Translating SNOMED CT Terminology into a Minor Language

Olatz Perez-de-Viñaspre and Maite Oronoz
Outline

1. Introduction

2. SNOMED CT

3. Translation Algorithm
   - Phase 1: Lexical Resources
   - Phase 2: Finite State Transducers and Biomedical Affixes

4. Results

5. Conclusions
Outline

1 Introduction

2 SNOMED CT

3 Translation Algorithm
   • Phase 1: Lexical Resources
   • Phase 2: Finite State Transducers and Biomedical Affixes

4 Results

5 Conclusions
Introduction. SNOMED CT and Basque

- **SNOMED Clinical Terms (SNOMED CT)**
  - Comprehensive, multilingual *clinical healthcare terminology*
  - Enables consistent representation of meaning in EHRs

- **Basque**
  - Spoken by the 27% of Basques (714,136 out of 2,648,998)
    - 663,035 in the Spanish part
    - 51,100 in the French part
  - Basque is a minority language in its standardization process and persists between two powerful languages, Spanish and French
  - Nowadays, co-official in some parts, during centuries out of educational systems, media, and industrial environments

  Written use of the Basque Language in the bio-sanitary system and in EHRs is low but co-official

---

**Figure 1:** Classification of dialects in Basque
(Koldo Zuazo, 2008)
Introduction. SNOMED CT and Basque

- **SNOMED Clinical Terms (SNOMED CT)**
  - Comprehensive, multilingual *clinical healthcare terminology*
  - Enables consistent representation of meaning in EHRs

- **Basque**
  - Spoken by the 27% of Basques (714,136 out of 2,648,998)
    - 663,035 in the Spanish part
    - 51,100 in the French part

- Basque is a minority language in its standardization process and persists between two powerful languages, Spanish and French
- Nowadays, co-official in some parts, during centuries out of educational systems, media, and industrial environments

Written use of the Basque Language in the bio-sanitary system and in EHRs is low but co-official

Figure 1: Classification of dialects in Basque (Koldo Zuazo, 2008)
Introduction. SNOMED CT and Basque

- SNOMED Clinical Terms (SNOMED CT)
  - Comprehensive, multilingual clinical healthcare terminology
  - Enables consistent representation of meaning in EHRs

- Basque
  - Spoken by the 27% of Basques (714,136 out of 2,648,998)
    - 663,035 in the Spanish part
    - 51,100 in the French part

- Basque is a minority language in its standardization process and persists between two powerful languages, Spanish and French
- Nowadays, co-official in some parts, during centuries out of educational systems, media, and industrial environments

Written use of the Basque Language in the bio-sanitary system and in EHRs is low but co-official

Figure 1: Classification of dialects in Basque (Koldo Zuazo, 2008)
Introduction. SNOMED CT and Basque

- SNOMED Clinical Terms (SNOMED CT)
  - Comprehensive, multilingual **clinical healthcare terminology**
  - Enables consistent representation of meaning in EHRs

- Basque
  - Spoken by the 27% of Basques (714,136 out of 2,648,998)
    - 663,035 in the Spanish part
    - 51,100 in the French part

- Basque is a **minority language** in its **standardization** process and persists between two powerful languages, Spanish and French
- Nowadays, co-official in some parts, during centuries out of educational systems, media, and industrial environments

 binnen

Written use of the Basque Language in the bio-sanitary system and in EHRs is low but co-official
Introduction. SNOMED CT and Basque

- **SNOMED Clinical Terms (SNOMED CT)**
  - Comprehensive, multilingual *clinical healthcare terminology*
  - Enables consistent representation of meaning in EHRs

- **Basque**
  - Spoken by the 27% of Basques (714,136 out of 2,648,998)
    - 663,035 in the Spanish part
    - 51,100 in the French part

  - Basque is a *minority language* in its *standardization* process and persists between two powerful languages, Spanish and French
  - Nowadays, co-official in some parts, during centuries out of educational systems, media, and industrial environments

  Written use of the Basque Language in the bio-sanitary system and in EHRs is low but co-official

---

*Figure 1: Classification of dialects in Basque*  
(Koldo Zuazo, 2008)
Introduction. Goals

Goals

- To offer a medical terminology in Basque to the bio-medical personnel to try to enforce the use of Basque in the bio-sanitary area
- To be able to access multilingual medical resources in Basque language

How?

Semi-automatically translating the terminology content of SNOMED CT
Introduction. Goals

Goals

- To offer a medical terminology in Basque to the bio-medical personnel to try to enforce the use of Basque in the bio-sanitary area
- To be able to access multilingual medical resources in Basque language

How?

