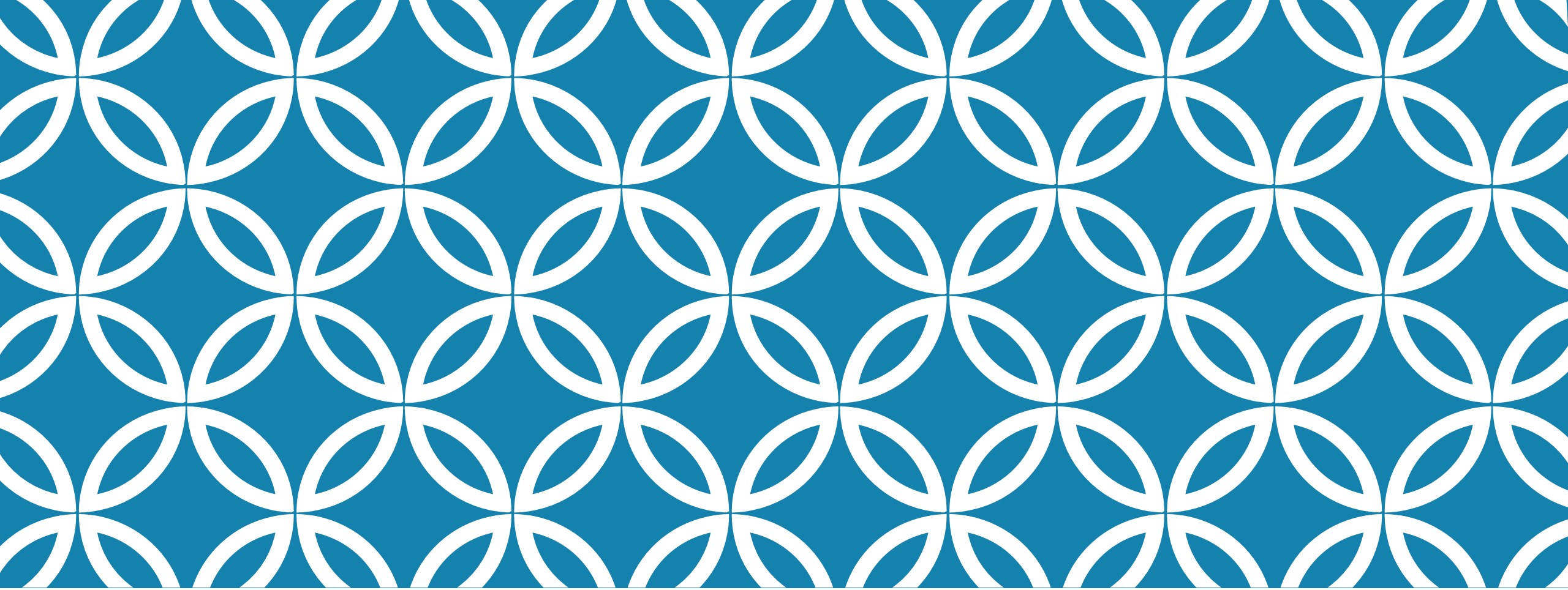


# THESIS SEMINAR MEETING

Philip Georgis

June 7, 2021

# OUTLINE OF PROGRESS

- Data entry from Arabic dataset
- Challenges with Turkic dataset
- New dataset: Polynesian languages
- Index of all language varieties
- New GitHub repository

# ARABIC DATASET (RATCLIFFE, 2020)

- Recall: semi-manual data entry/cleaning required due to source formatting

1.1 Data

| | | CA | Mor | Mlt | Cai | Dms | Irq | Skh | AqArb | Cyp | Glf | Ymn | Nig | Bux | Nub |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ALL | kull | koll | Kollu | kull | kəll | kull | ʧill | kəll, sa:ɣi:n | kull | kil | kull | ʧat | kullu | kulu |
| 2 | ASH | rama:d | rˁmadˁ | rmi:d | rama:dˁ* | rama:d | ruma:d | 1 | rˁama:d | ramat r–m–d | ramaad | 1 | 1 | ra'matt | ruman |
| 3 | BARK | qirfat | qeʃrˁa | ʔoʃra | ʔiʃr* | ʔəʃər | giʃra | kiʃr | səvi:ye | 1 | giʃra | giʃr | li:he, girfe | 1 | *kokobo lataka, (girifa) |
| 4 | BELLY | batˁn | kerʃ | zaʔʔ | batˁn* | batˁən | batˁin | batˁn, ʧarʃ | ʤoof | patn b–tˁ–n | batˁn | batˁn, karʃ | kirʃ | batin | batna |
| 5 | BIG | kabi:r | kbir | kibir | kibiir | kbi:r | ʧbi:r | ʧabi:r | gəbi:r | k–b–r | ʕood | Kabiir | kabi:r | ka'biir | kbir |
| 6 | BIRD | tˁa:ʔir | tˁir | ʕasfur | tˁeer* | tˁeer | tˁeer | tˁeer | ku:ʧka:ye | 2 | tˁeer | tˁajr | tˁe:ra | tayra | ter |
| 7 | BITE | ʕad | ʕed | gide | ʕadˁ | ʕadˁd | ʕaðˁð | ʕaðˁð | 1 | ʕaðð | ʕadˁd | lugus | adˁdˁa | yaʕazz | adi |

