god	kaeyen	achyie
Komen	Tamaroeij	neelden	paewa-oea
Blijven	Dunackad	binden	paani-i
Geven	Pahè	twilen	gersello-mello
opnen	Peparich-i	veremmen	paererao lewa
veronten	pemanabo	ruis en	lewa
draagen	pagekli	viszen	kakaleka
missen	popasij	maakten	fatjed-wau
Zien	popasij	huizen	nahafe
maken	pepovele-i-tyok-i	volgen	kahakapijio
Zaakten		namen	nej
opreken	dipaganan	wit	gahijka
eten	Faginono	rood	si-kovey
drinken	fajino wa lewa	blauw	kakiaki
roken	lewa	erwt	himono
	poerwa komi	zeel	kanoea

Woordenlijst.

- Jaramatrego.
- Edie-édie.
- Jarapoerpoera.
- Matrego.
- Tawawej.
- Panoekoe.
- Pahè.
- Pemanabo.
- Pàgehlie.
- Popo-ej.

Van der Straaten & Severijn (1855)

VERSLAG

VAN EEN IN 1854 BEWERKSTELLIGD ONDERZOK OP HET EILAND

ENGANO,

door den Luitenant ter zee 1^e klasse, kommandant van *Z. M. Schoenerbrik Sylph*,

J. VAN DER STRAATEN,

en den Kontrolleur der 1^e klasse bij het binnenlandsch bestuur ter westkust van Sumatra,

P. SEVERIJN.

Voorwoord.

In het tweede deel, bladz. 479 en volgg. van dit tijdschrift, is geplaatst een verslag omtrent het eiland Engano door den djoeragan Boewang.

Haar

Voorhoofd

Oogen

Mond

Tand

Neus

Tong

Hals

Borst

Buik

Arm

Voet

Teen

Laus

Oedoc.

Koc-oe.

Baka.

Ko-a.

Nienie.

Pano.

Dio.

Tioeano.

Berakidoija.

Toeka.

Poeafa.

Edjaauw.

Akinoijai.

Poké poké.

199 items

Walland (1864)

276 items

Hooge taal.

Eten.	<i>eat</i>	menoh.
Droomen.		à-nihai.
Kruipen.	<i>crawl</i>	o-oboh.
Zingen.	<i>sing</i>	khorah.
Slaan.		bai-djihie.
Krabben.		phio-noh.
Knijpen.		pie-ie.
Wrijven.	<i>rub</i>	paka-kahah.
Duwen.		kebadah.
Stooten.		ha-èh.

Lage taal.

pakehnonno.
kala-lehah.
patta haleh.
kanihonie.
fabieah.
khi-ioh.
nja-hai.

The only
19th-century author
claiming
speech-level system
in Enggano

Hooge taal. (high
language)

Lage taal. (low language)

Walland (1864)

Hoe is uw naam?	ha-nia-moe?
Van waar komt gij?	koeloe-ia-oh?
Kom hier.	adeh-waij dje-éh.
Blijf daar.	kaiama djidjah (ohbea djidjah).
Gaat spoedig.	ma-akoe kehedie.
Verwijder u.	kahi-ah.
Ik heb honger.	oewah hiko.
Ik wil eten.	ika manoh.
Ik wil slapen.	ika bahoea.
Ik wil mij baden.	ika baikeh.
Het is warm.	mokhoh kehah.
„ „ koud.	mokhoh hioh.
Geef mij pisang.	ano-kiaij-ajoh.
Gaat daar zitten.	hèhoe djidjah.
Zijt niet bang.	iada niekie niekie.
Raap dat op.	na-ma-ana.
Vaart-wel.	wah-dji-eh.

The first to
include
phrases and
clauses of
Enggano

17 items

Walland (1864)

The first to include phrases and clauses of Enggano (digitised & translated version)

category	English	Dutch	Enggano
phrases/clauses	What's your name?	Hoe is uw naam?	ha-nia-moe?
phrases/clauses	where do you come from?	Van waar komt gij?	koeloe-ia-oh
phrases/clauses	come here!	Kom hier	adeh-waij dje-éh
phrases/clauses	stay there!	Blijf daar	kaiama djidjah (ohbea djidjidah)
phrases/clauses	go soon!	Gaat spoedig	ma-akoe kehedie
phrases/clauses	move!	Verwijder u	kahi-ah
phrases/clauses	I'm hungry	Ik heb honger	oewah hiko
phrases/clauses	I want to eat	Ik wil eten	ika manoh
phrases/clauses	I want to sleep	Ik wil slapen	ika bahoea
phrases/clauses	I want to have a bath	Ik wil mij baden	ika baikeh

Francis (1870, in Oudemans 1889)

99 items

NEDERLANDSCH.	FRANCIS.	VAN DER STRAATEN EN SEVERIJN.	VON ROSENBERG.	BOEWANG.
zwemmen	kaäke	pakiea	poeroe lewo lewo	kahai
1 = 1	kahaie	egaij	daheij, dahei	adoea
2 = 2	adoea	adoloe	adoloe	akoloe
3 = 3	akoloe	koloe	agoloe	ampapa
4 = 4	afa	afa	aopa	aliema
5 = 5	aliema	lieba	aliema	akie akiema
6 = 6	kaikiene	akeno	akiaknio, akiakia	aliema adoe, aliemei a- doea
7 = 5 + 2	aliema é adoea	lieba doea	aliema agoloe, aliemei agoloe	aliema adoea
8 = 4 + 4 of 5 + 3	afa é afa	afa afa	aliema aopa, aliemei aopa	apa joepa
9 = 5 + 4	aliema é afa	lieba afa	tahapoeloe	apa apa adoea
10 = 10 of 4 + 4 + 2	kahafoeloe	tapoeloe	tahapoeloe alima	
11 = 10 + 1	kahafoeloe é kahaie		kahei taka	
12 = 10 + 2	kahafoeloe é adoea		kahei taka tahapoeloe	
13 = 10 + 3	kahafoeloe é akoloe		adoeei taka	
14 = 10 + 4	kahafoeloe é afa		adoeei taka tahapoeloe	
15 = 10 + 5	kahafoeloe é aliema		akoloe taka	
20 = 1 . 20	kahaie takka		akoloe taka tahapoeloe	
30 = 1 . 20 + 10	kahaie takka é kahafoeloe		aopeia taka	
40 = 2 . 20	adoea takka		aopeia taka tahapoeloe	
50 = 2 . 20 + 10	adoea takka é kahafoeloe		alimei taka	adoea taka tapoeloe (50)
60 = 3 . 20	akoloe takka			aliema takka (100)
70 = 3 . 20 + 10	akoloe takka é kahafoeloe			
80 = 4 . 20	afa takka			
90 = 4 . 20 + 10	afa takka é kahafoeloe			
100 = 5 . 20	aliema takka			
200 = 10 . 20	kahafoeloe takka			

Comparative
list with the
earlier
Enggano
word lists

Oudemans, Jean Abraham Chrétien. 1889. Engano (bewesten Sumatra), zijne geschiedenis, bewoners en voortbrengselen. *Tijdschrift van het Koninklijk*

Nederlandsch Aardrijkskundig Genootschap 6(1). 109–164.

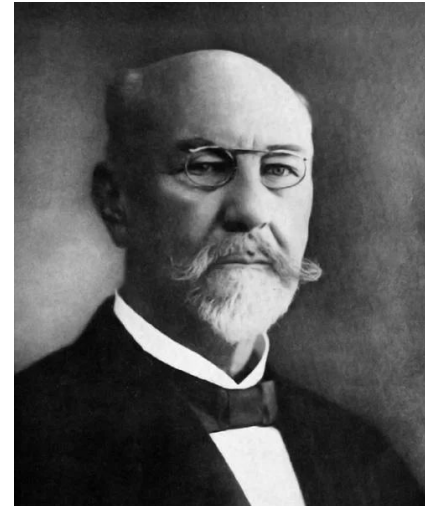
Oudemans (1879)

150 items

Comparative list with the other Barrier Island Languages (i.e., Mentawai and Nias)

WOORDENLIJST
VAN DE TALEN VAN ENGGANO, MENTAWEI EN NIAS
DOOR
J. A. C. OUDEMANS.