Semi-automatically translating the terminology content of SNOMED CT
Outline

1 Introduction

2 SNOMED CT

3 Translation Algorithm
   - Phase 1: Lexical Resources
   - Phase 2: Finite State Transducers and Biomedical Affixes

4 Results

5 Conclusions
SNOMED CT: core terminology for electronic health records with more than 296,000 active concepts and their corresponding terms (> 1 million terms)

Acceptable coverage of the terminology needed to record patients’ conditions (Humphreys et al., 1997)

Description Types:

| Concept: 95575002 - Obstruction of pelviureteric junction |
|-----------------------------------------------------------|
| **Descriptions in English**                                |
| **Description**                                           | **Type**               |
| Obstruction of pelviureteric junction (disorder)           | FSN                    |
| Obstruction of pelviureteric junction                      | Preferred Term         |
| PUJ - Pelviureteric obstruction                            | Synonym                |
| PUO - Pelviureteric obstruction                            | Synonym                |
| Pelviureteric obstruction                                 | Synonym                |
| UPJ - Ureteropelvic obstruction                            | Synonym                |
| Ureteropelvic obstruction                                 | Synonym                |
SNOMED CT

- SNOMED CT: core terminology for electronic health records with more than 296,000 active concepts and their corresponding terms (> 1 million terms)
- Acceptable coverage of the terminology needed to record patients' conditions (Humphreys et al., 1997)
- Description Types:

| Description in English                                                                 | Type          |
|--------------------------------------------------------------------------------------|---------------|
| Obstruction of pelviureteric junction (disorder)                                      | FSN           |
| Obstruction of pelviureteric junction                                                | Preferred Term|
| PUJ - Pelviureteric obstruction                                                      | Synonym       |
| PUO - Pelviureteric obstruction                                                      | Synonym       |
| Pelviureteric obstruction                                                            | Synonym       |
| UPJ - Ureteropelvic obstruction                                                      | Synonym       |
| Ureteropelvic obstruction                                                            | Synonym       |
**SNOMED CT**

- SNOMED CT: core terminology for electronic health records with more than 296,000 active concepts and their corresponding terms (> 1 million terms)
- Acceptable coverage of the terminology needed to record patients’ conditions (Humphreys et al., 1997)
- Description Types:

| Description | Type          |
|-------------|---------------|
| Obstruction of pelviureteric junction (disorder) | FSN Preferred Term |
| Obstruction of pelviureteric junction | Preferred Term |
| PUJ - Pelviureteric obstruction | Synonym |
| PUO - Pelviureteric obstruction | Synonym |
| Pelviureteric obstruction | Synonym |
| UPJ - Ureteropelvic obstruction | Synonym |
| Ureteropelvic obstruction | Synonym |

Louhi 2014 (Gothenburg, Sweden) IXA (http://ixa.si.ehu.es) Translating SNOMED CT into Basque
SNOMED CT

- SNOMED CT: core terminology for electronic health records with more than 296,000 active concepts and their corresponding terms (> 1 million terms)
- Acceptable coverage of the terminology needed to record patients conditions (Humphreys et al., 1997)
- Description Types:

| Concept: 95575002 - Obstruction of pelviureteric junction | |
|----------------------------------------------------------|---|
| **Descriptions in English**                              | **Type**        |
| **Description**                                          | **FSN**         |
| Obstruction of pelviureteric junction (disorder)          | Preferred Term  |
| Obstruction of pelviureteric junction                     | Synonym         |
| PUJ - Pelviureteric obstruction                           | Synonym         |
| PUO - Pelviureteric obstruction                            | Synonym         |
| Pelviureteric obstruction                                 | Synonym         |
| UPJ - Ureteropelvic obstruction                            | Synonym         |
| Ureteropelvic obstruction                                 | Synonym         |
| Hierarchy                        | Semantic Tag | Example                            |
|---------------------------------|--------------|------------------------------------|
| Clinical Finding/disorder       | disorder     | Myocardial infarction              |
|                                 | finding      | Hyperalphenaglobulinaemia          |
| Procedure/intervention          | procedure    | Eye structure transplantation      |
|                                 | regime/therapy | Pulsed electromagnetic energy to shoulder |
| Organism                        | organism     | Pelistega europaea                 |
| Body structure                  | body structure | Supratentorial brain structure     |
|                                 | morphologic abnormality | Acute erythremia                 |
|                                 | cell         | Umbrella cell                      |
|                                 | cell structure | Viral envelope                    |
| Substance                       | substance     | Bacterial agent                    |
| Pharmaceutical/biologic product | product      | Naratriptan                        |
| Qualifier value                 | qualifier value | Perinatal period                 |
| Observable entity               | observable entity | Postvaccination state           |
| Event                           | event        | Flood                              |
| Situation with explicit context | situation    | Mother smokes                      |
| Social context                  | occupation   | Hospital nurse                     |
|                                 | person       | Homosexual parents (family)         |
|                                 | ethnic group | Irish traveller                    |
|                                 | religion/philosophy | Nonconformist religion        |
|                                 | life style   | White collar thief                 |
|                                 | social concept | Upper class economic status       |
|                                 | racial group | American Indian race               |
| Physical object                 | physical object | Cardiac compression board         |
| Specimen                        | specimen     | Lumpectomy breast sample           |
| Environment/ geographical location | environment | Psychiatric intensive care unit |
|                                 | geographic location | Republic of Serbia          |
|                                 | environment/location | Environment or geographical location |
| Linkage concept                 | attribute    | Agent relationship                 |
|                                 | link assertion | Has problem member                |
|                                 | linkage concept | Linkage concept                   |
| Staging and scales              | assessment scale | Lequesne index                    |
|                                 | tumor staging | pM category                         |
|                                 | staging scale | Chest pain rating                  |
| Special concept                 | navigational concept | Enzymes A - L                    |
|                                 | namespace concept | Extension Namespace 1000001         |
|                                 | administrative concept | Appointment                      |
|                                 | special concept | Special concept                    |
| Record artifact                 | record artifact | Family history section             |

Louhi 2014 (Gothenburg, Sweden) IXA (http://ixa.si.ehu.es) Translating SNOMED CT into Basque 8 / 20
Two possible language sources: English and Spanish

We analyzed the RF2, Snapshot distributions dated 31-07-2012 (English) and 30-10-2012 (Spanish)

Analyzed aspects:

- General numbers of FSNs, PTs and Synonyms and their lacks:
  - The number of active concepts is the same: 296,433 (same file)
  - The number of terms in Spanish is smaller: 15,715 concepts lack of PTs and Synonyms

- Length of the terms in each language:
  - English: 6.76% (1 token), 23.28% (2 tokens) and 20.70% (3 tokens)
  - Spanish version: 33.79% (≤ 3 tokens), 66.21% (≥ 4 tokens)

Conclusions:

- The English version is more complete and consistent than the Spanish one
- The terms in the English version are shorter in length and, in consequence, simpler to translate

We decided to choose the English version as source
Two possible language sources: **English** and **Spanish**

We analyzed the RF2, Snapshot distributions dated 31-07-2012 (English) and 30-10-2012 (Spanish)

Analyzed aspects:

▶ General numbers of FSNs, PTs and Synonyms and their lacks:
  ★ The number of active **concepts** is the same: 296,433 (same file)
  ★ The number of **terms** in Spanish is smaller: 15,715 concepts lack of PTs and Synonyms

▶ **Length of the terms in each language:**
  ★ English: 6.76% (1 token), 23.28% (2 tokens) and 20.70% (3 tokens)
  ★ Spanish version: 33.79% (≤ 3 tokens), 66.21% (≥ 4 tokens)

**Conclusions:**

▶ The English version is more complete and consistent than the Spanish one
▶ The terms in the English version are shorter in length and, in consequence, simpler to translate

We decided to choose the English version as source
Two possible language sources: **English** and **Spanish**

We analyzed the RF2, Snapshot distributions dated 31-07-2012 (English) and 30-10-2012 (Spanish)

**Analyzed aspects:**

- **General numbers of FSNs, PTs and Synonyms and their lacks:**
  - The number of active **concepts** is the same: 296,433 (same file)
  - The number of **terms** in Spanish is smaller: 15,715 concepts lack of PTs and Synonyms

- **Length of the terms in each language:**
  - English: 6.76% (1 token), 23.28% (2 tokens) and 20.70% (3 tokens)
  - Spanish version: 33.79% (≤ 3 tokens), 66.21% (≥ 4 tokens)

**Conclusions:**

- The English version is more complete and consistent than the Spanish one
- The terms in the English version are shorter in length and, in consequence, simpler to translate

