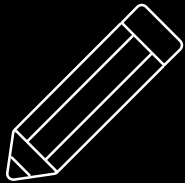


# Basque, NLP and Clinical domain



Ixa taldea





# Kaixo!

I'am **Olatz Perez de Viñaspre**

PhD in Computer Science and  
Researcher and lecturer at UPV/EHU's  
IXA NLP research group

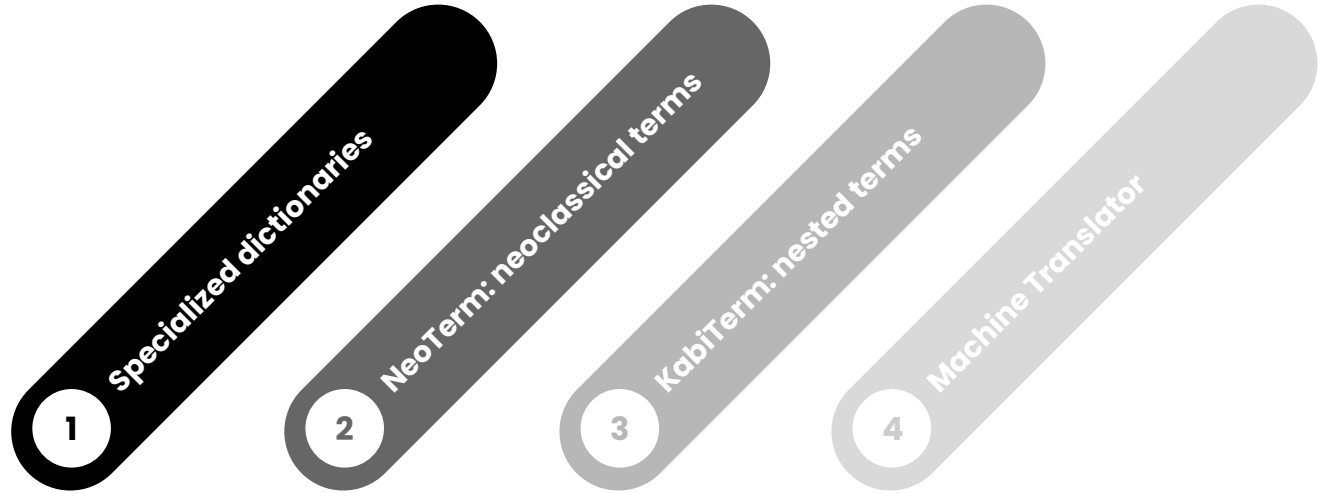


1

# **Terminology translation**



# Process



# Specialized dictionaries

Source: *Deoxyribonucleic acid*

Equivalents: azido desoxirribonukleiko  
(ZT Hiztegia, Euskalterm)

DNA  
(ZT Hiztegia, Euskalterm)



# Neoclassical terms

*Terms composed from combining forms derived from Latin and Greek roots*

Source-terms: *Radionecrosis*

Translation process:

Identified morphemes: radio+necr+osis

Translated morphemes: radio+nekr+osi

Created term: erradionekrosi



# Nested terms

Source term: *Fracture of elbow*

Translation process:

Identified structure: DISORDER + of + BODYPART

Translated structure: BODYPART+ren + DISORDER

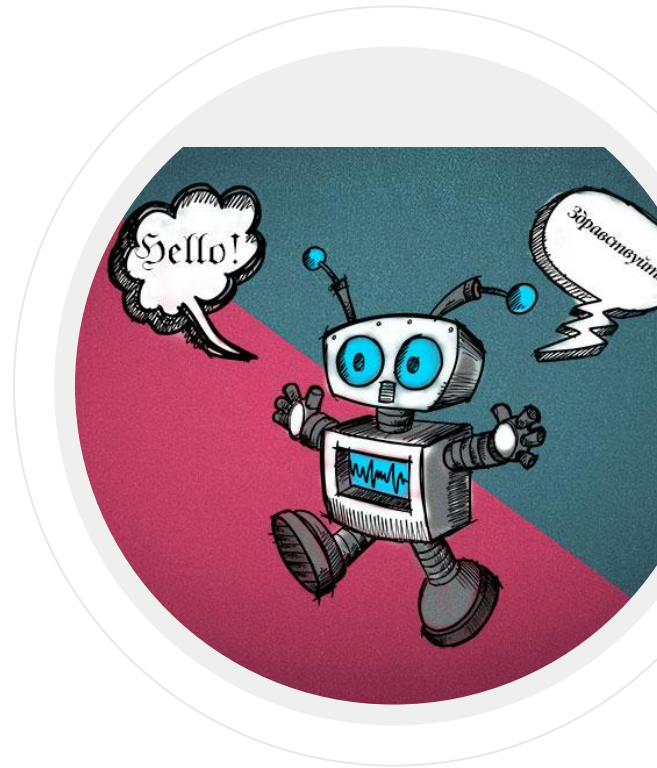
Created term: Ukondoaren haustura



# Machine Translation

Source term: *Lymphoma of lower esophagus*

Created term: Beheko esofagoko linfoma





# Working terminologies

## SNOMED CT

Systematized Nomenclature of Medicine - Clinical Terms

+ 300.000 Concepts

+1.000.000 Descriptions (en)

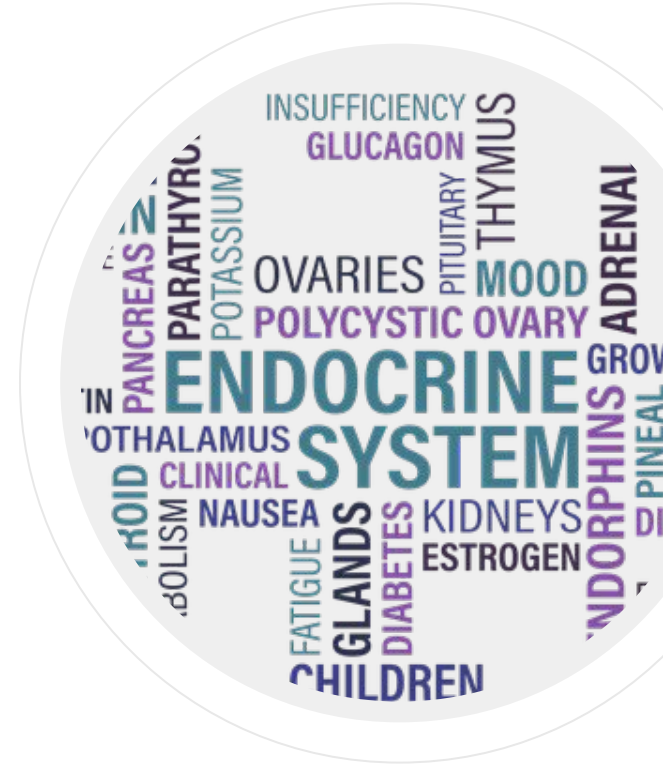
Reference terminology

## ICD-10-CM

International Classification of Diseases - 10th edition - Clinical Modification

+90.000 Concepts/Descriptions

Reference classification



A large white circle is centered on a black background. To its left, there are several overlapping circles of varying shades of gray, with the number '2' in white. To its right, there are several concentric white circles of varying diameters.

**2**

# **Machine translation of health record**



Machine Translation of health records?

