

# KabiTermICD: Nested Term Based Translation of the ICD-10-CM into a Minor Language

Olatz Perez-de-Vinaspre, Maite Oronoz and  
Natalia Elvira



# Motivation

# Basque

- Spoken in the Basque Country
- +700,000 speakers
- Official language in some regions
- Coexist with Spanish and French
- It has not related languages





# ICD-10-CM

International Classification of  
Diseases, Tenth Edition,  
Clinical Modification

## J12

### Viral pneumonia, not elsewhere classified

**Incl.:** bronchopneumonia due to viruses other than influenza viruses

**Excl.:** congenital rubella pneumonitis ([P35.0](#))

pneumonia:

- aspiration (due to):
  - NOS ([J69.0](#))
  - anaesthesia during:
    - labour and delivery ([O74.0](#))
    - pregnancy ([O29.0](#))
    - puerperium ([O89.0](#))
  - neonatal ([P24.9](#))
  - solids and liquids ([J69.-](#))
- in influenza ([J09](#), [J10.0](#), [J11.0](#))
- interstitial NOS ([J84.9](#))
- lipid ([J69.1](#))
- viral, congenital ([P23.0](#))

severe acute respiratory syndrome [SARS] ([U04.9](#))

#### J12.0

Adenoviral pneumonia

#### J12.1

Respiratory syncytial virus pneumonia

#### J12.2

Parainfluenza virus pneumonia

#### J12.3

Human metapneumovirus pneumonia

#### J12.8

Other viral pneumonia

#### J12.9

Viral pneumonia, unspecified



# Translation of ICD-10-CM into Basque

- Osakidetza: Public Health System of the Basque Autonomous Community
- 93,830 codes/descriptions
- Original ICD-10 version translated into Basque but not updated (+12,000 codes)
- Low personal resources
- Low bilingual corpora



# Translation of SNOMED CT into Basque

Previous work on automatic translation of SNOMED CT into Basque:

- Bilingual/multilingual mapping of lexical resources
- NeoTerm: morpheme based translation of neoclassical terms
- KabiTerm: nested term based translation of complex terms
- MatxinMed: adaptation of Matxin (general RBMT) into medical domain



# Nested terms

elbow Body structure  
left Qualifier elbow Body structure  
fracture Disorder of left elbow Body structure  
open Qualifier fracture of left elbow Disorder

# KabiTermICD



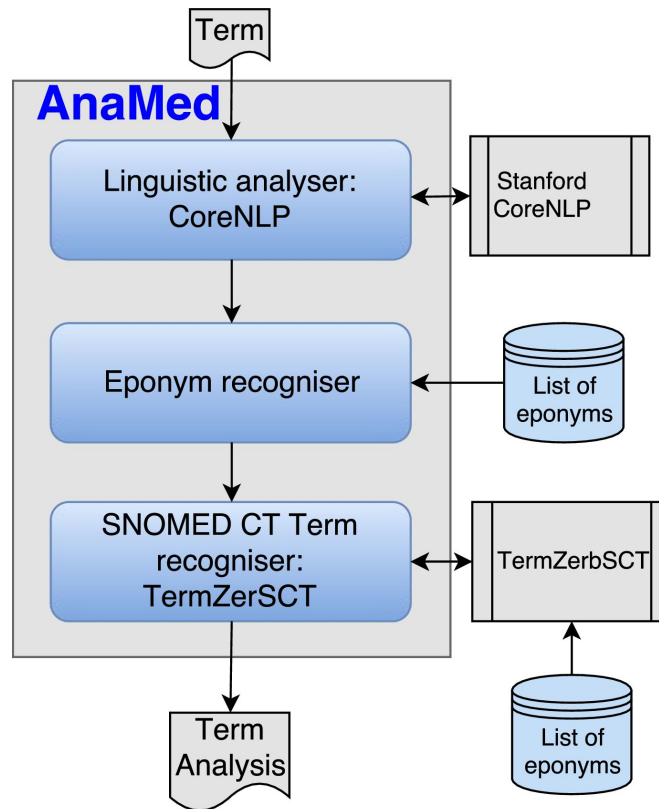


# KabiTermICD

- Finite State Transducers (Foma)
- Rule-Based term translator
- English -> Basque

