TICO-19 in HXLTM (public working draft)

Table of Contents

Public domain datasets of the [Translation Initiative for COVID-19](#) on the format HXLTM (Multilingual Terminology in Humanitarian Language Exchange).

1. Versions

- Online: https://tico-19-hxltm.etica.ai/

- PDF: https://tico-19-hxltm.etica.ai/tico-19-hxltm_eng-Latn.pdf

- Ebook EPUB: https://tico-19-hxltm.etica.ai/tico-19-hxltm_eng-Latn.epub

2. Tables

TICO-19 language pair	Source Language	Source language BCP47	Target language	Target language BCP47	Deterministic language pair
en-ar	en	en	ar	ar	en_ar
en-bn	en	en	bn	bn	en_bn
en-ckb	en	en	ckb	ckb	en_ckb
en-din	en	en	din	din	en_din
en-es-LA	en	en	es-LA	es-419	en_es-419
en-fa	en	en	fa	fa	en_fa
en-fr	en	en	fr	fr	en_fr
en-fuv	en	en	fuv	fuv	en_fuv
en-ha	en	en	ha	ha	en_ha
en-hi	en	en	hi	hi	en_hi
en-id	en	en	id	id	en_id
en-km	en	en	km	km	en_km
en-kr	en	en	kr	kr	en_kr
en-ku	en	en	ku	ku	en_ku
en-lg	en	en	lg	lg	en_lg
en-ln	en	en	ln	ln	en_ln
en-mr	en	en	mr	mr	en_mr
en-ms	en	en	ms	ms	en_ms
en-my	en	en	my	my	en_my
en-ne	en	en	ne	ne	en_ne
en-nus	en	en	nus	nus	en_nus
en-om	en	en	om	om	en_om

TICO-19 language pair	Source Language	Source language BCP47	Target language	Target language BCP47	Deterministic language pair
en-prs	en	en	prs	prs	en_prs
en-ps	en	en	ps	ps	en_ps
en-pt-BR	en	en	pt-BR	pt-BR	en_pt-BR
en-ru	en	en	ru	ru	en_ru
en-rw	en	en	rw	rw	en_rw
en-so	en	en	so	so	en_so
en-sw	en	en	sw	sw	en_sw
en-ta	en	en	ta	ta	en_ta
en-ti	en	en	ti	ti	en_ti
en-ti_ER	en	en	ti_ER	ti-ER	en_ti-ER
en-ti_ET	en	en	ti_ET	ti-ET	en_ti-ET
en-tl	en	en	tl	tl	en_tl
en-ur	en	en	ur	ur	en_ur
en-zh	en	en	zh	zh	en_zh
en-zu	en	en	zu	zu	en_zu

3. Quick explanations

4. Original data + minor changes

5. Appendix

5.1. A : Facebook dataset

- Source

  ◦ File: `fb_covid-19.zip/fb_covid-19/README.md`, date 2020-04-27

  ◦ Link: https://github.com/tico-19/tico-19.github.io/blob/master/data/fb_covid-19.zip

  ◦ Exact link: https://github.com/tico-19/tico-19.github.io/raw/f7357b64618f970a57f56aac50f66c8897afa852/data/fb_covid-19.zip

```
---
# COVID-19 Glossary translation
```

These files contain one term per line. These were translated by Facebook from English (en_XX) into