	ENGGANO.	MENTAWEL.	NIAS.
Hoofd	é oeloe	oté	heugen
Oor	kaliha	talinga	talinga
Neus	éfano	asak	idong
Mond	ékā	bibo	baba
Oog	ébāka	mata	mata
Kin	ékie	bakēla	bèmbè
Hoofdhaar	foedoeoe-oeloe	alai	boe-oe
Baard	fororoife	boeloe	gambi
Hals	hoenoe	lolokat	bāgi
Borst	émāha	karoeekat	baloedjoesoe
Buik	kiedāi	bācha	daloe
Arm	é afo	para	paño



Dr Antonie Cornelis Oudemans, Jzn.

[Source](#)

Oscar Louis Helfrich

The first providing detailed linguistic descriptions of Enggano in the 19th century



from Aardweg et al. (1938: 632)

Helfrich & Pieters (1891)

MALEISCH. NEDERLANDSCH. ENGANEESCH.

تڤكت toengkat.	stok.	obak ² .
تڤكر tengkar.	krakeel, harrewar- rerij.	paoewē.
تڤكس tangkis.	het afweren, af- wenden.	kieieja.
تڤكس toengkoes.	pakje, bundel.	pieohak ² .
تڤكڤ tangkap.	het aanvatten, vast- grijpen, pakken.	kaie.
تڤكڤ tangkoep.	plat op den buik liggen.	pēkieofo.
تڤكڤ tēngkoek.	de nek.	òkahaē.
تڤكو tongkau.	drie steenen voor treeft om het kook- sel op te zetten.	koenajo.

“the hitherto
longest and most
comprehensive
collection of the
Enggano language”

1,012 items

34 sentences

(Krauß, *under
review*: 26)

Helfrich & Pieters (1891)

- **Preceded by** ethnographic description of the Enggano island (**Helfrich 1888**), incl. (i) over 200 cultural terms and (ii) 18 words with dialectal variation

Helfrich, Oscar Louis. 1888. De eilandgroep Engano. *Tijdschrift van het Koninklijk Nederlandsch Aardrijkskundig Genootschap* 5. 272–314.

Helfrich, Oscar Louis. 1893. Aanvullingen en verbeteringen op de Maleisch-Nederlandsch-Enganeesch woordenlijst. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 35. 228–233.

Helfrich & Pieters (1891)

- **Followed by** correction in **Helfrich (1893)**

Helfrich, Oscar Louis. 1888. De eilandgroep Engano. *Tijdschrift van het Koninklijk Nederlandsch Aardrijkskundig Genootschap* 5. 272–314.

Helfrich, Oscar Louis. 1893. Aanvullingen en verbeteringen op de Maleisch-Nederlandsch-Enganeesch woordenlijst. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 35. 228–233.

Modigliani (1894)

266

Capitolo XVII

Another comparative list
but with Italian

ITALIANO	MALESE	NIAS	TOBA-BATACCO	ENGANESE
Abbassare	Turún	Manada	Munduc *	Paúho tópo
Abbastanza	Sampé	Irugi	Siccop *	Ho abacó
Abbastanza (Non)	Tida sampé	Loo irugi	Sowada siccop *	Chèfe iabacó
Abitazione grande	Rúma besár	Banúa sebúa	Bagas na balga	Èuba
Accanto	Decát	Bangái	Giala	Panéna
Accesso, piaga	Bárah	Fao sa	Cugang	Baro
Accettare	Trimá	Tema halo	Mandgialo *	Menaá
Accomodare	Membaíchi	Gama ²	Mangundgiam *	Parie ecáua
Acqua	Aier	Idáno	Aec	Ebò
A. mare	Aier laót	Aec ni tao * (vera- mente acqua di lago perchè i Toba non hanno mare).	Lebo. ² Eué

546 items (of Italian headwords)

Holle List of Enggano (collected in 1895)

The “Holle List” was a series of publication by Wim A. L. Stokhof

878 items (of
Enggano translation)

The Enggano one appears in Stokhof and Almanar (1987)

Rajeg, Gede Primahadi Wijaya. 2023a. Digitised, searchable Holle List in Stokhof (1980). University of Oxford, UK. <https://doi.org/10.5281/zenodo.7972273>.

Rajeg, Gede Primahadi Wijaya. 2023b. CLDF dataset of the Enggano word list from 1895 in Stokhof and Almanar's (1987) Holle List. Zenodo. <https://doi.org/10.5281/zenodo.8038975>.

Stokhof, W. A. L. (ed.). 1980. *Holle lists, vocabularies in languages of Indonesia, Vol. 1: Introductory Volume*. Vol. Materials in Languages of Indonesia. 1 vols. Canberra, A.C.T., Australia: Dept. of Linguistics, Research School of Pacific Studies, The Australian National University. <https://core.ac.uk/reader/159464813>. (21 May, 2023).

Stokhof, W. A. L. & Alma E. Almanar. 1987. *Holle lists, vocabularies in languages of Indonesia, Vol. 10/3: Islands off the west coast of Sumatra* (Pacific Linguistics (Series D) 76). Vol. Materials in Languages of Indonesia. 36 vols. Canberra, A.C.T., Australia: Dept. of Linguistics, Research School of Pacific Studies, The Australian National University. <http://hdl.handle.net/1885/144589>.

Materials from the 20th and 21st century

- Dictionaries (Helfrich 1916; Kähler 1987)
- Word list (Kasim et al. 1987) & loanwords (Nothofer 1992)
- Contemporary word list (Edwards et al. 2022)
- Lexicon (based on the corpora of the Documentation project) – the basis of online Enggano dictionary



Helfrich, Oscar Louis. 1916. Nadere bijdrage tot de kennis van het engganeesch. *Bijdragen tot de taal-, land- en volkenkund.* Leiden, The Netherlands: Brill 71(1). 472–555. <https://doi.org/10.1163/22134379-90001719>.

Kähler, Hans. 1987. *Enggano-Deutsches Wörterbuch* (Veröffentlichungen Des Seminars Für Indonesische Und Südseesprachen Der Universität Hamburg 14). Berlin; Hamburg: Dietrich Reimer Verlag.

Kasim, Yuslina, Marjusman Maksan, Syamsir Arifin & Zailoet Zailoet. 1987. *Pemetaan bahasa daerah di Sumatra Barat dan Bengkulu.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa. <https://repositori.kemdikbud.go.id/1684/>. (28 January, 2024).

Owen Edwards (collector), Engga Zakaria Sangian (consultant), I Gusti Ayu Gde Sosiowati (interviewer, transcriber), Ni Wayan Sukarini (interviewer, transcriber), 2022. *Enggano Wordlist.* MPEG/MP4/MXF/VND.WAV/EAF+XML/PDF/PLAIN. OCSEAN-ENO_20220712 at catalog.paradisec.org.au. <https://dx.doi.org/10.26278/R0N1-BG13>

Outlines

1. What are “lexical resources”? 
2. What are the existing linguistic materials for Enggano? 
3. How do we curate these materials into lexical resources?

- Digitisation
- Merging different data
- Translation
- Etymologisation
- Orthography transliteration

- Digitisation
- Merging different data
- Translation
- Etymologisation
- Orthography transliteration

Curating the Enggano Holle List (1895)

(Rajeg 2023a, b)

Rajeg, Gede Primahadi Wijaya. 2023a. Digitised, searchable Holle List in Stokhof (1980). University of Oxford, UK. <https://doi.org/10.5281/zenodo.7972273>.

Rajeg, Gede Primahadi Wijaya. 2023b. CLDF dataset of the Enggano word list from 1895 in Stokhof and Almanar's (1987) Holle List. Zenodo. <https://doi.org/10.5281/zenodo.8038975>.

