$\mu$  (pl.  $tezz\hat{a}het + \# : + ) \parallel neuf \parallel voir l'emploi des noms de nombre card. à : OE <math>kerad$ .

ΤE

- $\square \exists$  ețțeb  $\parallel$  v.  $E \square$  ețțeb.
- □∃ toubet (Ta. 2) | v. E□ toubet (Ta. 2).
- outtib  $\mathbb{E}\square * \mathrm{sm.}$  (pl.  $outtiben \mathbb{E}\square I$ ) || brique || se dit de briques de n'importe quelle espèce || syn. d' $\check{a}beller$  empl. d. ce s.
- gros tambour || l'ettebel est un gros tambour demi-sphérique qui n'a de peau que sur une face; il diffère du ganga « petit tambour plat ». Les principaux chefs touaregs, les ămenoûkal surtout, en possèdent de très

## Working without funding – a case study

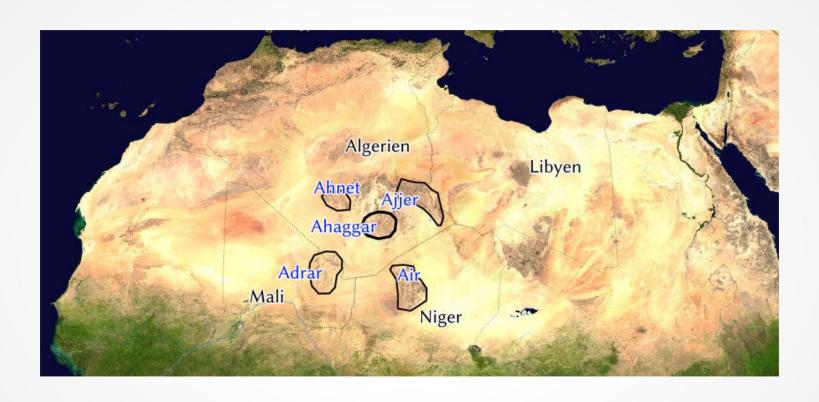
### Julian Jarosch

Academy of Sciences and Literature Johannes Gutenberg University Mainz

### Tuareg

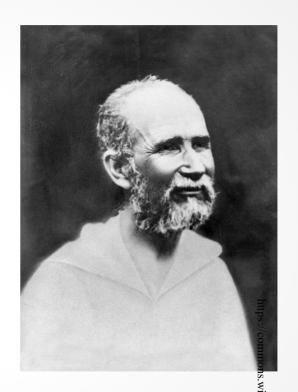
- Tamahaq (exonym: Tuareg)
  - southern Berber
    - Afroasiatic family
- 77,000 speakers (Ethnologue)
- present-day southern Algeria and extreme south-west Libya

# Geography



### Charles de Foucauld

- French military, explorer, and then monk
- lived in the Ahaggar for ca.11 years (1905–1916)
- Acquired fluent competence of Tamahaq



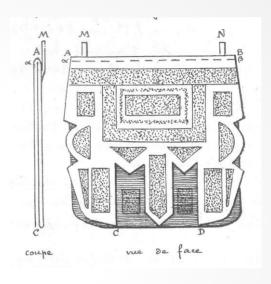
## Lexicographic works by Foucauld

- Dictionnaire abrégé touaregfrançais
  - completed 1913
  - published 1918–1920 (typeset)
- Dictionnaire abrégé touaregfrançais de noms propres
  - completed 1914
  - published 1940 (typeset)
- Dictionnaire touareg –



## Dictionnaire touareg – français

- 2028 manuscript pages
- unsurpassed milestone in Tuareg linguistics
- detailed definitions, verb inflection paradigms, and encyclopedic descriptions and illustrations



### Outline

- Manual Transcription in Wikisource
- Handwritten Text Recognition in Transkribus
- Georeferencing using GeoNames and OpenStreetMap

# Manual Transcription

### Wikisource

- fr.wikisource.org
- full text transcriptions of published, public domain works
- run by Wikimedia Foundation
  - i.e. sister project of Wikipedia
- note: language versions are semi-independent and apply different rules

### Backend



Accueil
Index des auteurs
Portails thématiques
Aide au lecteur
Contacter Wikisource
Texte au hasard