many languages.

```
Key Dialect
af_ZA   Afrikaans
am_ET   Amharic
ar_AR   Arabic
as_IN   Assamese
az_AZ   Azerbaijani
be_BY   Belarusian
bg_BG   Bulgarian
bn_IN   Bengali
bs_BA   Bosnian
ca_ES   Catalan
cb_IQ   Sorani Kurdish
cs_CZ   Czech
cx_PH   Cebuano
da_DK   Danish
de_DE   German
el_GR   Greek
es_XX   Spanish
et_EE   Estonian
fa_IR   Persian
fi_FI   Finnish
fr_XX   French
gu_IN   Gujarati
ha_NG   Hausa
he_IL   Hebrew
hi_IN   Hindi
hr_HR   Croatian
ht_HT   Haitian Creole
hu_HU   Hungarian
hy_AM   Armenian
id_ID   Indonesian
ig_NG   Igbo
is_IS   Icelandic
it_IT   Italian
ja_XX   Japanese
jv_ID   Javanese
ka_GE   Georgian
kk_KZ   Kazakh
km_KH   Khmer
kn_IN   Kannada
ko_KR   Korean
lg_UG   Ganda
ln_CD   Lingala
lo_LA   Lao
lt_LT   Lithuanian
lv_LV   Latvian
mg_MG   Malagasy
mk_MK   Macedonian
```

```
ml_IN    Malayalam
mn_MN    Mongolian
mr_IN    Marathi
ms_MY    Malay
my_MM    Burmese
ne_NP    Nepali
nl_XX    Dutch
no_XX    Norwegian
ns_ZA    Northern Sotho
om_KE    Oromo
pa_IN    Punjabi
pl_PL    Polish
ps_AF    Pashto
pt_XX    Portuguese
ro_RO    Romanian
ru_RU    Russian
si_LK    Sinhala
sk_SK    Slovak
sl_SI    Slovenian
so_SO    Somali
sq_AL    Albanian
sr_RS    Serbian
ss_SZ    Swazi
su_ID    Sundanese
sv_SE    Swedish
sw_KE    Swahili
ta_IN    Tamil
te_IN    Telugu
th_TH    Thai
tl_XX    Filipino
tn_BW    Tswana
tr_TR    Turkish
uk_UA    Ukrainian
ur_PK    Urdu
vi_VN    Vietnamese
wo_SN    Wolof
xh_ZA    Xhosa
yo_NG    Yoruba
zh_CN    Chinese (Simplified)
zh_TW    Chinese (Traditional)
zu_ZA    Zulu
```

5.2. B : Google datasets, readme.md

- Source
  - File: `google_covid-19.zip/google_covid-19/readme.md`, date 2020-04-27
  - Link: https://github.com/tico-19/tico-19.github.io/raw/master/data/google_covid-19.zip

```
File Format
Language and Locale Format
We use BCP-47 (https://tools.ietf.org/html/bcp47) as language and locale code format,
conforming to casing specs (https://tools.ietf.org/html/bcp47#section-3.1.4) and will
use hyphen to indicate locales or scripts.

We use two letter language code in most cases except for:

es-419, es-ES
fr-FR, fr-CA
pt-BR, pt-PT
zh-CN, zh-TW, zh-HK

File Format
CSV files with BCP-47 standard with following headers:
stringID | sourceLang | targetLang | pos | description | sourceString | targetString

Files are encoded in UTF-8.

pos tags will follow the spelled out pos names in POS Universal tags:
https://universaldependencies.org/u/pos/


File Naming Convention
sourceLang_targetLang
The file name should be all lower case. Example: en_af, en_pt-br

Tracking changes
After the initial batch of term commits, we will use the index.csv file to track the
following file change status:

Draft (the terms have been translated by professional translators but haven[]t been
independently reviewed) or Revised
Additional languages are being committed
Additional source terms are being added
Additional translations are being added

index.csv file has headers: file_name | status

Example:
en_af.csv | Draft
en_ms.csv | Revised

Translation Quality:

Translations have been created by professional translators.
```

> Some translations have not gone through independent review and are marked as draft, and translations with additional reviews have been marked as revised.
>
> All translations are provided as-is without warranty or any guarantees of correctness.