1 - The reference Holle List (Stokhof 1980)

5. THE NEW BASIC LIST (NBL)⁷

ID	DUTCH	ENGLISH	MALAY	1894	1904/11	1931
1.	lichaam	body	badan, tubuh	4	4	1
2.	hoofd	head	kepala	5	5	2
3.	gezicht, aange- zicht	face	muka, wajah	6	6	3
4.	voorhoofd	forehead	dahi	24	25	4
5.	schedel	skull	tempurung kepala	26	27	5
6.	hoofdhaar	hair	rambut	72	73	6
7.	kaal	bald	gundul, botak	-	-	7
8.	kruin (v.h.hoofd)	crown	hulu kepala	7	7	8
9.	oor	ear	telinga	8	8	9
10.	oorsmeer	earwax	tahi telinga	-	-	10
11.	oog	eye	mata	9	9	11

2 – The regional Holle List for Enggano (Stokhof & Almanar 1987)

2. THE ENGGANO LIST
ID ENGGANO
1. kārāhā, koedōdōkō
 2. èoeloe/èoedoe <1>
 3. èbaka
 4. èkoe(k)
 5. èaoeloe
 6. poeroeroe èoeloe
 8. kèlěkèlě poeěkā
 9. kariehā, kaliehě
 11. èbaka

Challenge 1:

The **reference** Holle List is
separated from the
regional Holle List

The IDs in the two lists need to be *matched!*

5. THE NEW BASIC LIST (NBL)⁷

ID	DUTCH	ENGLISH	MALAY
1.	lichaam	body	badan, tub
2.	hoofd	head	kepala
3.	gezicht, aange- zicht	face	muka, waja
4.	voorhoofd	forehead	dahi
5.	schedel	skull	tempurung
6.	hoofdhaar	hair	rambut
7.	kaal	bald	gundul, bo
8.	kruin (v.h.hoofd)	crown	hulu kepal
9.	oor	ear	telinga
10.	oorsmeer	earwax	tahi telin
11.	oog	eye	mata

2. THE ENGGANO LIST

ID	ENGGANO
1.	kārāhā, koedödökö
2.	èoeloe/èoedoe <1>
3.	èbaka
4.	èkoe(k)
5.	èaoeloe
6.	poeroeroe èoeloe
8.	kèlëkèlë poeëkā
9.	kariehā, kaliehě
11.	èbaka

The IDs in the two lists need to be *matched!*

5. THE NEW BASIC LIST (NBL)⁷

ID	DUTCH	ENGLISH
----	-------	---------

- | | | |
|-----|--------------------------|----------|
| 1. | lichaam | body |
| 2. | hoofd | head |
| 3. | gezicht, aange-
zicht | face |
| 4. | voorhoofd | forehead |
| 5. | schedel | skull |
| 6. | hoofdhaar | hair |
| 7. | kaal | bald |
| 8. | kruin (v.h.hoofd) | crown |
| 9. | oor | ear |
| 10. | oorsmeer | earwax |
| 11. | oog | eye |

MALAY

- | |
|------------|
| badan, tul |
| kepala |
| muka, waja |
| dahi |
| tempurung |
| rambut |
| gundul, bo |
| hulu kepal |
| telinga |
| tahi telin |
| mata |

2. THE ENGGANO LIST

ID	ENGGANO
----	---------

- | | |
|-----|-------------------|
| 1. | kārāhā, koedōdōkō |
| 2. | èoeloe/èoedoe <1> |
| 3. | èbaka |
| 4. | èkoe(k) |
| 5. | èaoeloe |
| 6. | poeroeroe èoeloe |
| 8. | kèlěkèlě poeěkā |
| 9. | kariehā, kaliehě |
| 11. | èbaka |

The IDs in the two lists need to be *matched*!

5. THE NEW BASIC LIST (NBL)⁷

ID	DUTCH	ENGLISH
----	-------	---------

- | | | |
|------|--------------------------|----------|
| → 1. | lichaam | body |
| 2. | hoofd | head |
| → 3. | gezicht, aange-
zicht | face |
| 4. | voorhoofd | forehead |
| 5. | schedel | skull |
| 6. | hoofdhaar | hair |
| 7. | kaal | bald |
| 8. | kruin (v.h.hoofd) | crown |
| 9. | oor | ear |
| 10. | oorsmeer | earwax |
| 11. | oog | eye |

MALAY

- | |
|------------|
| badan, tul |
| kepala |
| muka, waja |
| dahi |
| tempurung |
| rambut |
| gundul, bo |
| hulu kepal |
| telinga |
| tahi telin |
| mata |

2. THE ENGGANO LIST

ID	ENGGANO
----	---------

- | | |
|------|-------------------|
| → 1. | kārāhā, koedōdōkō |
| 2. | èoeloe/èoedoe <1> |
| 3. | èbaka |
| 4. | èkoe(k) |
| 5. | èaoeloe |
| 6. | poeroeroe èoeloe |
| 8. | kèlěkèlě poeěkā |
| 9. | kariehā, kaliehě |
| 11. | èbaka |

Challenge 2:

- > The lists are in PDFs
- > Manual matching is tedious
- > Aim for replicable workflow

1. Digitise the reference and regional Holle List

1. Digitise the reference and regional Holle List

- > O(ptical) C(haracter) R(ecognition)
- > manual correction for typos

Thanks to Mary Dalrymple for the OCR and Fitri Koemba, S.S. for the correction.

The digitised *reference* Holle List (Rajeg 2023a)

Interactive, web version (<https://engganolang.github.io/digitised-holle-list/>)

1. lichaam body
2. hoofd head
3. gezicht, aangezicht face
4. voorhoofd forehead
5. schedel skull
6. hoofdhaar hair
7. kaal bald
8. kruin (v.h.hoofd) crown
9. oor ear
10. oorsmeer earwax
11. oog eye