**We decided to choose the English version as source**
| Hierarchy                          | Semantic Tag (ST) | # FSN  | English version | Semantic Tag (ST) | # FSN |
|-----------------------------------|------------------|--------|----------------|------------------|-------|
| **Clinical Finding/disorder**     | disorder         | 94,242 | trastorno       | trastorno         | 82,725 |
|                                   | finding          | 45,401 | hallazgo        | hallazgo          | 36,625 |
| **Procedure/intervention**        | procedure        | 75,078 | procedimiento   | procedimiento     | 59,411 |
|                                   | regime/therapy   | 3,573  | régimen/terapia | régimen/terapia   | 2     |
|                                   | regime/tratamiento | 2,773 | régimen/tratamiento | régimen/tratamiento | 2,773 |
| **Organism**                      | organism         | 35,870 | organismo       | organismo         | 35,465 |
| **Body structure**                | body structure   | 26,960 | estructura corporal | estructura corporal | 26,747 |
|                                   | morphologic abnormality | 5,259 | anomalía morfológica | anomalía morfológica | 5,082 |
|                                   | cell             | 645    | célula          | célula            | 640   |
|                                   | cell structure   | 513    | estructura celular | estructura celular | 509   |
| **Substance**                     | substance        | 25,834 | sustancia       | sustancia         | 24,918 |
| **Pharmaceutical/biologic product** | product       | 24,379 | producto        | producto         | 23,854 |
| **Qualifier value**               | qualifier value  | 10,134 | calificador     | calificador       | 9,570 |
| **Observable entity**             | observable entity | 9,044   | entidad observable | entidad observable | 8,602 |
| **Event**                         | event            | 8,959  | evento          | evento            | 8,587 |
| **Situation with explicit context** | situation      | 8,716  | situación       | situación         | 5,785 |
| **Social context**                | occupation       | 6,460  | ocupación       | ocupación         | 4,650 |
|                                   | person           | 688    | persona         | persona           | 432   |
|                                   | ethnic group     | 366    | grupo étnico    | grupo étnico      | 283   |
|                                   | religion/philosophy | 227  | religión/filosofía | religión/filosofía | 217   |
|                                   | life style       | 30     | estilo de vida  | estilo de vida    | 25    |
|                                   | social concept   | 27     | contexto social | contexto social   | 26    |
|                                   | racial group     | 21     | grupo racial    | grupo racial      | 19    |
| **Physical object**               | physical object  | 5,148  | objeto físico   | objeto físico      | 4,747 |
| **Specimen**                      | specimen         | 1,455  | espéctimen      | espéctimen        | 1,386 |
| **Environment**                   | environment      | 1,253  | medio ambiente  | medio ambiente     | 1,162 |
| **geographical location**         | geographic location | 619 | localización geográfica | localización geográfica | 619 |
|                                   | environment/location | 1 | medio ambiente/localización | medio ambiente/localización | 1 |
| **Linkage concept**               | attribute        | 1,157  | atributo        | atributo          | 1,145 |
|                                   | link assertion   | 8      | relación asertiva | relación asertiva | 8    |
|                                   | linkage concept  | 1      | concepto de enlace | concepto de enlace | 1    |
| **Staging and scales**            | assessment scale | 1,125  | escala de evaluación | escala de evaluación | 1081 |
|                                   | tumor staging    | 261    | estadificación tumoral | estadificación tumoral | 249 |
|                                   | staging scale    | 41     | escala de estadificación | escala de estadificación | 16 |
| **Special concept**               | navigational concept | 732 | concepto para navegación | concepto para navegación | 725 |
|                                   | namespace concept | 153    | espacio de nombres | espacio de nombres | 153 |
|                                   | administrative concept | 80  | concepto administrativo | concepto administrativo | 31 |
|                                   | special concept  | 31     | concepto especial | concepto especial | 1 |

Louhi 2014 (Gothenburg, Sweden)
Outline

1 Introduction

2 SNOMED CT

3 Translation Algorithm
   - Phase 1: Lexical Resources
   - Phase 2: Finite State Transducers and Biomedical Affixes

4 Results

5 Conclusions
Translation Algorithm

Incremental approach

The design is for any language pair but some linguistic resources needed for source and objective languages

Our implementation:
- Input: 1 term in English
- Output: \( \geq 1 \) equivalent terms in Basque

The algorithm is applied at term-level (not at concept-level)

Algorithm: 4 phases
- The first 2 phases: developed and evaluated (quantitatively)
- Last 2 phases: in the very near future
Translation Algorithm

