

TALG Resources

Grupo TALG
Tecnoloxías e Aplicacións da Lingua Galega
Universidade de Vigo
<http://sli.uvigo.gal>

ReTeLe Workshop
Universidad de Salamanca, 2016

1 TALG Resources

2 Corpora

3 Lexical resources

1. TALG Resources

■ Corpora

- C1. CLUVI Parallel Corpus
- C2. CTG Galician Technical Corpus
- C3. SensoGal Parallel Corpus

■ Lexical resources

- L1. Galnet (Galician WordNet)
- L2. Termoteca (Terminology Knowledge Base)
- L3. English-Galician CLUVI Dictionary
- L4. Galician Synonym Dictionary
- L5. Galician DBpedia

2. Corpora

[C1]. CLUVI Parallel Corpus

- Linguality type: Multilingual
- Language: Basque, Catalan, English, French, Galician, German, Portuguese, Spanish
- Text format: plain
- Size
 - Number of sentences: 1,250,000
 - Number of words: 23,000,000
- Character encoding: ISO 8859-1

2. Corpora

[C1] CLUVI Parallel Corpus

- Classification
 - Text type: literary, subtitling and specialized (legal, tourism, computing, science divulgation, economy)
 - Register: formal
- Annotation: TMX
- Creation
 - Creation mode: computer assisted, human revision
- Version: 2.1
- Documentation: <http://sli.uvigo.gal/CLUVI>
- ISLRN: 910-993-402-072-9

[C1] CLUVI Parallel Corpus

- Distribution
 - Availability: <http://sli.uvigo.gal/CLUVI>
 - License
 - License: CC BY-NC-SA 3.0
 - Download location: Meta-Share
<http://hdl.handle.net/10230/20051>
 - Distribution Access/Medium: free online access
- Contact Person: Xavier Gómez Guinovart (xgg@uvigo.es)
- Resource Creation
 - Resource creator: TALG Group (UVigo)
 - Funding project: CLUVI (BFF2002-01385)

2. Corpora

[C2]. CTG Galician Technical Corpus

- Linguality type: Monolingual
- Language: Galician
- Text format: plain
- Size
 - Number of sentences: 800,000
 - Number of words: 18,000,000
- Character encoding: ISO 8859-1

[C2] CTG Galician Technical Corpus

- Classification
 - Text type: legal, computing, environment, economy, sociology, medicine
 - Register: formal
- Annotation: PoS, lemma, sense
- Creation
 - Creation mode: computer assisted
- Version: 1.0
- Documentation: <http://sli.uvigo.gal/CTG>
- ISLRN: 437-045-879-366-6

[C2] CTG Galician Technical Corpus

- Distribution
 - Availability: <http://sli.uvigo.gal/CTG>
 - License
 - License: CC BY-NC-SA 3.0
 - Download location:
 - Universal Dependencies 1.3 – LINDAT/CLARIN
<http://hdl.handle.net/11234/1-1699>
 - Distribution Access/Medium: free online access
- Contact Person: Xavier Gómez Guinovart (xgg@uvigo.es)
- Resource Creation
 - Resource creator: TALG Group (UVigo)
 - Funding project: SKATeR (TIN2012-38584-C06-04)

2. Corpora

[C3]. SensoGal Parallel Corpus

- Linguality type: Multilingual
- Language: English, Galician
- Text format: plain
- Size
 - Number of sentences: 3,000
 - Number of words: 60,000
- Character encoding: ISO 8859-1

[C3] SensoGal Parallel Corpus

- Classification
 - Text type: skill and hobbies, popular lore, belles lettres
 - Register: formal
- Annotation: TMX, PoS, lemma, sense
- Creation
 - Creation mode: computer assisted, human revision
- Version: 1.0 (under construction)
- Documentation: <http://sli.uvigo.gal/SensoGal>
- ISLRN: 653-144-288-768-2

[C3] SensoGal Parallel Corpus

- Distribution
 - Availability: <http://sli.uvigo.gal/SensoGal>
 - License
 - License: pending
 - Download location: pending
 - Distribution Access/Medium: free online access
- Contact Person: Xavier Gómez Guinovart (xgg@uvigo.es)
- Resource Creation
 - Resource creator: TALG Group (UVigo)
 - Funding project: TUNER (TIN2015-65308-C5-1-R)

3. Lexical resources

[L1] Galnet (Galician WordNet)

- Resource type: Lexicon
- Linguality type: Monolingual
- Languages: Galician
- Text format: SQL
- Size
 - Number of entries: 34,174 synsets (52,186 variants)
- Character encoding: UTF-8

3. Lexical resources

[L1] Galnet (Galician WordNet)

- Domains: general
- Encoding
 - Encoding level: lexical
 - Linguistic information: senses, variants, semantic relations, semantic classifications
- Creation mode: computer assisted, human editing
- Version: 3.0.22
- Documentation: <http://sli.uvigo.gal/galnet>
- ISLRN: 544-286-653-437-9

3. Lexical resources

[L1] Galnet (Galician WordNet)

- Distribution
- Availability: <http://sli.uvigo.gal/galnet>
- License
 - License: CC BY-NC-SA 3.0
 - Download:
 - Meta-Share (LMF): <http://hdl.handle.net/10230/22921>
 - MCR (SQL): <http://adimen.si.ehu.es/web/MCR/>
 - Distribution Access/Medium: free online access
- Contact Person: Xavier Gómez Guinovart (xgg@uvigo.es)
- Resource Creation
 - Resource creator: TALG Group (UVigo)
 - Funding project: SKATeR (TIN2012-38584-C06-04)