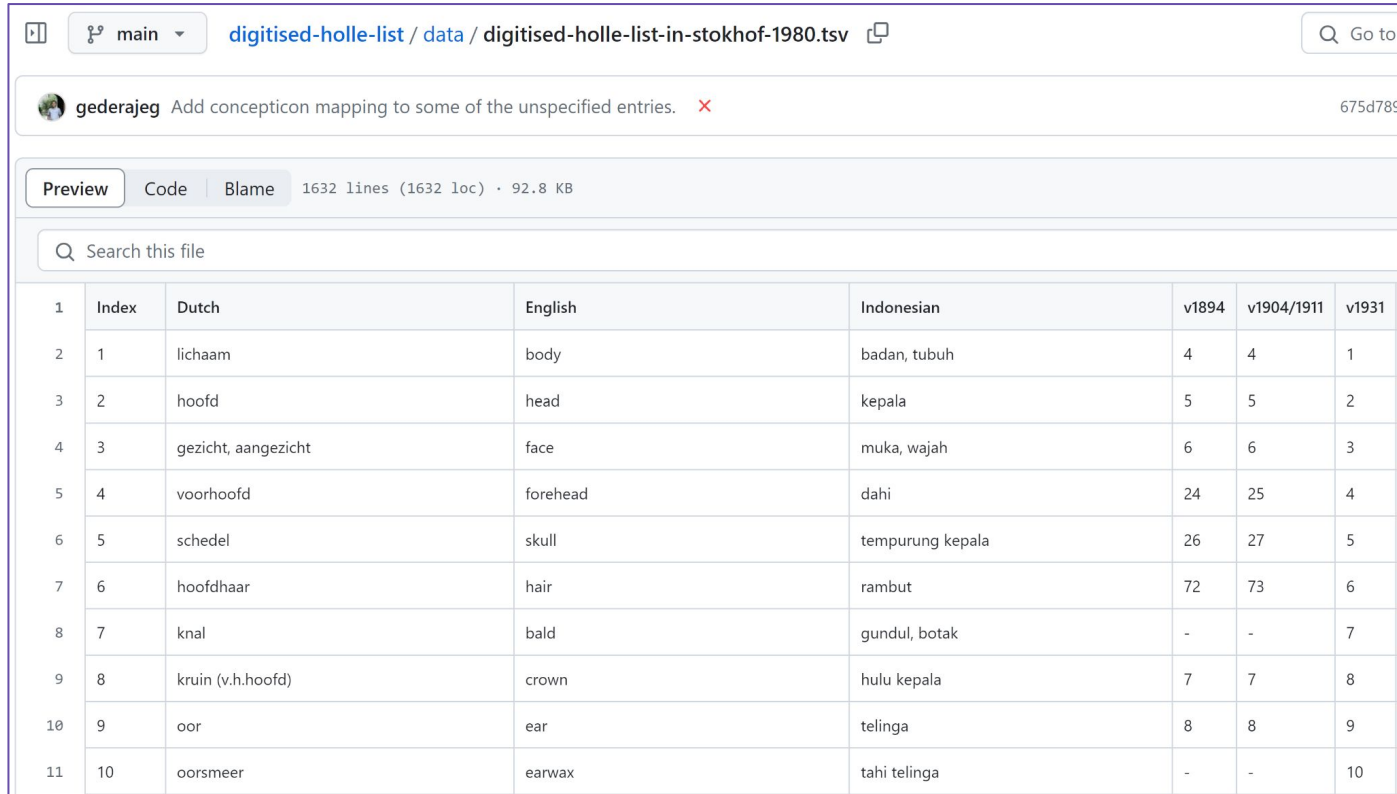
► Code

Table 1: The digitised, new basic list of the Holle List in Stokhof (1980: 22–72)

Index	Dutch	English	Indonesian	v1894	v1904/1911	v1931
1	lichaam	body	badan, tubuh	4	4	1
2	hoofd	head	kepala	5	5	2
3	gezicht, aangezicht	face	muka, wajah	6	6	3
4	voorhoofd	forehead	dahi	24	25	4
5	schedel	skull	tempurung kepala	26	27	5
6	hoofdhaar	hair	rambut	72	73	6
7	knal	bald	gundul, botak	-	-	7
8	kruin (v.h.hoofd)	crown	hulu kepala	7	7	8
9	oor	ear	telinga	8	8	9
10	oorsmeer	earwax	tahi telinga	-	-	10
11	oog	eye	mata	9	9	11

The digitised *reference* Holle List (Rajeg 2023a)

Raw plain-text file (<https://github.com/engganolang/digitised-holle-list/tree/main/data>)



The screenshot shows a GitHub file viewer for the file `digitised-holle-list / data / digitised-holle-list-in-stokhof-1980.tsv`. The file is 92.8 KB and contains 1632 lines of code. A search bar is present above the table. The table itself has 7 columns: Index, Dutch, English, Indonesian, v1894, v1904/1911, and v1931. It lists 10 entries related to body parts and head features.

	Index	Dutch	English	Indonesian	v1894	v1904/1911	v1931
1							
2	1	lichaam	body	badan, tubuh	4	4	1
3	2	hoofd	head	kepala	5	5	2
4	3	gezicht, aangezicht	face	muka, wajah	6	6	3
5	4	voorhoofd	forehead	dahi	24	25	4
6	5	schedel	skull	tempurung kepala	26	27	5
7	6	hoofdhaar	hair	rambut	72	73	6
8	7	knal	bald	gundul, botak	-	-	7
9	8	kruin (v.h.hoofd)	crown	hulu kepala	7	7	8
10	9	oor	ear	telinga	8	8	9
11	10	oorsmeer	earwax	tahi telinga	-	-	10

The digitised *Enggano* Holle List (Rajeg 2023b)

Raw plain-text file (<https://github.com/engganolang/holle-list-enggano-1895/tree/1.1.0/data-raw>)

2. THE ENGGANO LIST


1. kārāhā, koedödökö
2. èoeloe/èoedoe <1>
3. èbaka
4. èkoe(k)
5. èaoeloe
6. poeroeroe èoeloe
8. kèlèkèlè poeëkā
9. kariehā, kaliehě
11. èbaka



```
Enggano-Holle-List - Notepad
File Edit Format View Help
# Appearing in Stokhof and Almanar (1987: 189-205)
# Year of investigation: 1895
# Investigator: Abs vd Noord
# Place of investigation: Pulau Dua (in Enggano island)
# OCR by Mary Dalrymple
# Checked (proofread and corrected) and processed (annotated
footnotes and wrangled into a table) by Gede Primahadi Wijaya
Rajeg

<wordlist>
1. kārāhā, koedödökö
2. èoeloe/èoedoe <1>
3.     èbaka
4.     èkoe(k)
5.     èaoeloe
6.     poeroeroe èoeloe
8.     kèlèkèlè poeëkā
9.     kariehā, kaliehě
11.    èbaka
```

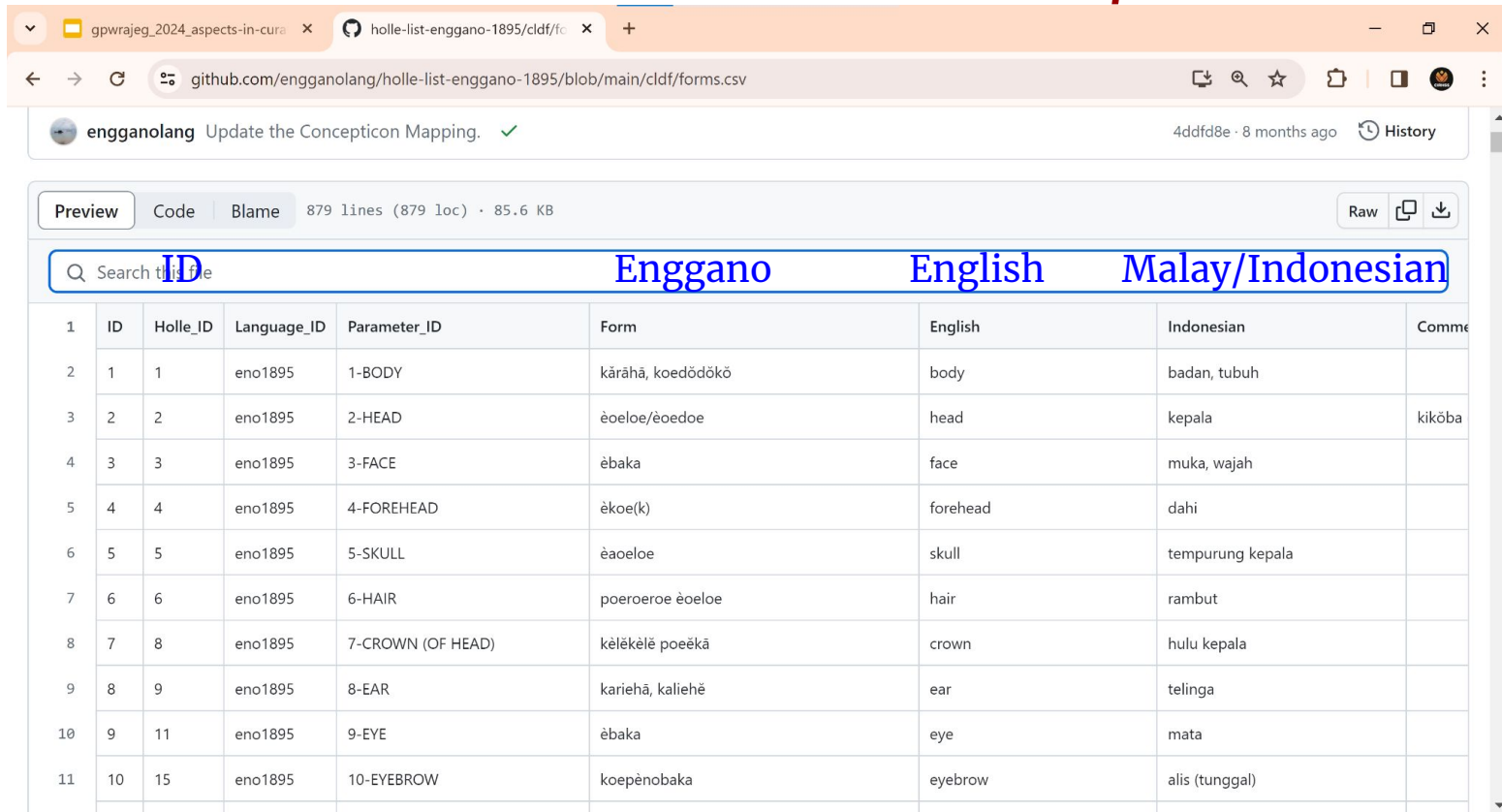
Ln 1, Col 1 100% Unix (LF) UTF-8

- Digitisation 
- Merging different data
- Translation
- Etymologisation
- Orthography transliteration