AnaMed analyser for preprocessing

# AnaMed



AnaMed analyser's architecture



# AnaMed: Linguistic analyser

## Stanford CoreNLP

### Tools used:

- Tokeniser
- Morphological analyser and tagger
- Named Entity Recognizer

### Information extracted:

- Tokens
- Offsets
- Lemma
- Part Of Speech
- Named Entity



## AnaMed: **Eponym recogniser**

From Wikipedia:

“An **eponym** is a person, place, or thing after whom or after which something is named, or believed to be named.”

**Down** syndrome  
**Alzheimer's** disease



AnaMed:

# SNOMED CT term recogniser

Systematized NOmenclature of  
MEDicine - Clinical Terms

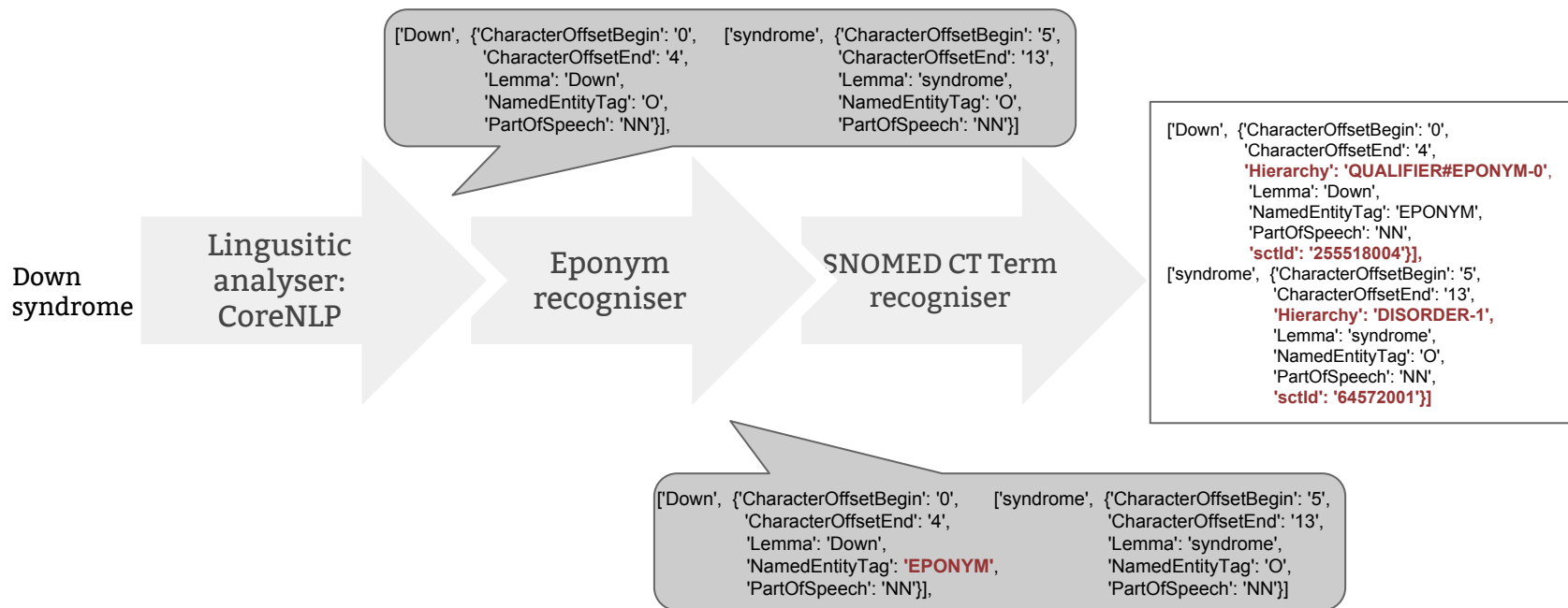
SNOMED CT:

- + 300,000 concepts
- + 1,000,000 descriptions

Term recogniser:

- Nested terms inside the main term
- open fracture of left arm
  - open fracture: Disorder
  - left arm: Body Structure

# AnaMed: example



# KabiTermICD





# KabiTermICD's translation process

malignant neoplasm of small intestine, unspecified

1. **AnaMed analyses the term and prepares the nested term groupings**  
malignant\_neoplasm of small intestine , unspecified
2. **Identification of Basque translation patterns and tagging nested terms**  
malignant\_neoplasm|DIS+a small |QUA intestine|BOD+areM , unspecified|ATZ &Last2First
3. **Rearrangement of the nested terms**  
small |QUA intestine|BOD+areM malignant\_neoplasm|DIS+a , unspecified|ATZ
4. **Translation of the nested terms into Basque**  
txiki&&&ADJK heste+areM neoplasia\_gaizto+a , zehaztugabea
5. **Rearrangement of the adjectives and determination of plurals**  
heste txiki+areM neoplasia\_gaizto+a , zehaztugabea
6. **Addition of declension to the nested terms**  
heste txikiaren neoplasia gaiztoa, zehaztugabea

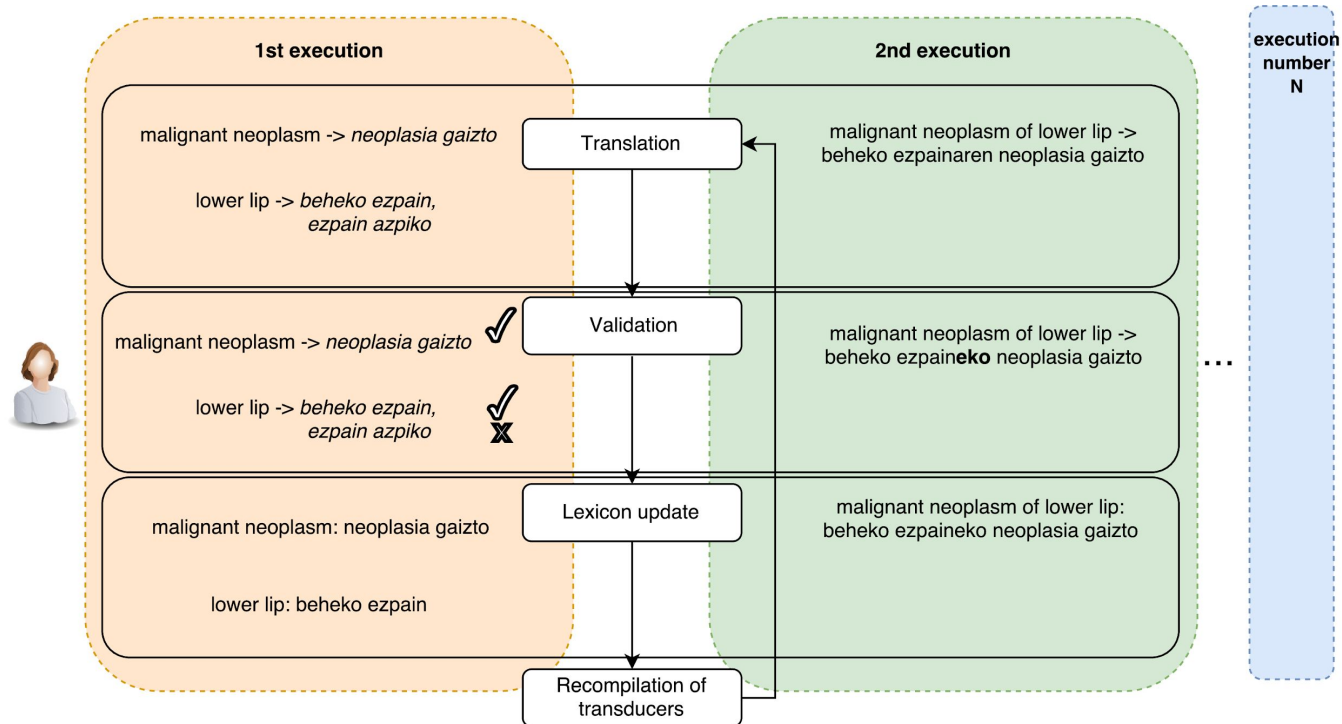


# KabiTermICD's transducers

1. AnaMed analyses the term and prepares the nested term groupings
2. **Identification of Basque translation patterns and tagging nested terms**
3. Rearrangement of the nested terms
4. **Translation of the nested terms into Basque**
5. Rearrangement of the adjectives and determination of plurals
6. **Addition of declension to the nested terms**

# Translation of ICD-10-CM

# Translation of ICD-10-CM





# Interface for the validation

**GNS10 KODEA:** C004

## **INGELESEZKOAK**

**Hobetsia:** Malignant neoplasm of lower lip, inner aspect

## **GAZTELANIAZKOAK**

**Hobetsia:** Neoplasia maligna de labio inferior, cara interna

## **EUSKARAZKOAK**



Aukeratua