# ARABIC DATASET (RATCLIFFE, 2020)

- Recall: semi-manual data entry/cleaning required due to source formatting

1.1 Data

| | | CA | Mor | Mlt | Cai | Dms | Irq | Skh | AqArb | Cyp | Glf | Ymn | Nig | Bux | Nub |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ALL | kull | koll | Kollu | kull | kəll | kull | tʃill | kall | kull | kil | kull | tʃat | kullu | kulu |
| 2 | ASH | ramad | rˁmadˁ | rmi:d | ramaːdˁ* | rama:d | ruma:d | | | | | | | | |
| 3 | BARK | qirfat | qeʃrˁa | ʔoʃra | ʔiʃr* | ʔəʃər | giʃra | | | | | | | | |
| 4 | BELLY | batˁn | kerʃ | zaʔʔ | batˁn* | batˁən | batˁin | | | | | | | | |
| 5 | BIG | kabi:r | kbir | kibir | kibiir | kbi:r | tʃbi:r | | | | | | | | |
| 6 | BIRD | tˁa:ʔir | tˁir | ʕasfur | tˁeer* | tˁeer | tˁeer | | | | | | | | |
| 7 | BITE | ʕad | ʕed | gide | ʕadˁ | ʕadˁd | ʕaðˁð | | | | | | | | |

| Index | Gloss | CA | Mor | Mlt | Cai | Dms | Irq | Skh | AqArb | Cyp | Glf | Ymn | Nig | Bux | Nub |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 40 | HEART | qalb | qelb | ʔalb | ʔalb | ʔalb | qalb | kalˁb | faːd | kilep | galb | galb | galib | qalb | gelba |
| 41 | HORN | qarn | qerˁn | ʔarn | ʔarn | ʔarən | girin | girin | | korne | garn | garn* | garn | ʃɔx | *gurun |
| 42 | I | ʔana: | ana | jiːn | ana | ʔana | ʔaːni | ana | ana | ʔ-n-a | ana | ʔana | ana | anaa | ʻana |
| 43 | KILL | qatala | qtel | ʔatel | mawwit | ʔatal | kital | katal | qatal | q-t-l | gital | gatal | katal | qa'tal | katul |
| 44 | KNEE | rukbat | rokba | rkobba | rukba* | rəkbe | rukba | rˁutʃbi | | | rikba | rukbih* | rukuba | | *rukuba |
| 45 | KNOW | ʕarafa | ʕrˁef | jaf | ʕirif | ʕəref | ʕiraf | ʕarˁaf | ʕərəf | ʕ-r-f | ʕaraf | diriʕaraf | ʕirif | ʔaraf | ʻarufu |
| | | | | | | | | | | | dara | | | | |
| 46 | LEAF | waraqat | werqa | werʔa | warʔa* | waraʔa | warga | waraka | waraqa | | wriga | waragih* | waraga | varaq | korofo |
| 47 | LIE | raqada | ttmed | mtedd | itmaddad* | tlaʔʔah | tmaddad | ttsallas | | | tamaddad | ragad* | ragad | daaʕa | |
| 48 | LIVER | kabid | kebda | fwiːd | kibda* | kəbəd | kabda | | | q-sˁ-b | tʃibd | kabdih* | kibde | kabda | *koobda |
| 49 | LONG | tˁawiːl | tˁwil | twil | tˁawiːl | tˁawiːl | tˁuwiːl | tˁawiːl | tˁ-w-l | tˁ-w-l | tˁawiil | tˁawiːl | tˁawil | ta'viil | towil |
| 50 | LOUSE | qamlah | qemla | ʔamla | ʔamla* | ʔamle | gamlˁa | kamlˁ | | q-m-l | gamla | gamleh | | qamla | |
| 51 | MAN | radʒul | rˁaʒel | radʒel | rˁaːgil | rədʒdʒaːl | ridʒdʒaːl | zalami | radʒəl | Intsan | rayyaal | radʒdʒaːl | radʒil | ʻzaaker | ragi |
| | | | | | | | | | | | | | | ra'ʒul | |
| 52 | MANY | kaθiːr | bezzaf | ħafna | kitiir | ktiːr | hwaːja | tʃθiːr | kətiːr | xtir | waadʒid | kaθiːr | katiːr | ka'siir | *milan |
| 53 | MEAT | laħm | lħam | laħam | laħma | (laħam) | laħam | laħm | laħəm | l-ħ-m | laħam | laħm | laːma | læħəm | laːm |
| 54 | MOON | qamar | gamrˁa | ʔamar | ʔamar | ʔamarˁ | gumar | kamarˁ | | q-m-r | gamar | gamar | gamar | qamar | *ʃa(h)ari |
| 55 | MOUNTAIN | dʒabal | ʒbel | muntanja | gabal* | dʒabal | dʒibal | tsabal | dʒabal | ʒipel | dʒibal | dʒabal* | hadʒar | dʒabal | dʒebel |
| 56 | MOUTH | fam | fomm | ħalʔ | buʔʔ | təmm | ħaliq | ħalk | təmm | *xankħ-n-k | ħaldʒ | fumm* | gaddu:ma | bæb | *kasma |
| | | | | | | | | bu:z | | | buuz | | | | |
| 57 | NAME | ism | smija | isem | ism | ʔəsəm | ʔisim | ism | əsəm | ʔ-s-m | isim | ʔism | isim | ʔisim | *asma |
| 58 | NECK | ʕunq | ʕenq | ʕonʔ | raʔaba* | raʔbe | rugba | rakaba | | r-q-b | rguba | rugbeh | ragaba | | *ragabtu |
| 59 | NEW | dʒadiːd | ʒdid | dʒdida | gidiid | dʒdiːd | dʒidiːd | tsadiːd | dʒədiːd | ʒtitdʒ-d-d | dʒadiid | dʒadiːd | dʒadiːd | dʒadiid | dʒedid |