Phase 0: Mapping of ICD-10

- Semi-automatic mapping between SNOMED CT and the ICD-10 (IHTSDO)
- By identifying the sense of a concept in SNOMED CT, the best semantic space in the ICD-10 for this concept is searched
- The corresponding Basque term for some of the SNOMED CT concepts is obtained through ICD-10
- To take into account:
  - At concept level, not at term level ⇒ Before executing the algorithm implementation
  - Different purposes: ICD-10 for classification and SNOMED CT for representation
- Fruitful for very specialised terms
Translation Algorithm

1. phase: Lexical Knowledge

- ItzulDB (XML): initialized with all the lexical resources available + the pairs generated in the translation process
- Dictionaries of the bio-medical domain and the ICD-10 classification

Example:

Input term: Deoxyribonucleic acid
Steps in figure number: 1, 2, 4
Translation: Azido desoxirribonukleiko, ADN, DNA
Translation Algorithm

2. phase: Morpheosemantics

- A term is analyzed at word-level and generation-rules are used to create the translation.
- We apply medical suffix and prefix equivalences and morphotactic rules, as well as transcription rules.

Example:

**Input term:** Photodermatitis  
**Steps in figure number:** 3, 5, 7, 6, 4  
**Applied rules:**  
- Identified parts: photo+dermat+itis  
- Translated parts: foto+dermat+itis  
**Translation:** Fotodermatitis
Translation Algorithm

1. Search the term in the translation pairs DB.
2. Is there any Basque term found?
   - Yes
   - No
3. Make a word-level analysis of the term.
4. Generate the Basque term(s).
5. Make a shadow-syntax analysis of the term.
6. Is there any generation-rule applied?
   - Yes
   - No
7. Is there any syntactic rule applied?
   - Yes
   - No
8. Store the Basque term(s).
9. Use Automatic Translator.

3. phase: Shallow Syntax (future)
   - Chunk-level generation rules
   - Hypothesis: some chunks will appear in ItzulDB

Example:

Input term: Deoxyribonucleic acid sample
Steps in figure number: 8, 9, 10, 6, 4
Chunks in ItzulDB:
1st chunk: Deoxyribonucleic acid
   Basque: azido desoxirribonukleiko, ADN, DNA
2nd chunk: sample
   Basque: lagin
Translation: Azido desoxirribonukleikoaren lagin, ADN lagin, DNA lagin
Translation Algorithm

4. phase: Machine Translation (future)

- Aim: to adapt a rule-based automatic translation system called Matxin (Mayor et al., 2011) to the medical domain

Example:

- **Input term:** Partial excision of oesophagus and interposition of colon
- **Steps in figure number:** 12, 4
- **Translation:** Esofagoaren zati baten excisiona eta interpositiona bi puntua
Translation Algorithm

1. Search the term in the translation pairs DB
2. Is there any Basque term found?
3. No
4. Store the Basque term(s)
5. Make a word-level analysis of the term
6. Generate the Basque term
7. Is there any generation-rule applied?
8. No
9. Make a shadow-syntax analysis of the term
10. Yes
11. Is there any syntactic rule applied?
12. Use Automatic Translator

Feedback

- All the processes finish in step 4
- The Basque equivalents with their original English terms are stored in an XML document that follows the TermBase eXchange
- Itzu1DB (lexical resources) is enriched with the translation pairs generated that overcome a confidence threshold ⇒ Help in the translation of new terms
Phase 1: Lexical Resources (English-Basque pairs)

Resources used to initialize ItzuIDB

- **ZT Dictionary**: Science and technology (medicine, biochemistry, biology...). 13,764 English-Basque equivalences
- **Nursing Dictionary**: 5,393 entries
- **Glossary of Anatomy**: Anatomical terminology used by University experts in their lectures. 2,578 useful entries
- **ICD-10**: translated into Basque in 1996. Also available in English and in Spanish. 7,061 equivalences
- **EuskalTerm**: Terminology bank contains 75,860 entries from which 26,597 are from the biomedical domain
- **Elhuyar Dictionary**: English-Basque dictionary. 39,164 equivalences
Phase 2: Finite State Transducers and Biomedical Affixes

- FSTs used to identify the affixes in English Medical terms and by means of affix translation pairs, to generate the equivalent terms in Basque

**Input term:** symphysiolysis

**Identified affixes:** sym+physio+lysis, sym+physi+o+lysis

**Translation of the affixes:** sim+fisio+lisi, sim+fisi+o+lisi

**Morphotactics output term:** sinfisiolisi

**First approach (Perez-de-Viñaspre et al., 2013):**
- 826 prefixes and 143 suffixes with medical meanings manually translated
- Evaluation: Gold Standard of 885 English-Basque pairs: precision of 93% and recall of 41%
- Only SNOMED CT terms for which all the prefixes and suffixes were identified were translated
  - For instance, the “hypophosphatemia” was not translated
  - “hypo”, “phos” and “emia” affixes identified
  - But “phat” not identified