beheko ezpaineiko neoplasia gaiztoa, barne itxura



Egokia



Errepasatzeko

**Ballabidea:** |pat\_gn\_14



Aukeratua

beheko ezpaineiko neoplasia gaiztoa, barrualdea



Egokia



Errepasatzeko

# Results

# 60%

Of not postedited descriptions  
(3,379)

# Edit distance of the post-edited terms

Edit distance	Quantity	Percentage	Edit distance	Quantity	Percentage
<b>1</b>	109	4.77	<b>10-14</b>	225	9.85
<b>2</b>	39	1.71	<b>15-19</b>	198	8.67
<b>3</b>	216	9.46	<b>20-24</b>	67	2.93
<b>4</b>	986	43.17	<b>25-29</b>	38	1.66
<b>5</b>	74	3.24	<b>30-34</b>	13	0.57
<b>6</b>	59	2.58	<b>35-39</b>	9	0.39
<b>7</b>	63	2.76	<b>40-44</b>	3	0.13
<b>8</b>	113	4.95	<b>45</b>	1	0.04
<b>9</b>	71	3.11	<b>Total</b>	2,284	100



# Error types and examples

Edit distance	Error type	Source	Correction
1	Spelling error	<i>descemetozelea, eskuineko begia</i>	<i>deszemetozelea, eskuineko begia</i>
1	Singular article overproction	<i>atrofia optikoa primarioa, aldebikoa</i>	<i>atrofia optiko primarioa, aldebikoa</i>
2	Incorrect plural mark	<i>beste parafili<b>ak</b> batzuk</i>	<i>beste parafilia batzuk</i>
3-4	Incorrect genitive mark	<i>kornea<b>ren</b> neoplasia gaizto</i>	<i>kornea<b>ko</b> neoplasia gaizto</i>
5+	Error combination	<i>orbita<b>ren</b> zelulitis<b>ak</b></i>	<i>orbita<b>ko</b> zelulitis</i>
5+	Incorrect nested term	<i><b>urradura</b>, eskuineko aldaka</i>	<i><b>abrasioa</b>, eskuineko aldaka</i>

# Conclusions



## Conclusions

- KabiTermICD is a Rule-Based translator for ICD-10-CM descriptions into Basque
- 60% of descriptions with no post-edition need
- No previous work (known) on automatic translation of ICD-10-CM
  - New promising path to improve expert's translations work
- Solution for less resourced languages without enough corpus

# Thank you!

Eskerrik asko!