

Parallel texts from the Swedish Migration Agency

SND-ID: ext0329-1.

Is part of collection at SND: [Parallel Texts from Public Agencies](#)

Creator/Principal investigator(s)

Simon Dahlberg - Institute for Language and Folklore, Language Council of Sweden
Institute for Language and Folklore, Language Council of Sweden

Research principal

[Institute for Language and Folklore](#) - Language Council of Sweden

Description

Parallel texts downloaded with "w3m -dump" from an ubuntu shell, from the website of the Swedish Migration Agency.

The texts have been downloaded using the command 'w3m -dump' from an ubuntu shell, whereafter the resulting text files were stripped to contain only the interesting text (no menus and such).

Sampling procedure

Multilingual parallel content.

Data format / data structure

[Text](#)

Data collection 1

- Mode of collection: Self-administered writings and/or diaries: web-based
- Time period(s) for data collection: 2019-01-01 - 2019-01-31

Resource type

Corpus

Foreseen use

NLP application

Text corpus

- Linguality
 - Multilingual
- Language
 - Swedish (swe)
Texts: 33
 - Amharic (amh)
Texts: 23
 - Arabic (ara)
Texts: 33

- Azerbaijani (aze)
Texts: 27
- Central Kurdish (ckb)
Texts: 29
- English (eng)
Texts: 33
- Persian (fas)
Texts: 32
- Croatian (hrv)
Texts: 23
- Armenian (hye)
Texts: 24
- Georgian (kat)
Texts: 1
- Northern Kurdish (kmr)
Texts: 28
- Mongolian (mon)
Texts: 25
- Dari (prs)
Texts: 28
- Pushto (pus)
Texts: 28
- Romany (rom)
Arli (dialect)
Texts: 24
- Russian (rus)
Texts: 33
- Somali (som)
Texts: 29
- Spanish (spa)
Texts: 31
- Albanian (sqi)
Texts: 27
- Thai (tha)
Texts: 4
- Tigrinya (tir)
Texts: 29
- Turkish (tur)
Texts: 2
- Uzbek (uzb)

Texts: 25

- Chinese (zho)

Texts: 3

- French (fra)

Texts: 31

[More..](#)

- Modality

Written Language

- Size

Words: 29008 (swe)

Texts: 33 (swe)

Words: 438614 (TOT)

Texts: 580 (TOT)

- Original source

migrationsverket

www.migrationsverket.se

Geographic spread

Geographic location: [Sweden](#)

Responsible department/unit

Language Council of Sweden

Contributor(s)

Institute for Language and Folklore, Language Council of Sweden

Identifiers

URL:

<http://liljeholmen.sprakochfolkminnen.se/sprakresurser/version/20190124/myndighetsdata/texter>

URL:

<http://liljeholmen.sprakochfolkminnen.se/sprakresurser/version/20190124/myndighetsdata/texter/Migrationsverket>

Research area

[Society and culture](#) (CESSDA Topic Classification)

[Legislation and legal systems](#) (CESSDA Topic Classification)

[Conflict, security and peace](#) (CESSDA Topic Classification)

[International politics and organisations](#) (CESSDA Topic Classification)

[Social sciences](#) (Standard för svensk indelning av forskningsämnen 2011)

[Languages and literature](#) (Standard för svensk indelning av forskningsämnen 2011)

[Social welfare policy and systems](#) (CESSDA Topic Classification)

Keywords

[Immigration policy](#), [Emigration and immigration](#)

Accessibility level

Access to data through an external actor
Data are freely accessible

Homepage

[Link to Texts from Public Agencies](#)

[Link to resource](#)

Contact for questions about the data

Gunnar Eriksson

gunnar.eriksson@sprakochfolkminnen.se

This resource has the following relations

Is part of [Texts from Public Agencies](#)

Related research data in SND's catalogue

[Parallel texts from public agencies](#)

Is part of collection at SND

[Parallel Texts from Public Agencies](#)

CLARIN Virtual Collection Registry

[Add to collection](#)

A virtual collection is connected to a specific research purpose and contains links to data resources from various digital archives. It is easy to create, access, and cite the collection.

Read more about virtual collections on the [CLARIN website](#).

Download metadata

[DataCite](#)

[MetaShare](#)

[MetaShare-CMDI](#)

[DDI 2.5](#)

[DDI 3.3](#)

[DCAT-AP-SE 2.0](#)

[JSON-LD](#)

[PDF](#)

[Citation \(CLS\)](#)