Auteur au hasard

Contribuer

Scriptorium Forum des nouveaux

Communauté Livre au hasard Modifications récentes Faire un don

Imprimer / exporter

Télécharger en EPUB Télécharger en MOBI Télécharger en PDF Autres formats Citer ce texte

Outils

Pages liées
Suivi des pages liées
Téléverser un fichier
Pages spéciales
Version imprimable
Lien permanent
Informations sur la
page
Citer cet article
Sous-pages

Ò

Langues

#### Livre:Foucauld, Dictionnaire touareg.djvu

Avancement

Titre Dictionnaire touareg - français =

Sous-titre Dialecte de l'Ăhaggar

Auteur Charles de Foucauld =

Éditeur André Basset

Maison d'édition Imprimerie nationale de France

Lieu d'édition Paris

Bibliothèque Universitätsbilbiothek Mainz (bibliothèque universitaire de Mayence, ISIL DE-77)

À corriger

Fac-similés djvu

#### Pages

Index Discussion Styles

towning - français

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486

#### Table des matières

Lire Modifier Voir l'historique | ★ Plus V Rechercher sur Wikisource

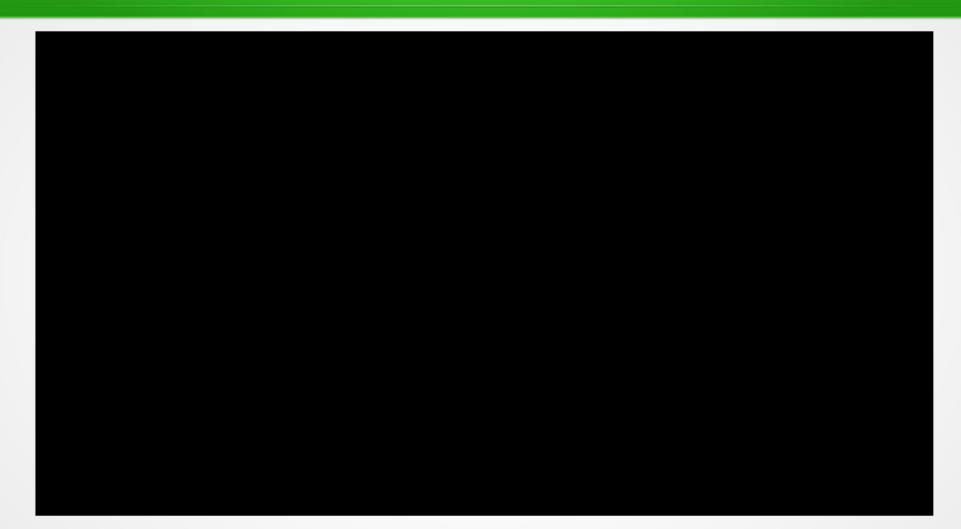
Discussion Préférences Bêta Liste de suivi Contributions Se déconnecter