**Current approach:**
- We have increased the number of affixes and transcription rules
- New numbers: 1,703 prefixes and 630 suffixes and 40 rules for transcription
- We are able to translate terms even though all their parts are not identified
- We now translate “hypophosphatemia” into “hipofosfatemia”
Outline

1 Introduction

2 SNOMED CT

3 Translation Algorithm
   - Phase 1: Lexical Resources
   - Phase 2: Finite State Transducers and Biomedical Affixes

4 Results

5 Conclusions
Results in the translation: Dictionary matching and morphosemantics.

- **Phase 1: Dictionary matching**
  - Evaluation in terms of **quantity**, not of **quality**
  - Dictionaries manually generated by lexicographers. The quality is assumed

- **Phase 2: Morphosemantics**
  - 93% precision and 41% recall
  - # Syn: The number of obtained Basque terms
  - # Matches: The number of English terms translated
  - The same input terms may have synonyms or even the same equivalent term given by different dictionaries.
    Example: “alopatia” obtained in ZT and Nursing.

| Disorder | Finding | Body Structure | Procedure |
|----------|---------|---------------|-----------|
|          | #Syn    | #Matches      | #Syn      | #Matches |
| ICD-10 mapping | 11,227 | - | 1,878 | - | 0 | - |
| In dictionaries | 4,804 | 3,488 | 1,836 | 915 | 5,896 | 2,992 | 778 | 473 |
| ZT Dictionary | 1,104 | 883 | 367 | 311 | 1,812 | 1,212 | 293 | 253 |
| Nursing Dictionary | 437 | 350 | 340 | 245 | 978 | 725 | 199 | 157 |
| Glossary of Anatomy | 3 | 3 | 10 | 8 | 1,982 | 1,431 | 2 | 2 |
| ICD-10 | 2,434 | 2,308 | 216 | 195 | 410 | 370 | 5 | 4 |
| EuskalTerm | 906 | 596 | 442 | 306 | 2,346 | 1,423 | 202 | 155 |
| Elhuyar | 299 | 135 | 956 | 300 | 1,090 | 367 | 270 | 91 |
| Morphosemantics | 2,620 | 2,184 | 705 | 578 | 970 | 779 | 1,551 | 1,362 |
| Total | 17,627 | 5,672 | 4,419 | 1,493 | 6,866 | 3,771 | 2,329 | 1,835 |
## Overall Results

|                     | Disorder | Finding | Body Structure | Procedure |
|---------------------|----------|---------|----------------|-----------|
| Translated Concepts | 14,125   | 2,777   | 3,231          | 1,502     |
| Concepts in total   | 65,386   | 33,204  | 31,105         | 82,069    |
| Percentage          | 21.60%   | 8.36%   | 10.39%         | 1.83%     |

- **Disorder**: 21.60% of the translated. Good. Thanks to the ICD-10 (11,227 synonyms) and morphosemantics (81.53% of the simple terms)
- **Finding**: the most balanced
- **Body Structure**: the Glossary of Anatomy only contributes in this hierarchy (previous table)
- **Procedure**: dictionaries do not help much, in contrast, morphosemantics contribution allows to translate the 87.84% of the simple terms
Outline

1 Introduction

2 SNOMED CT

3 Translation Algorithm
   - Phase 1: Lexical Resources
   - Phase 2: Finite State Transducers and Biomedical Affixes

4 Results

5 Conclusions
Conclusions

- We have designed a translation algorithm for the multilingual terminology content of SNOMED CT and we have implemented the first two phases
  1. Lexical resources feed our database
  2. Basque equivalents are generated using transducers and medical and biological affixes
- Dictionaries provide Basque equivalents of any term length while transducers get as input unique token terms
- Results are provided for the most populated hierarchies are shown even though both methods are applied for all the hierarchies in SNOMED CT
- Results are promising. We obtained the equivalents in Basque of 21.60% of the disorders
- Future Work:
  - Specialist in medical terminology can check the quality of the obtained terms and correct them
  - Implement the remainder of the phases in the algorithm: Shallow Syntax and Machine Translation
Translating SNOMED CT Terminology into a Minor Language

Olatz Perez-de-Viñaspre and Maite Oronoz

University of the Basque Country

IXA Taldea