# ARABIC DATASET (RATCLIFFE, 2020)

- Recall: semi-manual data entry/cleaning required due to source formatting

1.1 Data

| | | CA | Mor | Mlt | Cai | Dms | Irq | Skh | AqArb | Cyp | Glf | Ymn | Nig | Bux | Nub |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ALL | kull | koll | Kollu | kull | kəll | kull | tʃill | kəll. | kull | kil | kull | tʃat | kullu | kulu |
| 2 | ASH | rama:d | rˁmadˁ | rmi:d | rama:dˁ* | rama:d | ruma:d | | | | | | | | |
| 3 | BARK | qirfat | qeʃrˁa | ʔoʃra | ʔiʃr* | ʔəʃər | giʃra | | | | | | | | |
| 4 | BELLY | batˁn | kerʃ | zaʔʔ | batˁn* | batˁən | batˁin | | | | | | | | |
| 5 | BIG | kabi:r | kbir | kibir | kibiir | kbi:r | tʃbi:r | | | | | | | | |
| 6 | BIRD | tˁa:ʔir | tˁir | ʕasfur | tˁeer* | tˁeer | tˁeer | | | | | | | | |
| 7 | BITE | ʕad | ʕed | gide | ʕadˁ | ʕadˁd | ʕaðˁð | | | | | | | | |

| Index | Gloss | CA | Mor | Mlt | Cai | Dms | Irq |
|---|---|---|---|---|---|---|---|
| 40 | HEART | qalb | qelb | ʔalb | ʔalb | ʔalb | qalb |
| 41 | HORN | qarn | qerˁn | ʔarn | ʔarn | ʔarən | girin |
| 42 | I | ʔana: | ana | ji:n | ana | ʔana | ʔa:ni |
| 43 | KILL | qatala | qtel | ʔatel | mawwit | ʔatal | kital |
| 44 | KNEE | rukbat | rokba | rkobba | rukba* | rəkbe | rukb |
| 45 | KNOW | ʕarafa | ʕrˁef | jaf | ʕirif | ʕəref | ʕiraf |
| | | | | | | | |
| 46 | LEAF | waraqat | werqa | werʔa | warʔa* | waraʔa | warg |
| 47 | LIE | raqada | ttmed | mtedd | itmaddad* | tlaʔʔah | tmad |
| 48 | LIVER | kabid | kebda | fwi:d | kibda* | kəbəd | kabd |
| 49 | LONG | tˁawi:l | tˁwil | twil | tˁawi:l | tˁawi:l | tˁuw |
| 50 | LOUSE | qamlah | qemla | ʔamla | ʔamla* | ʔamle | gaml |
| 51 | MAN | radʒul | rˁaʒel | radʒel | rˁa:gil | rədʒdʒa:l | ridʒd |
| | | | | | | | |
| 52 | MANY | kaθi:r | bezzaf | ħafna | kitiir | kti:r | hwa: |
| 53 | MEAT | laħm | lħam | laħam | laħma | (laħam) | laħa |
| 54 | MOON | qamar | gamrˁa | ʔamar | ʔamar | ʔamarˁ | gum |
| 55 | MOUNTAIN | dʒabal | ʒbel | muntanja | gabal* | dʒabal | dʒiba |
| 56 | MOUTH | fam | fomm | ħalʔ | buʔʔ | təmm | ħaliq |
| | | | | | | | |
| 57 | NAME | ism | smija | isem | ism | ʔəsəm | ʔisim |
| 58 | NECK | ʕunq | ʕenq | ʕonʔ | raʔaba* | raʔbe | rugb |
| 59 | NEW | dʒadi:d | ʒdid | dʒdida | gidiid | dʒdi:d | dʒidi |