*Write **R codes** to
merge the (i)
reference HL and (ii)
the Enggano HL by
matching IDs*

<https://github.com/engganolang/holle-list-enggano-1895/tree/main/code>

The Enggano Holle List with English & Indonesian translation (now searchable & manipulatable)



github.com/engganolang/holle-list-enggano-1895/blob/main/cldf/forms.csv

engganolang Update the Concepticon Mapping. ✓ 4ddfd8e · 8 months ago History



Preview Code Blame 879 lines (879 loc) · 85.6 KB Raw Copy Download

Search this file

	ID	Enggano	English	Malay/Indonesian				
1	ID	Holle_ID	Language_ID	Parameter_ID	Form	English	Indonesian	Comment
2	1	1	eno1895	1-BODY	kārāhā, koedōdōkō	body	badan, tubuh	
3	2	2	eno1895	2-HEAD	èoeloe/èoedoe	head	kepala	kikōba
4	3	3	eno1895	3-FACE	èbaka	face	muka, wajah	
5	4	4	eno1895	4-FOREHEAD	èkoe(k)	forehead	dahi	
6	5	5	eno1895	5-SKULL	èaoeloe	skull	tempurung kepala	
7	6	6	eno1895	6-HAIR	poeroeroe èoeloe	hair	rambut	
8	7	8	eno1895	7-CROWN (OF HEAD)	kèlèkèlè poeëkà	crown	hulu kepala	
9	8	9	eno1895	8-EAR	karihā, kaliehè	ear	telinga	
10	9	11	eno1895	9-EYE	èbaka	eye	mata	
11	10	15	eno1895	10-EYEBROW	koepènobaka	eyebrow	alis (tunggal)	

<https://github.com/engganolang/holle-list-enggano-1895/blob/main/cldf/forms.csv>

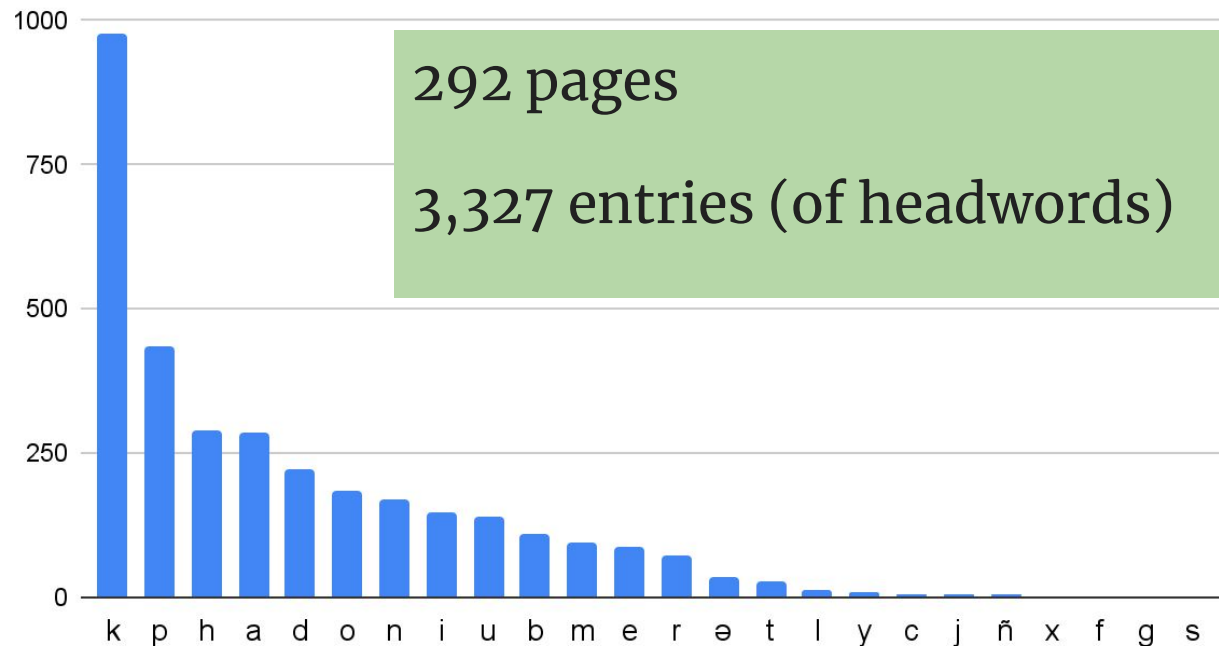
Now, the other (digitised) regional HLLs can be matched computationally with the reference HLL to gain their translations.

- Digitisation 
- Merging different data 
- Translation
- Etymologisation
- Orthography transliteration

Retro digitising the
Enggano-German
dictionary
(Kähler 1987)

Enggano-German (Kähler 1987)

Number of stems/headwords per alphabet (manual count)



Enggano-German (Kähler 1987: 276)

tē	(DIA) = tō, dō	Art und Weise (§ 31a)
	kitē = kidō	sein wie
tebe:	itebe = iðebe (DIA)	an der Oberseite, oben (§ 26b)
teʔei:	kaʔiteʔei	jmd schlechte Speisen anbieten
teo	kiteo ekitai	überreden
təhəda	(DIA) = dəhəda	
tixoi	= didixoi, didiki: kitixoi	etwas umwickeln
tikī	kahatikī	(mit etwas Biegsamem) schlagen
°etirī	(< ML sirih)	Betelbissen
tō	= dō, tē (DIA)	wie (beim Vergleich; § 31a)
	kitoiya < kitō eiya	wie das Wesen sein

Challenge 1:

deciphering the
components of a
dictionary entry

Enggano-German (Kähler 1987: 276)

tē	(DIA) = tō, dō	Art und Weise (§ 31a)
	kitē = kidō	sein wie
tebe:	itebe = iðebe (DIA)	an der Oberseite, oben (§ 26b)
teʔei:	kaʔiteʔei	jmd schlechte Speisen anbieten
teo	Enggano HEADWORDS	
təhəda	(DIA) = dəhəda	
tixoi	= didixoi, didiki: kitixoi	etwas umwickeln
tikī	kahatikī	(mit etwas Biegsamem) schlagen
°etirī	(< ML sirih)	Betelbissen
tō	= dō, tē (DIA)	wie (beim Vergleich; § 31a)
	kitoiya < kitō eiya	wie das Wesen sein

Enggano-German (Kähler 1987: 276)

tē	(DIA) = tō, dō	Art und Weise (§ 31a)
	kitē = kidō	sein wie
tebe:	itebe = iðebe (DIA)	VARIANTS (=) (b)
teʔei:	kaʔiteʔei	jmd schlechte Speisen anbieten
teo	kiteo ekital	überreden
təhəda	(DIA) = dəhəda	
tixoi	= didixoi, didiki:	kitixoi etwas umwickeln
tikī	kahatikī	(mit etwas Biegsamem) schlagen
°etirī	(< ML sirih)	Betelbissen
tō	= dō, tē (DIA)	wie (beim Vergleich; § 31a)
	kitoiya ‘ kitō eiya	wie das Wesen sein

Enggano-German (Kähler 1987: 276)

tē	(DIA) = tō, dō	Art und Weise (§ 31a)
	kitē = kidō	sein wie
tebe:	itebe = iðebe (DIA)	an der Oberseite, oben (§ 26b)
teʔei:	kaʔiteʔei	jmd schlechte Speisen anbieten
teo	kiteo ekitai	überreden
təhəda	(DIA) = dəhəda	
tixoi	= didixoi, didiki: kitixoi	etwas umwickeln
tikī	kahatikī	(mit etwas Biegsamem) schlagen
°etirī	(< ML sirih)	Betelbissen
tō	= dō, tē (DIA)	wie (beim Vergleich; § 31a)
	kitoiya < kitō eiya	wie das Wesen sein