3. Lexical resources

[L2] Termoteca (Terminology Knowledge Base)

- Resource type: Terminology Knowledge Base
- Linguality type: Multilingual
- Languages: English, French, Galician, Portuguese, Spanish
- Text format: plain, XML
- Size
 - Number of entries: 8,000 concepts (16,000 terms)
- Character encoding: ISO 8859-1

3. Lexical resources

[L2] Termoteca (Terminology Knowledge Base)

- Domains: legal, computing, environment, economy, sociology, medicine
- Encoding
 - Encoding level: lexical and multilexical
 - Linguistic information: multilingual equivalences, corpus-based examples, semantic relations
- Creation mode: computer assisted, human editing
- Version: 2.0
- Documentation: <http://sli.uvigo.gal/termoteca>
- ISLRN: 013-411-392-664-8

3. Lexical resources

[L2] Termoteca (Terminology Knowledge Base)

- Distribution
- Availability: <http://sli.uvigo.gal/termoteca>
- License
 - License: CC BY-NC-SA 3.0
 - Download: Meta-Share (LMF version)
<http://hdl.handle.net/10230/17104>
 - Distribution Access/Medium: free online access
- Contact Person: Xavier Gómez Guinovart (xgg@uvigo.es)
- Resource Creation
 - Resource creator: TALG Group (UVigo)
 - Funding project: CLUVI (BFF2002-01385)

3. Lexical resources

[L3] English-Galician CLUVI Dictionary

- Resource type: Lexicon
- Linguality type: Multilingual
- Languages: English, Galician
- Text format: plain, XML
- Size
 - Number of entries: 20,000 entries (30,000 equivalences)
- Character encoding: ISO 8859-1

3. Lexical resources

[L3] English-Galician CLUVI Dictionary

- Domains: general
- Encoding
 - Encoding level: lexical and multilexical
 - Linguistic information: bilingual equivalences, corpus-based translation examples, grammatical information, usage notes
- Creation mode: computer assisted, human editing
- Version: 2.2
- Documentation: <http://sli.uvigo.gal/diccionario>
- ISLRN: 814-297-765-781-0

3. Lexical resources

[L3] English-Galician CLUVI Dictionary

- Distribution
- Availability: <http://sli.uvigo.gal/diccionario>
- License
 - License: CC BY-NC-SA 3.0
 - Download: Meta-Share (LMF version)
<http://hdl.handle.net/10230/20051>
 - Distribution Access/Medium: free online access
- Contact Person: Xavier Gómez Guinovart (xgg@uvigo.es)
- Resource Creation
 - Resource creator: TALG Group (UVigo)
 - Funding project: CLUVI (BFF2002-01385)

3. Lexical resources

[L4] Galician Synonym Dictionary

- Resource type: Lexicon
- Linguality type: Monolingual
- Languages: Galician
- Text format: plain, XML
- Size
 - Number of entries: 27,000 entries (45,000 senses, 200,000 synonyms)
- Character encoding: ISO 8859-1

3. Lexical resources

[L4] Galician Synonym Dictionary

- Domains: general
- Encoding
 - Encoding level: lexical and multilexical
 - Linguistic information: senses, synonyms, grammatical information
- Creation mode: computer assisted, human editing
- Version: 300714
- Documentation: <http://sli.uvigo.gal/sinonimos>
- ISLRN: 378-916-911-644-3

3. Lexical resources

[L4] Galician Synonym Dictionary

- Distribution
- Availability: <http://sli.uvigo.gal/sinonimos>
- License
 - License: CC BY-NC-SA 3.0
 - Download: pending
 - Distribution Access/Medium: free online access
- Contact Person: Xavier Gómez Guinovart (xgg@uvigo.es)
- Resource Creation
 - Resource creator: TALG Group (UVigo)
 - Funding project: SKATeR (TIN2012-38584-C06-04)

3. Lexical resources

[L5] Galician DBpedia

- Resource type: Lexicon
- Linguality type: Monolingual
- Languages: Galician
- Text format: RDF
- Size
 - Number of entries: 11,000,000 RDF triples
- Character encoding: UTF-8

3. Lexical resources

[L5] Galician DBpedia

- Domains: encyclopedic
- Encoding
 - Encoding level: lexical and multilexical
 - Linguistic information: encyclopedic definition, equivalences, semantic classification
- Creation mode: computer assisted
- Version: 2015-10 (current)
- Documentation: <http://gl.dbpedia.org>

3. Lexical resources

[L5] Galician DBpedia

- Distribution
- Availability: <http://gl.dbpedia.org>
- License
 - License: CC BY-NC-SA 3.0
 - Download: DBpedia.org
<http://downloads.dbpedia.org/current/core-i18n/gl/>
 - Distribution Access/Medium: free online access
- Contact Person: Xavier Gómez Guinovart (xgg@uvigo.es)
- Resource Creation
 - Resource creator: TALG Group (UVigo)
 - Funding project: TECANDALI (CN 2014/007)

TALG Resources

Grupo TALG
Tecnoloxías e Aplicacións da Lingua Galega
Universidade de Vigo
<http://sli.uvigo.gal>

ReTeLe Workshop
Universidad de Salamanca, 2016