

Gold standard for English-Swedish Europarl data (GES)

SND-ID: ext0283-1.

Creator/Principal investigator(s)

Lars Ahrenberg - Linköping University, Department of Computer and Information Science

Maria Holmqvist - Linköping University, Department of Computer and Information Science

Research principal

[Linköping University](#) - Department of Computer and Information Science

Description

Reference corpus for word linking, divided into training data and test data. The sentences come from the English and Swedish parts of Europarl.

Data are created from the English-Swedish part of the Europarl corpus. For each sentence pair in the selected subset, token correspondences are stated as pairs of integral token identifiers

Data format / data structure

[Numeric](#)

[Text](#)

Resource type

Corpus

Foreseen use

NLP application

Text corpus

- Linguality
 - Bilingual
 - Language
 - English (eng)
 - Swedish (swe)
- Sentences: 1164

[More..](#)

- Modality
 - Written Language
- Size
 - Sentences: 1164
- Annotation
 - Alignment
 - Manual annotation

Responsible department/unit

Department of Computer and Information Science

Research area

[Engineering and technology](#) (Standard för svensk indelning av forskningsämnen 2011)

[Language technology \(computational linguistics\)](#) (Standard för svensk indelning av forskningsämnen 2011)

Keywords

[Translation](#), [Natural language processing](#)

Publications

Maria Holmqvist and Lars Ahrenberg (2011). A Gold Standard for English-Swedish Word Alignment. In Proceedings of the 18th Nordic Conference on Computational Linguistics, Riga, Latvia, May 11-13, 2011.

If you have published anything based on these data, [please notify us](#) with a reference to your publication(s). If you are responsible for the catalogue entry, you can update the metadata/data description in DORIS.

Accessibility level

Access to data through an external actor
Data are freely accessible

License

[CC BY 4.0](#)

Homepage

[A Gold Standard Word Alignment for English-Swedish](#)

Contact for questions about the data

Lars Ahrenberg
lars.ahrenberg@liu.se

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 \(CLS\)](#)