Cai.txt

```
ALL/kull/kull
ASH/rama:dˁ/rama:dˁ
BARK/ʔiʃr*/ʔiʃr
BELLY/batˁn*/batˁn
BIG/kibiir/kibi:r
BIRD/tˁeer*/tˁe:r
BITE /ʕadˁdˁ/ʕadˁdˁ
BLACK/iswid/iswid
BLOOD/damm/damm
BONE/ʕadma/ʕadma
BREAST/sidir*/sidir
BURN/ħaraʔ/ħaraʔ
CLAW/dˁifrˁ*/dˁifrˁ
CLOUD/saħaab/saħa:b
COLD/bard/bard
COME/ga/ga
DIE/maat/ma:t
DOG/kalb/kalb
DRINK/ʃirib/ʃirib
DRY/na:ʃif*/na:ʃif
EAR/widn/widn
EARTH,DIRT/ardˁ/ardˁ
EAT/akal/akal
EGG/beedˁa/be:dˁa
EYE/ʕeen/ʕe:n
FAT,N./ʃaħm/ʃaħm
FEATHER/ri:ʃa*/ri:ʃa
FIRE/na:r/na:r
FISH/samak/samak
FLY/tˁa:rˁ/tˁa:rˁ
```

# ARABIC DATASET (RATCLIFFE, 2020)

- Will likely need to exclude some of the (smaller) varieties from comparison, e.g.
  - Cypriot Arabic
    - Much data missing, most only in form of triconsonantal roots vs. full word forms in other lects
  - Soukhne Arabic (rural Syrian dialect from between Damascus and Baghdad)
    - Cannot find a (unique) Glottolog entry
    - Unclear whether it is a Levantine or Eastern Arabic variety
    - Found ASJP wordlist for "Soukhne Syrian Arabic", but cross-referenced same as Iraqi Arabic (https://asjp.clld.org/languages/SOUKHNE_SYRIAN_ARABIC)

# TURKIC DATASET: CHALLENGES

- Mostly pre-processed and loaded the Turkic dataset (Savelyev & Robeets, 2020)

| | A | B | W | X | Y | Z | AA | AB | AC | AD | AE | AF | AG | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | **Meaning** | **Root** | **Gagauz** | | **KaraKalpak** | | **KarachayBalka** | | **Karaim** | | **Kazakh** | | **Khakas** | |
| 2 | | | **Standard romar** | **IPA** | **Standard romar** | **IPA** | **Standard romar** | **IPA** | **Standard romar** | **IPA** | **Standard romar** | **IPA** | **Standard romar** | **IPA** |
| 3 | **fire (n.)** | *o:t | | | ot | ot | ot | ot | ot | ot | ot | ot | ot | ot |
| 4 | | *kaŋ- | | | | | | | | | | | | |
| 5 | **nose (n.)** | *burUn | burnu | burnu | murïn | muruun | burun | burun | burun | burun | murïn | muruun | purun | purun |
| 6 | | *tumčuk | | | | | | | | | | | | |
| 7 | | *bar- | | | ? bar | ? bar | bar | bar | bar | bar | bar | bar | par | par |
| 8 | **go (v.)** | *jorï- / *jör(i)- | | | | | | | | | | | | |
| 9 | | *kejt- | git | git | ? ket | ? ket | | | ket | ket | ket | ket | | |
| 10 | **water (n.)** | *sɨb | su | su | suw | suw | suw | suw | suv | suv | su | su | suɣ | suɣ |
| 11 | | *agɨŕ | a:z | a:z | awïz | awuuz | awuz | awuz | avuz | avuz | awïz | awuuz | aas | aas |
| 12 | **mouth (n.)** | *ańak | | | | | | | | | | | | |
| 13 | | *jü:ŕ | | | | | | | | | | | | |
| 14 | **tongue (n.)** | *tIl | dil | dil | til | til | til | til | til | til | tïl | təl | təl | təl |
| 15 | **blood (n.)** | *kia:n | kan | kan | ḳan | qan | ḳan | qan | kan, (k) ḳan | kan, (k) qan | ḳan | qan | χan | χan |
| 16 | **bone (n.)** | *siŋök | | | süyek | syjek | süyek | syjek | sivek, (k) süyek | sivek, (k) syjek | süyek | syjek | söök | søøk |
| 17 | | *kemük | kemik | kemik | | | | | | | | | | |
| 18 | **2SG pronoun** | *sę | | | | | | | | | | | | |
| 19 | | *sɐn | sän | sɐn | sen | sen | sen | sen | sen | sen | sen | sen | sin | sin |
| 20 | | *tąmor | | | tamïr | tamuur | tamïr | tamuur | tamur | tamur | tamïr | tamuur | | |
| 21 | | *jïldïŕ | | | | | | | | | | | | |
| 22 | **root (n.)** | *taŕɨl = *tasɨl | | | | | | | | | | | | |
| 23 | | *kök | kök | køk | | | | | | | | | | |
| 24 | | *tüp | | | | | | | | | | | | |
| 25 | | *ö:ŕ | | | | | | | | | | | | |
| 26 | **come (v.)** | *gęl | gel | gel | kel | kel | kel | kel | kel | kel | kel | kel | kil | kil |
| 27 | **breast (n.) // (chest) (n.)** | *köküR(ek) | gü:s | gy:s | kökürek | køkyrek | kökürek | køkyrek | kökräk, kekrek, k | køkrɛk, kekrek, k | kökrek | køkrek | kögɔs, köksɔ | køgɔs, |
| 28 | | *tiö:Ĺ | | | ? tös | ? tøs | töš | tøʃ | töš, tes | tøʃ, tes | tös | tøs | | |
| 29 | | *kebde | | | gewde | gewde | | | | | kewde | kewde | | |
| 30 | | *jag- | ya:mur | ja:mur | žawïn, (žamɣïr) | ʒawuun, (ʒamɣuur | ǰaŋur, (Balk.) ǰaw | ʤaŋur, (Balk.) ʤ | yamɣur | jamɣur | žaŋbïr, žawïn | ʒaŋbuur, ʒawuun | naŋmïr | naŋmu |