GERMAN

Enggano-German (Kähler 1987: 276)

tē	(DIA) = tō, dō	Art und Weise (§ 31a)
kitē	= kidō	sein wie
tebe:	itebe = iðebe (DIA)	an
teʔei:	kaʔiteʔei	jmd schlechte
teo	kiteo ekitai	überreden
təhəda	(DIA) = dəhəda	
tixoi	= didixoi, didiki: kitixoi	etwas umwickeln
tikī	kahatikī	(mit etwas Biegsamem) schlagen
°etirī	(< ML sirih)	Betelbissen
tō	= dō, tē (DIA)	wie (beim Vergleich; § 31a)
kitoiya	< kitō eiya	wie das Wesen sein

DERIVED
FORMS/EXAMPLES

Enggano-German (Kähler 1987: 276)

tē	(DIA) = tō, dō	Art und Weise (§ 31a)
kitē	= kidō	sein wie
tebe:	itebe = iðebe (DIA)	an
te [?] ei:	ka [?] ite [?] ei	jmd schlechte Speisen anbieten
teo	kiteo ekitai	überreden
təhəda	(DIA) = dəhəda	
tixoi	= didixoi, didiki: kitixoi	etwas umwickeln
tikī	kahatikī	(mit etwas Biegsamem) schlagen
^o etirī	(< ML sirih)	Betelbissen
tō	= dō, tē (DIA)	wie (beim Vergleich; § 31a)
kitoiya	< kitō eiya	wie das Wesen sein

CROSS-REF.



Enggano-German (Kähler 1987: 276)

tē (DIA) = tō, dō Art und Weise (§ 31a)

kitē = kidō

tebe: itebe = iðe

teʔei: kaʔiteʔei

teo kiteo ekita

təhəda (DIA) = dəhə

tixoi = didixoi,

tikī kahatikī

°etirī (< ML sirih) Betelbissen

tō = dō, tē (DIA) wie (beim Vergleich; § 31a)

kitoiya < kitō eiya wie das Wesen sein

Many more:

- Loanword
- Etymology
- Remark/Note
- Direction of derivation
- Homograph/Homonym ID

Challenge 2:

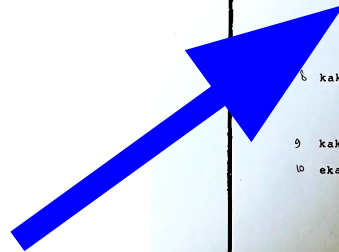
Lexicographic training
for transcribers

Reverse-engineering of
the entry to be
transcribed

Reverse-Engineering 1: Manual identification of the entry component in the print out.

104

-ka?a	unser (in), von mir und euch: Possessivsuffix der 1.Pl in (§ 11)
eubāka?a	unser (in) Haus
eka?a?aho	= eka?aho, eka?a?oho Baumart (ML kasai, Pometia pinnata Forst.)
Ka?a?āi	Ort auf Enggano
ka?abəha	(ka + abəha) und nachdem (Konjunktion)
ka?abu	kika?abu vergessen
hō buka?abu	vergaß bereits
kika?abui	jmd, etwas vergessen
dibuka?abui	sie vergaßen etwas
eka?abua	Vergessensort
eaka?abu	= eka?abu das Vergessen
eka?abuía?a	d Vergessene
eka?abuía?au edobu	mein Vergessenes war
Sachen	
eka?abuiyo	d zu Vergessende



117

1 kakahao	← kahao same form
2 kaka?fōhōi	← ka?fōhōi same form
3 kakanāka	← kākā same form
4 kakano?a	← kano?a same form
5 kakape	→ kikakapei bestreiten, leugnen → kikakapə?a etwas leugnen → kikakapə?ə?ə?ə? (132z14) herausgeben
6 kakapi	← kapi same form
7 kakarai	= kakarei, (kakadei (DIA): kikakarai (ver)folgen, verjagen, jagen; fortziehen (intr)) → ?ua kikakarai ebeo e?ana ich verjage den Hund → kabukakaraika kia er jagte sie fort → edikakarai d Fortgejagte → kakareica (?) zu erlangen trachten ← kari: kikakarīa?a sich umdrehen → kipakakarīa?a etwas umdrehen, rotieren lassen → edipakarīa?a d Umgewendete
9 kakarodi	← karodi same form
10 ekakaroku	Baumart; daraus gewonnenes Fischgift (Derris; XIz17f: Fischgift ... stellte man aus den Früchten des gleichnamigen Baumes her, die zu Pulver zerrieben, zu handgroßen Bündeln in Blätter verpackt und über dem Feuer erhitzt wurden. Diese wurden bei Ebbe in die Wasserlöcher des Rifles geworfen, und nach etwa einer Stunde konnte man die betäubten Fische herausnehmen.)
11 kakāupe	= kāupe bevor
12 kaki	→ kikaki hobeln (mit dem Haumesser)

Each page was checked for accuracy

Huge thanks to the Transcription Team



Putu Dea Indah Kartini



Fitriani Putri Koemba



Ni Putu Wulan Lestari



Yul Fulgensia Rusman Pita

<https://www.ling-phil.ox.ac.uk/news/2023/05/28/retro-digitalisation-work-enggano-german-dictionary-udayana-university-indonesia>



Ida Bagus Made Ari Segara



Putu Wahyu Widiatmika



I Gede Semara Dharma Putra

around two and a half months work

Challenge 3:

How to transcribe in
real-time into a single
database storage for all?

Reverse-Engineering 2

*Build an online
database entry
system*



Cokorda Pramatha
(Comp. Sci., Udayana U)



Ida Bagus Gede
Sarasvananda
(Comp. Sci., INSTIKI)

*Crowdsource the
transcription
process in real-time*



Menu

Users

Input Data Kamus

Questionnaire

Logout

Total Stem
3328

Total Example
5711

Recent Stem
nãpũ

Total Users
17

Hello Gede Primahadi

Metadata

Alphabet Page in
Entry No

Input Stem

Stem Form
German Translation
homonymID

Recent Stem

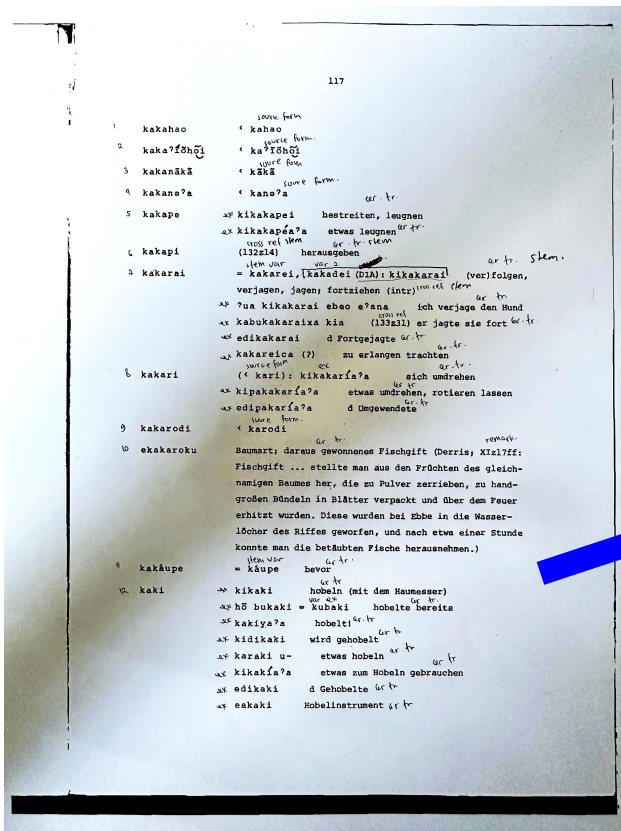
nãpũ
by Putu Dea Indah Kartini

nãpũ
by Putu Dea Indah Kartini

napõnã?ã
by Putu Dea Indah Kartini

Reverse-Engineering 3

Type in (by hand) the identified component into the relevant slot



Metadata

Alphabet

K

Page in
Kähler

157

Entry No

9

Input Stem

Stem Form

kodo

German
Translation
homonymID

Challenge 4:

Processing the transcribed data into excel-like format

Handling strange characters in the output

Snippet of the digitised Enggano-German dictionary



1_stem_german_translation-to-check



File Edit View Insert Format Data Tools Extensions Help Accessibility



Share



90%



%

.0

.00

123

Default...