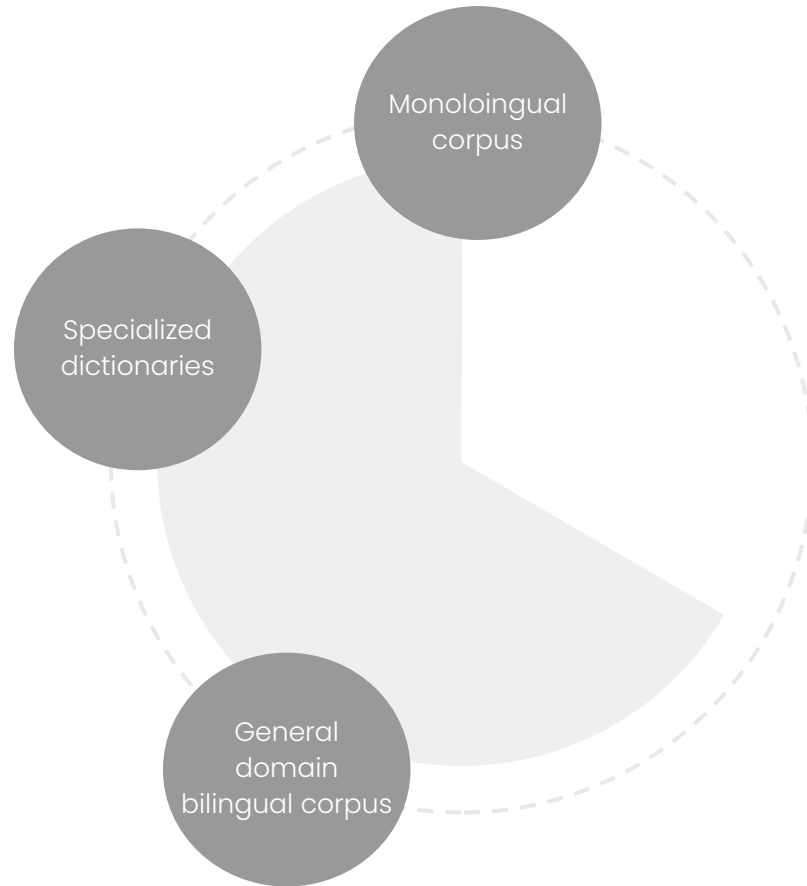


## Basque and Spanish



Corpus?

# Without bilingual clinical records...



# Current

**(eu)** fibrinolisiaren ondoren egoera klinikoa ez da askorik aldatu

**(es)** tras fibrinólisis el estado clínico no ha variado significativamente

*(en) after fibrinolysis the clinical status has not changed significantly*



# Even harder...

**(eu)** hipertentsioaren aurkako tratamenduarekin hasi da, tentsioak neurri egokian mantenduz; hipergluzemiarako joera antzeman da egonaldian

**(es)** se ha iniciado tratamiento antihipertensivo, manteniendo tensiones adecuadas; se ha detectado tendencia a la hiperglucemia durante su estancia





# Not all is gold...

**(eu)** egun batzuetatik hona **eginkariak**  
beltzak direla iruditu zaio

**(es)** desde unos días, se le ha parecido que  
son negros

*(en) since a few days, it has seemed to him  
that they are black*



# Or even worse...

(eu) ondorengo egunetan ez zuen botaka egin, eta **beherakoak** gutxituz eta giltzurrunen funtzioa hobetuz joan ziren

(es) en los días posteriores no presentaba vómitos, con disminución de la **tos** y mejoría de la función renal

(en) in the following days, he did not present vomiting, with a decrease in **cough** and an improvement in kidney function