Q

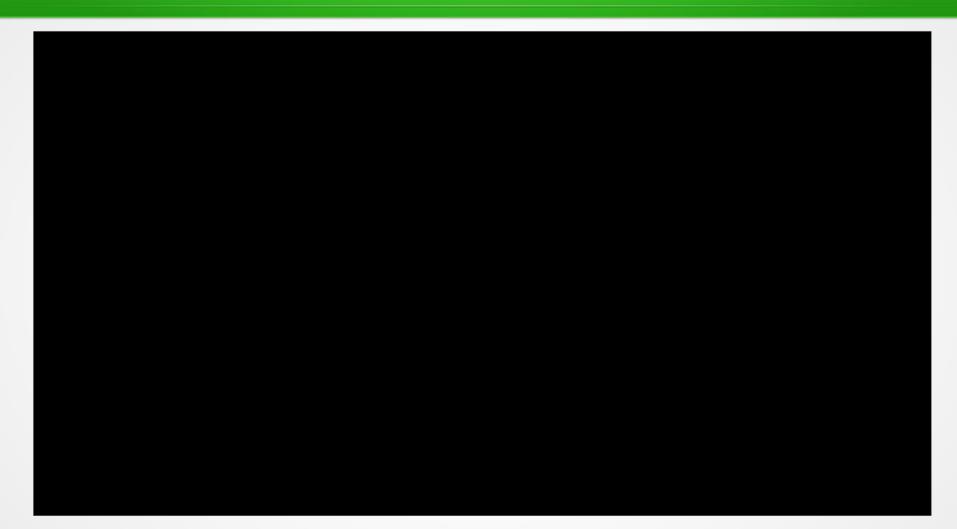
O

# Avertissement.	page 2
Dictionnaire.	
■ B ⊖	
■ CH G	120
□ D Λ	
■ D E	251
a F X	294
■ G X, Ġ T	374
₀ H i	495
ω Ι Σ	
⊞ J X	
■ K :-	714
⊞ K	945
■ K ::	947
. Ё И	971
⊞ M C	1.131
⊞ N I. Ñ ≢	1.279
• OU 8	1.440
# RO	1.547
⊞ R :	1.683
# S O	1.794
# T +	1.876
⊞ T E	
a Z X. Ż X	1.925
Appendice. Conjugaison des verbes.	
I. Verbes réguliers.	1.999
II. Verbes irréguliers.	
Table des matières.	2.028

# **Basic Transcription**



# Adding Markup



### Frontend

Charles de Foucauld
Dictionnaire touareg - français
Dialecte de l'Ăhaggar
Texte établi par André Basset, Imprimerie nationale de France.