10



B

I



A



A1

stem_id

	A	B	C	D	E	F	G	H	I
1	stem_id	kms_Alphabet	kms_page	kms_entry_no	stem_form	German	German_corrected	category	English
2	11_1685428252	a	1	1	a-	leitet Objektssätze ein		stem_GermanTr:	introduces object sentences
3	11_1685428330	a	1	2	a-	wenn, falls; denn, weil		stem_GermanTr:	if, if; because of
4	11_1685428509	a	1	3	[?]a-	Präfix, das Wortstämme erweitert		stem_GermanTr:	Prefix that expands word ste
5	11_1685429142	a	1	4	-[C]a	kennzeichnet Zukunft bei Verben		stem_GermanTr:	indicates future in verbs
6	11_1685429642	a	1	6	?aa	Hallo! He! (Anruf)		stem_GermanTr:	Hello! He! (call)
7	11_1685430716	a	2	1	a?a	geh!		stem_GermanTr:	go!
8	11_1685430762	a	2	2	éa?a	jene(r,s) - korrelat zur 3.P		stem_GermanTr:	those - correlative to the 3rd
9	11_1685430833	a	2	3	-a?a	kennzeichnet die komplexe Variation bei Verben		stem_GermanTr:	marks the complex variation
10	11_1685430954	a	2	4	-a?a	Suffix an Verbalstämmen, das die Aufforderung an die 2.PI kennzeichnet		stem_GermanTr:	Suffix on verbal stems that ir
11	11_1685431789	a	2	10	?a?anō	einige(s), sovieler		stem_GermanTr:	some, so many
12	11_1685431862	a	2	11	eaba	Grasart		stem_GermanTr:	Grass type
13	11_1685431896	a	2	12	e?aba?a:ua	Rochen		stem_GermanTr:	Rays
14	11_1685432004	a	2	13	aba:i kaha:i?i	("es kommt [noch] eines" =) neun		stem_GermanTr:	("It comes [still] one" =) nine
15	11_1685432114	a	2	14	e?abakoko	Dämon, der kleine Kinder entführt		stem_GermanTr:	Demon who kidnapped your
16	11_1685497746	a	3	1	abe	wenn, falls		stem_GermanTr:	if

Write codes (in progress for over 9 months)



main

kahler-1987 / code /



gederajeg Add editing code and de-activate the access to Google

Name



..



1-pre-processing.R



2-read-the-translation-sheets.R



3-read-the-translation-sheet-EXAMPLE.R



4-pre-processing-the-checked-translation.R



5-combine-original-and-checked-translation.R



6-check-stems-and-examples-without-German-translation.R



7-checking-loanword-data.R

<https://github.com/engganolang/kahler-1987>

Some distributional
insights from the
digitised dictionary

Count of etymological info/marker for the headword

```
# A tibble: 7 × 2
```

```
  lang                                     n  
  <chr>                                <int>
```

```
1 PAN (Proto-Austronesian)                86  
2 PMP (Proto-Malayo-Polynesian)          42  
3 ML (Malay [now BI])                     6  
4 SMtw (Sudmentawai [South Mentawai])    2  
5 BAR (Bare`e [Austronesian, Sulawesi])  1  
6 MKB (Minangkabau [Austronesian, Sumatra]) 1  
7 Nias                                     1
```

Count of forms marked as variant (=)

A tibble: 532 × 3

	stem_form	stem_formVariant	variant_forms
	<chr>	<chr>	<int>
1	adiu	adúú ; ariu ; arúú ; aliu	4
2	a?io?oi	ai?io?oi ; aya?ioi ; aya?io?oi ; ayo?oi	4
3	da?a	kida?a ; kinã?ã ; kida?a:o ; kita?a:o	4
4	?adiba	?ariba ; ?aliba ; ?alibe	3
5	arúú	adúú ; adiu ; ariu	3
6	aya?ioi	ayo?oi ; a?io?oi ; aya?ioi	3
7	ci?a	cici?a ; jiji?a ; ici?a	3
8	-da	-ta ; -ra ; -nã	3
9	-da?a	-ra?a ; -nã?ã ; -ta?a	3
10	doo	nõõ ; too ; tee	3

Revamping *EnoLEX*, a diachronic lexical database for Enggano

(Rajeg, Krauße, & Pramarta, to be
presented at *the AsiaLEX 2024* [Tokyo])

<https://github.com/engganolang/enolex>

Snippet of the diachronic lexical database (Krauß, Rajeg, & Pramarta) – wide format

Semantic domain & Glosses

Sources of the words in diachrony

A	D	E	F	G	H	I	J	K
Semantic field	Indonesian gloss	English gloss	Enggano (Brouwer early 1850s)	Enggano (Boewang 1854)	Enggano South? (v. Rosenberg 1855)	Enggano Northwest? (vd Straten & S. 1855)	Enggano North (Walland 1864)	Enggano (Francis 1870)
human objects	rumah	house	djoeba[-beri-berri], joeba		joewa	tjoeba, djoeba	oebah, oebah[-dobah]	oeba, eoeba [kiemo], oeba [ejobie]

word forms

Challenge 1:

Etymologising: grouping
words of the same origin
(in progress by Daniel Krauß)

Etymologising results: A case for 'house'

A tibble: 20 × 5

	id	year	words	gloss_by_year
	<int>	<fct>	<chr>	<chr>
1	1043	<1855	djoeba[-beri-berri] ; joeba	house
2	1043	1855	joewa	house
3	1043	1855	tjoeba ; djoeba	house
4	1043	1864	oebah ; oebah[-dobah]	house
5	1043	1870	oeba ; eoeba [kiemo] ; oeba [ejobie]	house
6	1043	1879	oeba ; eoeba[-kiemo]	house
7	1043	1888	èoeba	house, cage
8	1043	1891	èoebě	house, chest, place
9	1043	1891	èoeba	house, chest, place
10	1043	1894	èuba	house
11	1043	1895	èoba ; èobě ; èoeba ; èoebadiê	house
12	1043	1916	èoeba ; èoebě	house, home, chest, suitcase
13	1043	1982	uba	house
14	1043	1987	euba ; euabai	house
15	1043	2011	iub	house
16	1043	2019	yu	house
17	1044	1894	cacario	house
18	1045	<1855	praauw	house
19	1045	1855	praauw	house
20	1045	1864	prahoe	house

Etymologising results: A case for ‘house’

1	<u>1043</u>	<1855	djoeba[-beri-berrri] ; joeba	house
2	<u>1043</u>	1855	joewa	house
3	<u>1043</u>	1855	tjoeba ; djoeba	house
4	<u>1043</u>	1864	oebah ; oebah[-dobah]	house
5	<u>1043</u>	1870	oeba ; eoeba [kiemo] ; oeba [ejobie]	house
6	<u>1043</u>	1879	oeba ; eoeba[-kiemo]	house
7	<u>1043</u>	1888	èoeba	house, cage
8	<u>1043</u>	1891	èoebě	house, chest, place
9	<u>1043</u>	1891	èoeba	house, chest, place
10	<u>1043</u>	1894	èuba	house
11	<u>1043</u>	1895	èoba ; èobě ; èoeba ; èoebadiê	house
12	<u>1043</u>	1916	èoeba ; èoebě	house, home, chest, suitcase
13	<u>1043</u>	1982	uba	house
14	<u>1043</u>	1987	euba ; euabai	house
15	<u>1043</u>	2011	iub	house
16	<u>1043</u>	2019	yu	house

