A. GBI

B. TLI

external data inputs

logical + design curation step

logical datasets

statistical curation step

statistical datasets

195 features

THE WORLD ATLAS OF LANGUAGE STRUCTURES ONLINE

AUTOTYP

['fɔɪ.bɬ]

scripts/TLI/compile-external-input.R

838 features

(curated_data/TLI/compiled_external_input_features.csv)

(192 [W] + 545 [A] + 60 [L] + 41 [P])

95 retentions

100 exclusions

4 exclusions w/o replacement

96 exclusions with replacement

6 new features derived from 1 input feature

102 new features derived from >1 input features

54 meta-features

48 input features, conditioned

scripts/*dataset*/decisions-log.csv, modification.type ∈ (logical, design, design.automated, design.manual))

102 retentions (15 + 52 + 17 + 18)

736 exclusions (177 + 493 + 43 + 23)

254 automated exclusions (22 + 227 + 5 + 0)

79 features with < 100 languages (17 + 62 + 0 + 0)

175 features insufficiently variable (5 + 165 + 5 + 0)

482 manual exclusions (155 + 266 + 38 + 23)

282 exclusions w/o replacement (46 + 209 + 26 + 1)

200 exclusions with replacement (109 + 57 + 12 + 22)

174 new features derived from >1 input features

73 new features derived from 1 input feature (65 + 8 + 0 + 0)

52 combined features (6 [W] + 6 [A] + 15 [W, A] + 6 [W, L] + 19 [W, P])

122 input features, conditioned (82 + 33 + 3 + 4)

scripts/GBI/create-full-datasets.R

GBI logical: 203 features

curated_data/GBI/logicalGBI/**logicalGBI.csv**

curated_data/GBI/logicalGBI/**cldf**

scripts/*dataset*/feature-recode-patterns.csv, design.logical = TRUE

scripts/TLI/create-full-datasets.R

TLI logical: 349 features

curated_data/TLI/logicalTLI/full/**logicalTLI.csv**

curated_data/TLI/logicalTLI/**cldf**

157 retentions

46 exclusions

3 exclusions w/o replacement

43 exclusions with replacement

39 new features derived from >1 input features

5 combined features

34 input features, conditioned

scripts/*dataset*/decisions-log.csv modification.type = statistical

300 retentions (147 [W] + 79 [A] + 17 [L] + 18 [P] + 15 [W, A] + 6 [W, L] + 18 [W, P])

49 exclusions (21 [W] + 20 [A] + 3 [L] + 4 [P] + 1 [W, P])

13 exclusions w/o replacement (5 + 7 + 1 + 0)

36 exclusions with replacement (16 [W] + 13 [A] + 2 [L] + 4 [P] + 1 [W, P])

33 new features derived from >1 input features

3 combined features (1 [A] + 1 [P] + 1 [L])

30 input features, conditioned (16 [W] + 11 [A] + 2 [P] + 1 [W, P])

scripts/GBI/create-full-datasets.R

GBI statistical: 196 features

curated_data/GBI/statisticalGBI/**statisticalGBI.csv**

curated_data/GBI/statisticalGBI/**cldf**

scripts/*dataset*/feature-recode-patterns.csv, design.logical.statistical = T

scripts/TLI/create-full-datasets.R

TLI statistical: 349 features

curated_data/TLI/statisticalTLI/full/**statisticalTLI.csv**

curated_data/TLI/statisticalTLI/**cldf**