# TURKIC DATASET: CHALLENGES

- Possible issues with some of the IPA transcriptions
  - Orthographic symbols sometimes used in place of IPA characters

    → Possible to replace wherever clear what the intended IPA symbols were

    e.g. <š> = /ʃ/, <ɯ> = /ɯ/, or Chuvash <ö> = /əʷ/

    …but some are less clear, and resources may not exist to disambiguate.

  - Turkish transcriptions call into question the validity/quality of transcriptions
    - Transcribes phonemes which do not exist in standard Turkish, e.g. /q/, /ɣ/
    - Some words are mistaken: e.g. **üštünde* /yʃtynde/ 'above'
    - → Maybe some non-standard dialect of Turkish?
    - Could correct the issues in Turkish, but I don't know enough about other varieties to check/correct them

  - Annotations within transcriptions: need to spend more time to sort these out

# TURKIC DATASET: CHALLENGES

- Could alternatively use a different dataset for Turkic:
  - **NorthEuraLex**
    - Only includes 8 Turkic languages
    - Turkish, Azeri, Uzbek, Kazakh, Bashkir, Tatar, Sakha, Chuvash

  - **"Culture Words for Turkic"** from Diachronic Atlas of Comparative Linguistics (DiACL)
    - Only includes 9 Turkic languages
    - Turkish, Gagauz, Turkmen, Azeri, Crimean Tatar, Kazakh, Kumyk, Tuvan, Southern Altai
    - Not a Swadesh list

  … vs. 31 Turkic languages in Savelyev & Robeets (2020) dataset, which also includes cognate coding.

  → would strongly prefer to find solution with this dataset

# POLYNESIAN DATASET (WALWORTH, 2018)

- One final dataset including 30 modern Polynesian languages + Proto-Polynesian reconstructed forms
- Source is well formatted/machine readable, no issues in IPA transcriptions, includes Glottolog and ISO references, cognate coding
- No major issues found → fully preprocessed

# OVERVIEW OF DATASETS

| Family | Source Name | Reference | Number of Varieties |
|---|---|---|---|
| Arabic | Glottometrics of Arabic | Ratcliffe (2020) | 14 |
| Italic | Global Lexicostatistic Database | Saenko (2016) | 58 |
| (Balto-)Slavic | NorthEuraLex | Dellert et al. (2019) | 9 (+2 Baltic) |
| Uralic | NorthEuraLex | Dellert et al. (2019) | 26 |
| Polynesian | Polynesian Segmented Data | Walworth (2018) | 31 |
| Sinitic | Collection of basic vocabulary words and characteristic dialect words in modern Chinese dialects | Líu et al. (2007) | 19 |
| Turkic | Basic vocabulary datasets for the Turkic languages | Savelyev & Robbeets (2020) | 31 |