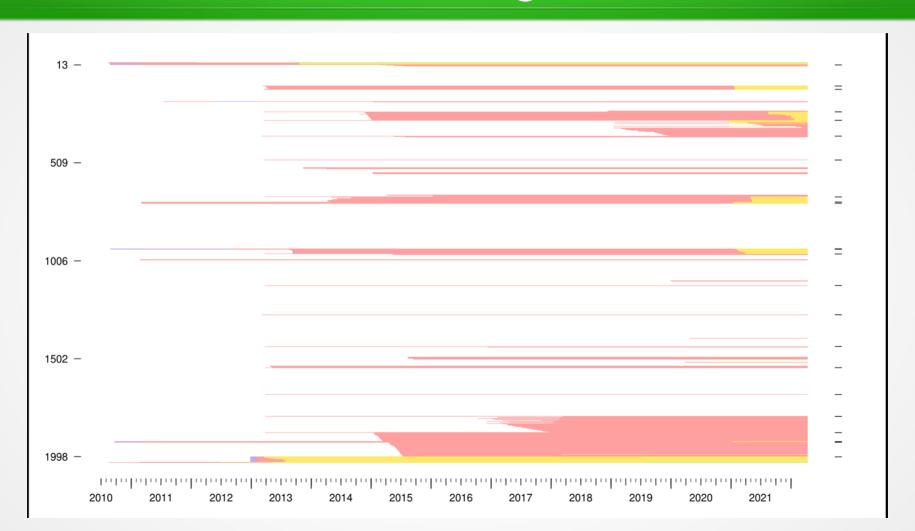
#### Table des matières

Avertissement.  Dictionnaire.	page 2
	13
	120
	137
	251
	294
	374
	495
	682
	945
	971
	1.131
	1.440
	1.547
	1.683
	1.794
	1.876
	1.922
<b>⋴</b> zж, ҳж	1.925
■ Appendice. Conjugaison	
I. Verbes réguliers	1.999
II. Verbes irréguliers.	2.025
* Table des matières	2.028
~~~~~	~~~~~~~~
Rec	hercher dans ce livre

- taftek -+H+ sf. nv. prim.; φ (pl. tifettår :+H+), dar tëftek (tåftek), dar tfettår || m. s. q. le pr. || p. ext. « tapis (tapis proprement dit, couverture, tissu, natte, peau, paille, ch. qlconque qu'on étend pour servir de tapis) ». D. ce s., est syn. d'åsefter et plus us, que lui.
- ăsefter :+ HO sm. nv. f. 1; φ (pl. iseftîren I:+ HO), dar seftîren || fait de faire tendre com. tapis : fait de faire tendre de tapis.
- ătefter :+ H+ sm. nv. f. 3<sup>bis</sup> ; \( \text{p} \) (pl. iteftiren | :+ H+), dar teftiren || fait d'être tendu com. tapis ; fait d'être tendu de tapis ; fait de se tendre com. tapis : fait de se tendre de tapis.
- äsefter :+ HO sm. φ (pl. iseftår :+ HO), dar seftår || tapis (tapis proprement dit, couverture, tissu, natte, peau, paille, ch. qlconque qu'on étend pour servir de tapis) || syn. de taftek empl. d. ce s., et moins us. que taftek.
- O+H téfetest +10+H+ sf. φ (pl. tifetesin I0+H+), dar täfetest (těfetest), dar třetesin || ocre rouge sous forme de pierre tendre || la téfetest, réduite en poudre, sert de fard pour le visage des femmes et qlqf. de remède pour les yeux || diffère de makara « ocre jaune sous forme de pierre tendre » || diffère de tämeghoit « ocre rouge sous forme de terme»