5.3. B : Google datasets, index.csv

- Source

    ◦ File: `google_covid-19.zip/google_covid-19/index.csv`, date 2020-04-27

    ◦ Link: https://github.com/tico-19/tico-19.github.io/raw/master/data/google_covid-19.zip

    ◦ Exact link: https://github.com/tico-19/tico-19.github.io/raw/f7357b64618f970a57f56aac50f66c8897afa852/data/google_covid-19.zip

```
ar_en.csv    Draft
bn_en.csv    Draft
cs_en.csv    Draft
da_en.csv    Draft
de_en.csv    Draft
en_af.csv    Draft
en_am.csv    Draft
en_ar.csv    Draft
en_az.csv    Draft
en_be.csv    Draft
en_bg.csv    Draft
en_bn.csv    Draft
en_bs.csv    Draft
en_ca.csv    Draft
en_ceb.csv   Draft
en_co.csv    Draft
en_cs.csv    Draft
en_cy.csv    Draft
en_da.csv    Draft
en_de.csv    Draft
en_el.csv    Draft
en_eo.csv    Draft
en_es-419.csv    Draft
en_et.csv    Draft
en_eu.csv    Draft
en_fa.csv    Draft
en_fi.csv    Draft
en_fil.csv   Draft
en_fr-FR.csv     Draft
en_fy.csv    Draft
en_ga.csv    Draft
en_gd.csv    Draft
en_gl.csv    Draft
en_gu.csv    Draft
en_ha.csv    Draft
```

```
en_he.csv    Draft
en_hi.csv    Draft
en_hmn.csv   Draft
en_hr.csv    Draft
en_ht.csv    Draft
en_hu.csv    Draft
en_hy.csv    Draft
en_id.csv    Draft
en_ig.csv    Draft
en_is.csv    Draft
en_it.csv    Draft
en_ja.csv    Draft
en_jv.csv    Draft
en_ka.csv    Draft
en_kk.csv    Draft
en_km.csv    Draft
en_kn.csv    Draft
en_ko.csv    Draft
en_ku.csv    Draft
en_ky.csv    Draft
en_la.csv    Draft
en_lb.csv    Draft
en_lo.csv    Draft
en_lt.csv    Draft
en_lv.csv    Draft
en_mg.csv    Draft
en_mk.csv    Draft
en_ml.csv    Draft
en_mn.csv    Draft
en_mr.csv    Draft
en_ms.csv    Draft
en_my.csv    Draft
en_nb.csv    Draft
en_ne.csv    Draft
en_nl.csv    Draft
en_ny.csv    Draft
en_pa.csv    Draft
en_pl.csv    Draft
en_ps.csv    Draft
en_pt-BR.csv    Draft
en_ro.csv    Draft
en_ru.csv    Draft
en_sd.csv    Draft
en_si.csv    Draft
en_sk.csv    Draft
en_sl.csv    Draft
en_sm.csv    Draft
en_sn.csv    Draft
en_so.csv    Draft
en_sq.csv    Draft
en_sr.csv    Draft
```

```
en_st.csv    Draft
en_su.csv    Draft
en_sv.csv    Draft
en_sw.csv    Draft
en_ta.csv    Draft
en_te.csv    Draft
en_tg.csv    Draft
en_th.csv    Draft
en_tr.csv    Draft
en_uk.csv    Draft
en_ur.csv    Draft
en_uz.csv    Draft
en_vi.csv    Draft
en_xh.csv    Draft
en_yi.csv    Draft
en_yo.csv    Draft
en_zh-CN.csv     Draft
en_zh-TW.csv     Draft
en_zu.csv    Draft
es-419_en.csv    Draft
es-ES_en.csv     Draft
fa_en.csv    Draft
fr_en.csv    Draft
hi_en.csv    Draft
id_en.csv    Draft
it_en.csv    Draft
iw_en.csv    Draft
ja_en.csv    Draft
ko_en.csv    Draft
ms_en.csv    Draft
nl_en.csv    Draft
no_en.csv    Draft
pt-BR_en.csv     Draft
pt-PT_en.csv     Draft
ru_en.csv    Draft
sv_en.csv    Draft
th_en.csv    Draft
tr_en.csv    Draft
vi_en.csv    Draft
zh-CN_en.csv     Draft
zh-TW_en.csv     Draft
```

6. License

The EticaAI has dedicated the work to the public domain by waiving all of their rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law. You can copy, modify, distribute and perform the work, even for commercial purposes, all

without asking permission.