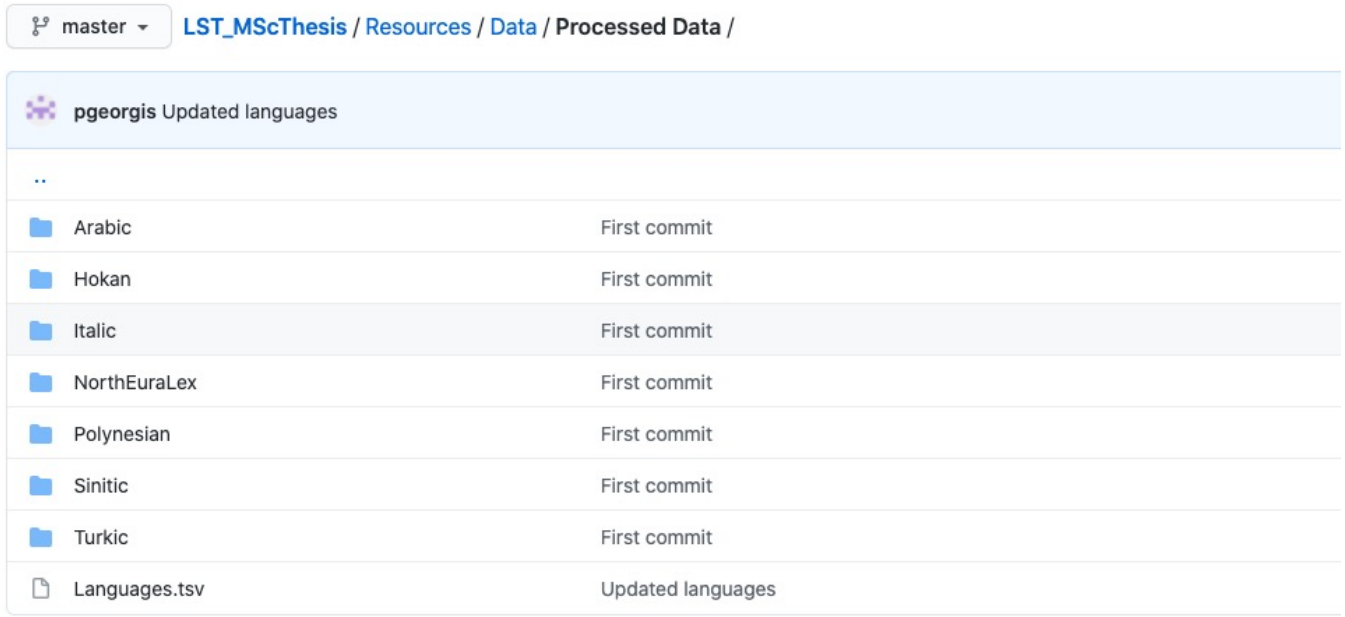
# INDEX OF LANGUAGE VARIETIES

- Cross-referenced all varieties from datasets with Glottolog and ISO codes in order to be able to compare results with gold standard trees
- Some varieties could not be found → exclude from study (e.g. specific Italic dialects)