Cognate set (ID 1043)

Etymologising results: A case for 'house'

A tibble: 20 × 5

	id	year	words	gloss_by_year
	<int>	<fct>	<chr>	<chr>
1	1043	<1855	djoeba[-beri-berri] ; joeba	house
2	1043	1855	joewa	house
3	1043	1855	tjoeba ; djoeba	house
4	1043	1864	oebah ; oebah[-dobah]	house
5	1043	1870	oeba ; eoeba [kiemo] ; oeba [ejobie]	house
6	1043	1879	oeba ; eoeba[-kiemo]	house
7	1043	1888	èoeba	house, cage
8	1043	1891	èoebě	house, chest, place
9	1043	1891	èoeba	house, chest, place
10	1043	1894	èuba	house
11	1043	1895	èoba ; èobě ; èoeba ; èoebadiê	house
12	1043	1916	èoeba ; èoebě	house, home, chest, suitcase
13	1043	1982	uba	house
14	1043	1987	euba ; euabai	house
15	1043	2011	iub	house
16	1043	2019	yu	house
17	1044	1894	cacario	house
18	1045	<1855	praauw	house
19	1045	1855	praauw	house
20	1045	1864	prahoe	house

Etymologising results: A case for 'house'

17 1044 1894 cacario house

Cognate set (ID 1044)

18 1045 <1855 praauw house

19 1045 1855 praauw house

20 1045 1864 prahoe house

Cognate set (ID 1045)

Challenge 2:

Processing the **complex**
notes on the different
meaning/gloss of a word per
period

Detailed notes about the meaning of word (component) for HOUSE across period.

English gloss	Remarks
house	<p>meaning1888 'house, cage'_meaning1891 'house, chest, place'_meaning1916 'house, home, chest, suitcase'_note<1855: first word means 'ship', second word means 'village', second element in first word unclear but cf. boat_note1855vdS: first word means 'boat', second word means 'village'_note1864: first word means 'house', second word means 'box'_note1864: second element of second word means 'thing'_note1870: first word means 'house', second word means 'sailboat', third word means 'steamboat'_note1879: first word means 'house', second word means 'ship', second word literally means 'house of the sea'_note1895: first word means 'house', second word means 'house', third word means 'proa, canoe, case for cloth', fourth word means 'box, case'_note1987K: first word means 'house, container, nest', second word means 'coffin, closed container'</p>

Write R codes to deal with the Remarks (c. 2 months work between Jan.- Feb. 2024)

```
# A tibble: 20 × 5
  id year words gloss_by_year note_year
<int> <fct> <chr> <chr> <chr>
1 1043 <1855 djoeba[-beri-berri] ; joeba house "djoeba[-beri-berri] means 'ship' ; joeba means...
2 1043 1855 joewa house NA
3 1043 1855 tjoeba ; djoeba house "tjoeba means 'boat' ; djoeba means 'village...
4 1043 1864 oebah ; oebah[-dobah] house "oebah means 'house' ; oebah[-dobah] means 'box...
5 1043 1870 oeba ; eoeba [kiemo] ; oeba [ejobie] house "oeba means 'house' ; eoeba [kiemo] means 'sa...
6 1043 1879 oeba ; eoeba[-kiemo] house "oeba means 'house' ; eoeba[-kiemo] means 's...
7 1043 1888 èoeba house, cage NA
8 1043 1891 èoebě house, chest, place NA
9 1043 1891 èoeba house, chest, place NA
10 1043 1894 èuba house NA
11 1043 1895 èoba ; èobě ; èoeba ; èoebadiě house "èoba means 'house' ; èobě means 'house' ; èoe...
12 1043 1916 èoeba ; èoebě house, home, chest, suitcase NA
13 1043 1982 uba house NA
14 1043 1987 euba ; euabai house "euba means 'house, container, nest' ; euabai ...
15 1043 2011 iub house NA
16 1043 2019 yu house NA
17 1044 1894 cacario house NA
18 1045 <1855 praauw house NA
19 1045 1855 praauw house NA
20 1045 1864 prahoe house NA
```

Write R codes to deal with this (c. 2 months work between Jan.- Feb. 2024)

gloss_by_year

<chr>

house

house

house

house

house

house

house, cage

house, chest, place

house, chest, place

house

house

house, home, chest, suitcase

house

house

house

house

house

house

house

house

note_year

<chr>

"djoeba[-beri-berri] means 'ship' ; joeba means...

NA

"tjoeba means 'boat' ; djoeba means 'village...

"oebah means 'house' ; oebah[-dobah] means 'box...

"oeba means 'house' ; eoeba [kiemo] means 'sa...

"oeba means 'house' ; eoeba[-kiemo] means 's...

NA

NA

NA

NA

"èoba means 'house' ; èobě means 'house' ; èoe...

NA

Tidying up and restructuring remarks on meaning variation of forms per year (gloss_by_year) and meaning of word component (note_year)

Challenge 3:

Standardising
orthography across word
list & IPA mapping

Orthography standardisation

- Use `qlcData` R package (Cysouw 2018) to create a **skeleton of orthography profile** from a word list
 - orthography profile: occurrences of distinct types of characters/letters

Orthography profile skeleton for Brouwer (early 1850s)

Left	Grapheme	Right	Class	Replacement
	oh	\$		o
	oe			u
	oé			u
	ou			o:u
	ie			ii
	ié			ii
	dj			j
	eh	\b		e

Standardised orthography with tokenisation/segmentation

Left	Grapheme	Right	Class	Replacement
	oh	\$		o

1	words	commons	commons_tokenised
11	kokoh	kOkO	k O k o
102	okoha	OkOha	O k O h a

Mapping IPA phoneme with standardised/common orthography



Enggano_Transcription_LongFormat



File Edit View Insert Format Data Tools Extensions Help Accessibility



125%

£

%

.0

.00

123

Defaul...

-

10

+

B

I



E454



fx

Brouwer <1855

	A	B	C	D	E
1	cats	phoneme	grapheme	commons	source
247	vowel	o	oh	o	Brouwer <1855
268	vowel	ɔ	o	O	Brouwer <1855
287	vowel	õ		õ	Brouwer <1855
306	vowel	u	oe	u	Brouwer <1855

Mapping IPA phoneme with standardised/common orthography (Brouwer early 1850s)




1	words	commons	commons_tokenised	ipa	ipa_tokenised
11	kokoh	kOko	k O k o	kɔko	k ɔ k o
102	okoha	OkOha	O k O h a	ɔkɔha	ɔ k ɔ h a

Mapping IPA phoneme with standardised/common orthography (Brouwer early 1850s)

1	words	commons	commons_tokenised	ipa	ipa_tokenised
11	kokoh	kOko	k O k o	kɔko	k ɔ k o
102	okoha	OkOha	O k O h a	ɔkɔha	ɔ k ɔ h a

Segmentised/tokenised sound segments are (i) standard in modern lexical database and (ii) allow quantitative historical language comparison (List & Forkel 2022)

Outlines

1. What are “lexical resources”? 
2. What are the existing linguistic materials for Enggano? 
3. How do we curate these materials into lexical resources? 

Closing

Enggano has rich resources of legacy materials for historical lexical database

Closing

Digitisation and curation processes
involve a range of domain-specific
knowledge:

linguistics
computational/coding
data(base) management

Acknowledgment

UK Research and Innovation (UKRI): Arts and Humanities
Research Council (AHRC) (grant ID: [AH/W007290/1](#))

Udayana University:

Bachelor of English Literature ; Linguistics Postgraduate
Programs (Master & Doctoral) ; Faculty of Humanities

Acknowledgment : mè èm nè'ah Enggano core team & community





Aspects in curating lexical resources for Enggano

Gede Primahadi Wijaya Rajeg

 <https://orcid.org/0000-0002-2047-8621>

Enggano Research Seminar at the Faculty of Humanities, Udayana University (18 March 2024)