- #+H ăfteżżen I#+H sm. φ (n. d'u. et col.) (pl. de div. ou p. n. ifteżżân I#+H), dar ĕfteżżân || nom d'une plante non persistante (« zilla macroptera Cosson » (B.T.)) (ar. « chebreg »).
- ăfteżżen em Makket +: EIX+ H (m. à m. « ăfteżżen de La Mecque ») sm. φ (col. s. n. d'u.) (pl. de div. ifteżżân em Makket +: EIX+ H), dar ĕfteżżân em Makket || pois chiches.
- ΣΕΉ tăfettouit | v. ΣΕΉ fedei.
- KH effež XH va. prim.; conj. 27 « eddel »; (ifféž, iffáž, éd iffež, our iffíž) || mâcher (broyer avec les dents) || a aussi les s. pas. et pron. « être mâché » et « se mâcher ».
- neffeż \*\*XII vn. f. 4 ; conj. 99 « bereġ » ; (innĕffeż, iennîffeż, éd ienneffeż, our inneffeż] || être mâché || ne s'empl. qu'au fig., le suj. étant un membre de p. ou d'an., une étoffe, une peau, comprimés par accident entre 2 corps durs, pour exprimer qu'ils ont été meurtris et broissés et com. mâchés.
- tâffeż ※光+ va. f. 6; conj. 226 « tâddel »; (itâffeż, our iteffeż) ||
  mâcher hab. || a aussi les s. pas. et pron.

[373]

## Rate of Progress

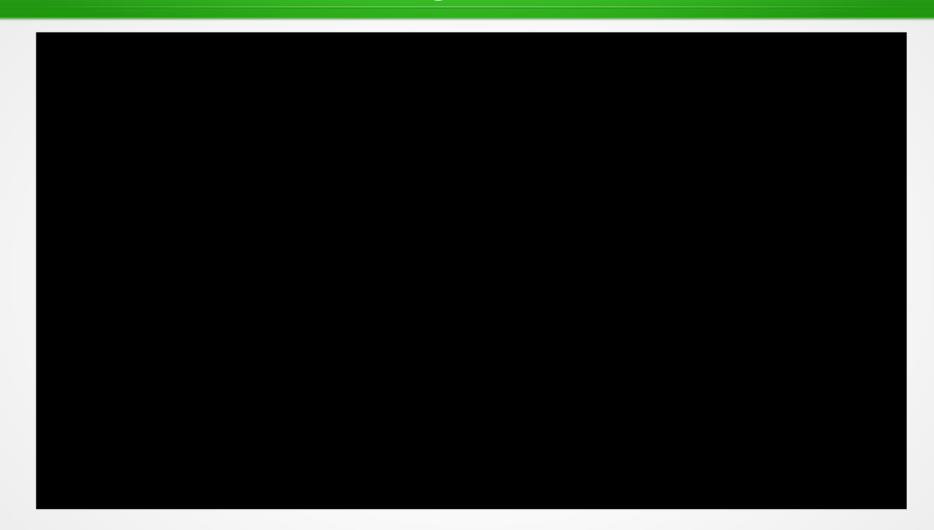


## Handwritten Text Recognition

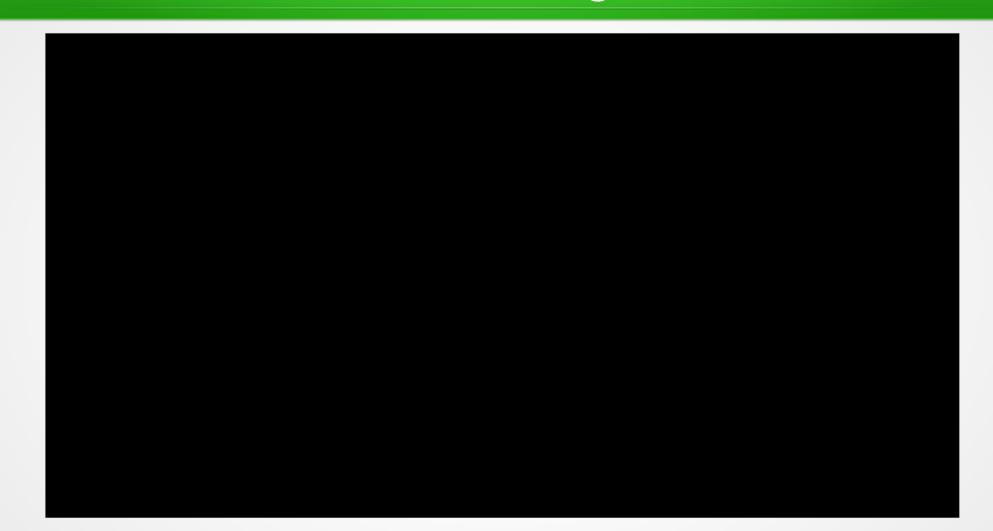
### Transkribus

- readcoop.eu/transkribus
- powered by machine learning
  - needs training data
  - is flexible about the type of script (handwriting, print, ... writing direction)
- Transkribus offers limited free use to new users;
   has to be paid after the first few hundred pages

# Line Segmentation



# Automatic Recognition



### Precision