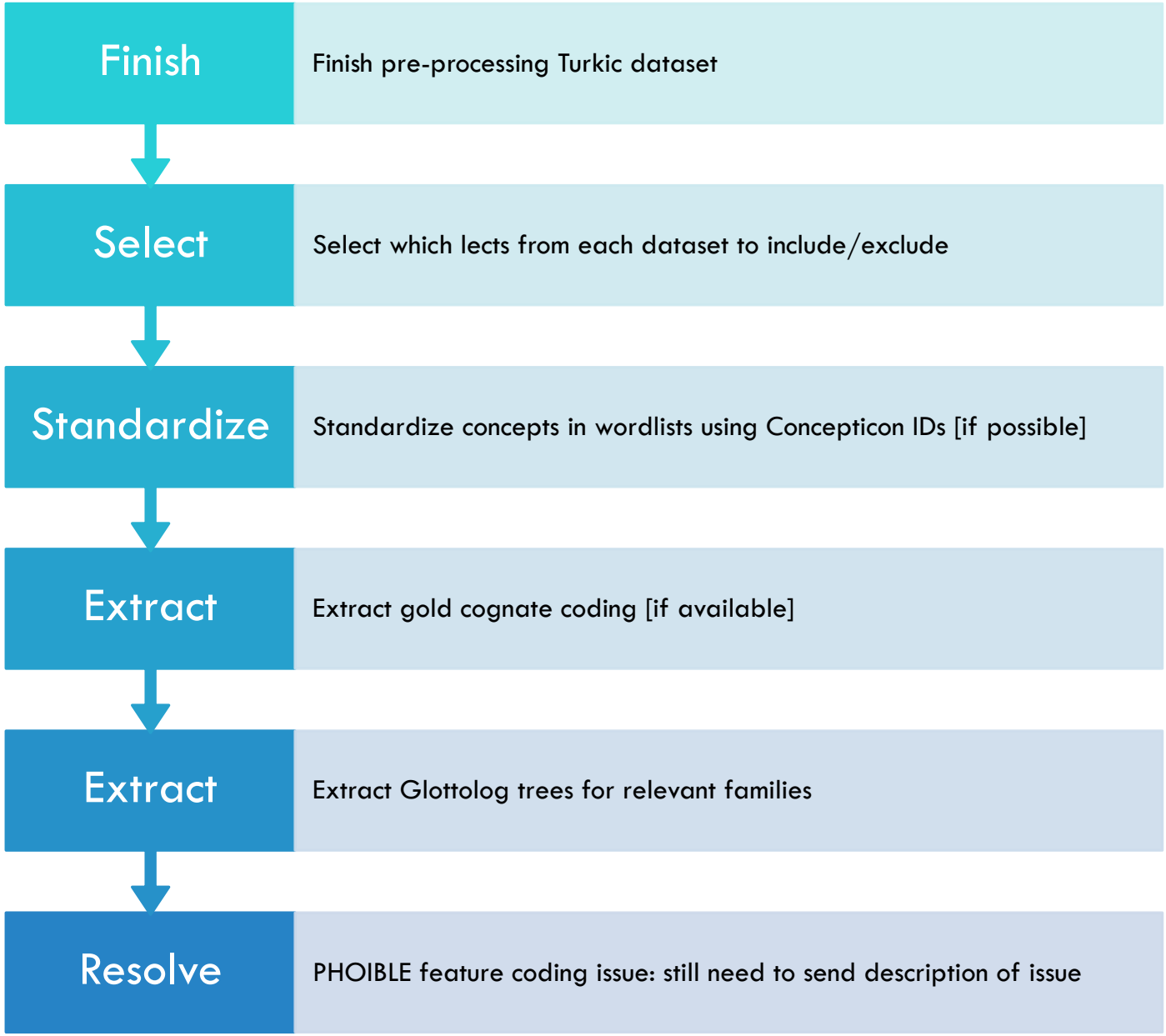
| 1 | Dataset | Family | Name | Source Name | Glottolog Name | Glottocode | ISO 639-3 | ISO 639-3 Name | Reference | Notes |
|---|---|---|---|---|---|---|---|---|---|---|
| 171 | NorthEuraLex | Uralic, Finnic | Estonian | Estonian | Estonian | esto1258 | ekk | Standard Estonian | Dellert et al., 2019 | |
| 172 | NorthEuraLex | Uralic, Finnic | Finnish | Finnish | Finnish | finn1318 | fin | Finnish | Dellert et al., 2019 | |
| 173 | NorthEuraLex | Uralic, Finnic | Livonian | Livonian | Liv | livv1244 | liv | Liv | Dellert et al., 2019 | |
| 174 | NorthEuraLex | Uralic, Finnic | North Karelian | North Karelian | Karelian | kare1335 | krl | Karelian | Dellert et al., 2019 | |
| 175 | NorthEuraLex | Uralic, Finnic | Olonets Karelian | Olonets Karelian | Livvi | livv1243 | olo | Livvi | Dellert et al., 2019 | |
| 176 | NorthEuraLex | Uralic, Finnic | Veps | Veps | Veps | veps1250 | vep | Veps | Dellert et al., 2019 | |
| 177 | NorthEuraLex | Uralic, Hungarian | Hungarian | Hungarian | Hungarian | hung1274 | hun | Hungarian | Dellert et al., 2019 | |
| 178 | NorthEuraLex | Uralic, Khantyic | Northern Khanty | Northern Khanty | Kazym-Berezover-Suryskarer Khanty | khan1273 | kca | Khanty | Dellert et al., 2019 | |
| 179 | NorthEuraLex | Uralic, Mansic | Northern Mansi | Northern Mansi | Northern Mansi | mans1258 | mns | Mansi | Dellert et al., 2019 | |
| 180 | NorthEuraLex | Uralic, Mari | Hill Mari | Hill Mari | Western Mari | west2392 | mrj | Western Mari | Dellert et al., 2019 | |
| 181 | NorthEuraLex | Uralic, Mari | Meadow Mari | Meadow Mari | Eastern Mari | east2328 | mhr | Eastern Mari | Dellert et al., 2019 | |
| 182 | NorthEuraLex | Uralic, Mordvin | Erzya | Erzya | Erzya | erzy1239 | myv | Erzya | Dellert et al., 2019 | |
| 183 | NorthEuraLex | Uralic, Mordvin | Moksha | Moksha | Moksha | moks1248 | mdf | Moksha | Dellert et al., 2019 | |
| 184 | NorthEuraLex | Uralic, Permian | Komi-Permyak | Komi-Permyak | Komi-Permyak | komi1269 | koi | Komi-Permyak | Dellert et al., 2019 | |
| 185 | NorthEuraLex | Uralic, Permian | Komi-Zyrian | Komi-Zyrian | Komi-Zyrian | komi1268 | kpv | Komi-Zyrian | Dellert et al., 2019 | |
| 186 | NorthEuraLex | Uralic, Permian | Udmurt | Udmurt | Udmurt | udmu1245 | udm | Udmurt | Dellert et al., 2019 | |
| 187 | NorthEuraLex | Uralic, Saami | Inari Sami | Inari Sami | Inari Saami | inar1241 | smn | Inari Sami | Dellert et al., 2019 | |
| 188 | NorthEuraLex | Uralic, Saami | Kildin Sami | Kildin Sami | Kildin Saami | kild1236 | sjd | Kildin Sami | Dellert et al., 2019 | |
| 189 | NorthEuraLex | Uralic, Saami | Lule Sami | Lule Sami | Lule Saami | lule1254 | smj | Lule Sami | Dellert et al., 2019 | |
| 190 | NorthEuraLex | Uralic, Saami | Northern Sami | Northern Sami | North Saami | nort2671 | sme | Northern Sami | Dellert et al., 2019 | |