```
← 73
fait d'être tendu com. tapis : fait d'être tendu de tapis : fait de se
                                                                                                            fait d'être tendu com. tapis : fait d'être tendu de tapis : fait de se
tendre com, tapis : fait de se tendre de tapis.
                                                                                                            tendre com, tapis : fait de se tendre de tapis.
- ăsefter ⊙H+: sm. φ (pl. iseftâr ⊙H+:), dar seftâr ∥ tapis (tapis
                                                                                                            - ăsefter ⊙H+: sm. φ (pl. iseftâr ⊙H+:), dar seftâr ∥ tapis (tapis
proprement dit. couverture, tissu, natte, peau, paille, ch. glconque
                                                                                                            proprement dit, couverture, tissu, natte, peau, paille, ch. glconque
ou'on étend pour servir de tapis) | syn. de taftek empl. d. ce s.. et
                                                                                                            qu'on étend pour servir de tapis) | syn. de taftek empl. d. ce s.. et
moins us. que taftek.
                                                                                                            moins us. que taftek.
H+O téfetest +H+O+ sf. φ (pl. tifetesîn +H+OI), dar tăfetest (těfetest).
                                                                                                            H+O téfetest +H+O+ sf. φ (pl. tifetesin +H+OI), dar tăfetest (těfetest).
dar tfetesin ∥ ocre rouge sous forme de pierre tendre ∥ la téfetest.
                                                                                                            dar tfetesîn ∥ ocre rouge sous forme de pierre tendre ∥ la téfetest.
réduite en poudre, sert de fard pour le visage des femmes et
                                                                                                            réduite en poudre, sert de fard pour le visage des femmes et
qlqf. de remède pour les yeux | diffère de makara "ocre jaune
                                                                                                            qlqf. de remède pour les yeux | diffère de makara "ocre jaune
sous forme de pierre tendre" | diffère de tămeghoit "ocre rouge
                                                                                                            sous forme de pierre tendre" | diffère de tămeghoit "ocre rouge
sous forme de terre".
                                                                                                            sous forme de terre".
H+X ăfteżżen H+XI sm. φ (n. d'u. et col.) (pl. de div. ou p. n. ifteżżân
                                                                                                            H+米 ăfteżżen H+米I sm. φ (n. d'u. et col.) (pl. de div. ou p. n. ifteżżân
H+米I), dar ēfteżżān ∥ nom d'une plante non persistante ("zilla
                                                                                                            H+XI), dar ĕfteżżân ∥ nom d'une plante non persistante ("zilla
macroptera Cosson" (B.T.)) (ar. "chebreg").
                                                                                                          ← macroptera Cosson" (B. T.)) (ar. "chebreg").
- ăfteżżen em Makket X+XIC:+ (m. à m. "ăfteżżen de La Mecque")
                                                                                                            - ăfteżżen em Makket H+IC:+ (m. à m. ăfteżżen de la Mecqué)
sm. φ (col. s. n. d'u.) (pl. de div. ifteżżân em Makket Η+ЖΙС:+), dar
                                                                                                            sm. φ (col. s. n. d'u.) (pl. de div. ifteżżân em Makket H+IC++), dar
ěfteżżân em Makket ∥ pois chiches.
                                                                                                            ēfteżżân em Makket | pois chiches.
HEΣ tăfettouit | v. HEΣ fedei.
                                                                                                            HEΣ tăfettouit | v. HEΣ fedei.
H米 effež H米 va. prim. : coni. 27 "eddel" : (iffěž. iffåž. éd iffež. our
                                                                                                            H米 effež H米 va. prim. : coni. 27 "eddel" : (iffěž. iffåž. éd iffež. our
iffiž) | macher (brover avec les dents) | a aussi les s. pas. et pron.
                                                                                                            iffiž) | macher (brover avec les dents) | a aussi les s. pas. et pron.
"être mâché" et "se mâcher".
                                                                                                            "être mâché" et "se mâcher".
- neffeż l光業 vn. f. 4 ; conj. 99 "bereġ" ; (inněffeż, iennîffeż, éd
                                                                                                            - neffeż l光業 vn. f. 4 ; conj. 99 "bereġ" ; (inněffeż, iennîffeż, éd
ienneffeż, our inneffeż) ∥ être mâché ∥ ne s'empl. qu'au fig., le suj.
                                                                                                            ienneffeż, our inneffeż) ∥ être mâché ∥ ne s'empl. qu'au fig., le suj.
étant un membre de p. ou d'an., une étoffe, une peau, comprimés
                                                                                                            étant un membre de p. ou d'an., une étoffe, une peau, comprimés
par accident entre 2 corps durs, pour exprimer qu'ils ont été
                                                                                                            par accident entre 2 corps durs, pour exprimer qu'ils ont été
meurtris et broisses et com. mâches.
                                                                                                          + meurtris et broisses et com. mâches.
- tâffež 十光米 va. f. 6 : conj. 226 "tâddel" : (itâffež. our iteffež) ∥
                                                                                                            - tâffeż +H≭ va. f. 6 : conj. 226 "tâddel" : (itâffeż, our iteffeż) ∥
mâcher hab. | a aussi les s. pas. et pron.
                                                                                                            mâcher hab. | a aussi les s. pas. et pron.
- tâneffâż 十日光米 vn. f. 4.7 ; conj. 230 "târeġâh" ; (itâneffâż, our
                                                                                                            - tâneffâż +IHX vn. f. 4.7 ; conj. 230 "târeġâh" ; (itâneffâż, our
iteneffiż) | être hab. mâché | a le s. c. à c. de la f. 4.
                                                                                                            iteneffiż) | être hab. mâché | a le s. c. à c. de la f. 4.
- oûfoûż H≭ sm. nv. prim. : (pl. oûfoûżen H≭I) | fait de mâcher |
                                                                                                            - oûfoûż H\ sm. nv. prim. : (pl. oûfoûżen H\|) | fait de mâcher |
a aussi les s. pas. et pron. "fait d'être mâché" et "fait de se
                                                                                                            a aussi les s. pas. et pron. "fait d'être mâché" et "fait de se
mâcher".
                                                                                                            mâcher".
- ăneffeż ΙΗЖ sm. nv. f. 4 ; φ (pl. ineffîżen ΙΗЖΙ), dar neffîżen
                                                                                                            - ăneffeż IHЖ sm. nv. f. 4 ; φ (pl. ineffîżen IHЖI), dar neffîżen |
fait d'être mâché | a aussi a le s. c. à c. de la f. 4.
                                                                                                            fait d'être mâché | a aussi a le s. c. à c. de la f. 4.
HX feżż HX sm. (s. s. pl.) ∥ "feżż" ; mot sans signification figurant le
                                                                                                            HX feżż HX sm. (s. s. pl.) ∥ "feżż"; mot sans signification figurant le
bruit d'un liquide ou d'un gaz enfermés dans un récipient qui
                                                                                                            bruit d'un liquide ou d'un gaz enfermés dans un récipient qui
en jaillissent avec force, ou celui d'une ch. qui est projetée à
                                                                                                            en jaillissent avec force, ou celui d'une ch. qui est projetée à
terre avec violence | sign. aussi "bruit produit par un vent
                                                                                                            terre avec violence | sign. aussi "bruit produit par un vent
```

### Results

- character error rate below 1%
  - ca. 26 errors in a typical page of 3000 characters
- transcription time for one page reduced from 30–45 minutes to ca. 10–15 minutes
  - with higher accuracy
- partial migration to a completely cost-free tool is underway

# Georeferencing

### Dictionnaire [...] de noms propres

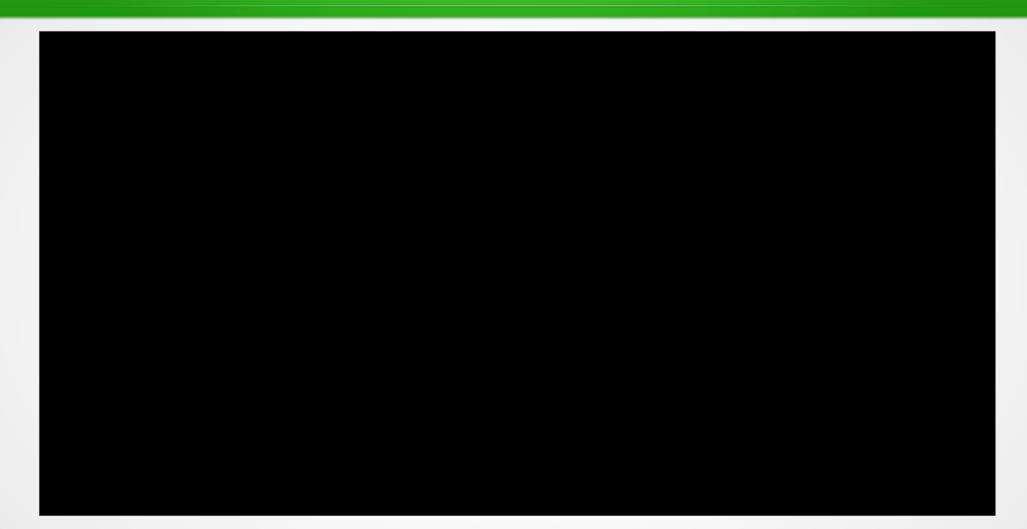
- largest part: toponyms (over 2700 entries)
  - with reference to the named places, i.e. a gazetteer
  - ca. 2300 unique names (types) for ca. 3500 places (tokens)
- toponyms are coarsely located in regions
  - ... thousands of square kilometers large

## Linking Geographical Data

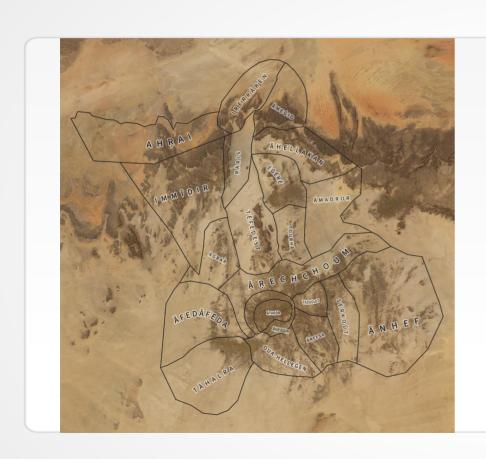
- geometries (lines, points, areas) reside in OpenStreetMap
- unique identifiers are taken from GeoNames
- linking happens in two steps, with GeoNames serving as a tertium comparationis

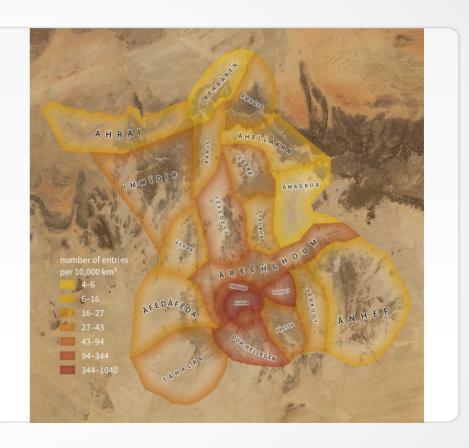


# Mapping in QGIS



## Result





# Concluding Remarks

## Conclusions

- it works
- it really is cost free\*
  - \*except Transkribus