| 191 | NorthEuraLex | Uralic, Saami | Skolt Sami | Skolt Sami | Skolt Saami | skol1241 | sms | Skolt Sami | Dellert et al., 2019 | |
| 192 | NorthEuraLex | Uralic, Saami | Southern Sami | Southern Sami | South Saami | sout2674 | sma | Southern Sami | Dellert et al., 2019 | |
| 193 | NorthEuraLex | Uralic, Samoyedic | Forest Enets | Forest Enets | Forest Enets | fore1265 | enf | Forest Enets | Dellert et al., 2019 | |
| 194 | NorthEuraLex | Uralic, Samoyedic | Nganasan | Nganasan | Nganasan | ngan1291 | nio | Nganasan | Dellert et al., 2019 | |
| 195 | NorthEuraLex | Uralic, Samoyedic | Northern Selkup | Northern Selkup | Selkup | selk1253 | sel | Selkup | Dellert et al., 2019 | |
| 196 | NorthEuraLex | Uralic, Samoyedic | Tundra Nenets | Tundra Nenets | Tundra Nenets | nene1249 | yrk | Nenets | Dellert et al., 2019 | |
| 197 | NorthEuraLex | Yeniseian, Northern Yeniseian | Ket | Ket | Ket | kett1243 | ket | Ket | Dellert et al., 2019 | |
| 198 | NorthEuraLex | Yukaghir, Kolymic | Southern Yukaghir | Southern Yukaghir | Southern Yukaghir | sout2750 | yux | Southern Yukaghir | Dellert et al., 2019 | |
| 199 | NorthEuraLex | Yukaghir, Northern Yukaghir | Northern Yukaghir | Northern Yukaghir | Northern Yukaghir | nort2745 | ykg | Northern Yukaghir | Dellert et al., 2019 | |
| 200 | Polynesian | Austronesian, Polynesian | Proto-Polynesian | Polynesian | Polynesian | poly1242 | | | Walworth, 2018 | |
| 201 | Polynesian | Austronesian, Polynesian, Anuta | Anuta | Anuta | Anuta | anut1237 | aud | Anuta | Walworth, 2018 | |
| 202 | Polynesian | Austronesian, Polynesian, Carolinean Outlier Polynesian | Kapingamarangi | Kapingamarangi | Kapingamarangi | kapi1249 | kpg | Kapingamarangi | Walworth, 2018 | |
| 203 | Polynesian | Austronesian, Polynesian, East Futuna | East Futuna | EastFutuna | East Futuna | east2447 | fud | East Futuna | Walworth, 2018 | |
| 204 | Polynesian | Austronesian, Polynesian, East Polynesian | Austral A | AustralA | Austral | aust1304 | aut | Austral | Walworth, 2018 | |
| 205 | Polynesian | Austronesian, Polynesian, East Polynesian | Austral B | AustralB | Austral | aust1304 | aut | Austral | Walworth, 2018 | |

# NEW GITHUB REPOSITORY

- Created new GitHub repository under my personal account: pgeorgis/LST_MScThesis
- Currently private, but I can grant access
- Will use to store update presentations, processed data, future scripts, etc.



# NEXT STEPS


Finish
Finish pre-processing Turkic dataset
Select
Select which lects from each dataset to include/exclude
Standardize
Standardize concepts in wordlists using Concepticon IDs [if possible]
Extract
Extract gold cognate coding [if available]
Extract
Extract Glottolog trees for relevant families
Resolve
PHOIBLE feature coding issue: still need to send description of issue