# Objectives

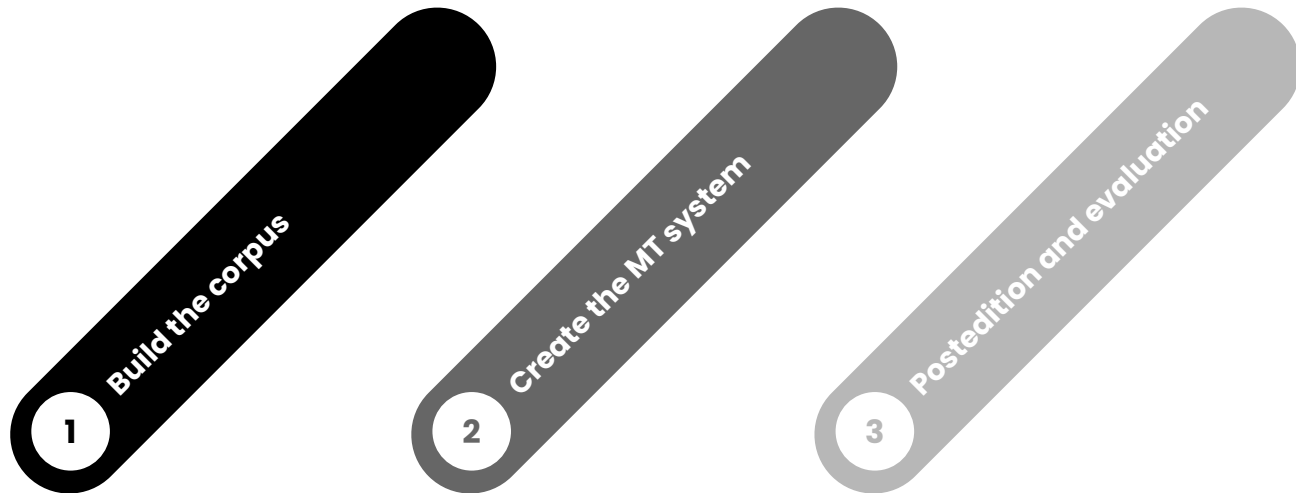
Online MT system for health records

- Basque -> Spanish
  - Backtranslation
- Adapted to:
  - Specialization
  - Record type





# Process



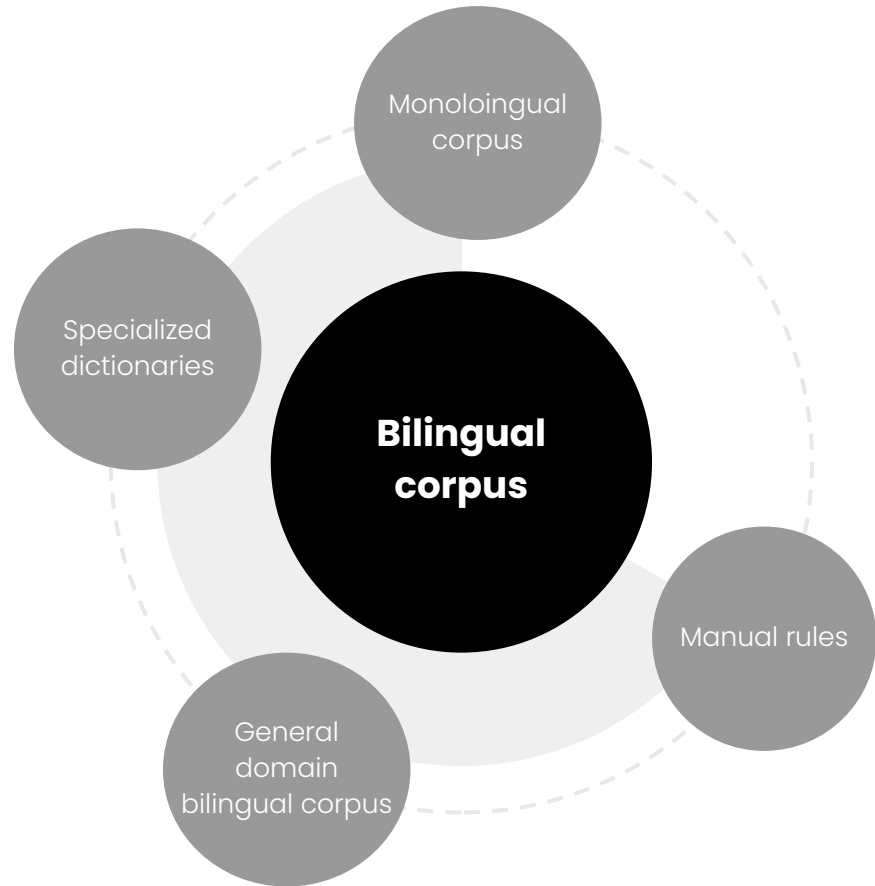
# Help of clinical professionals

Advantages of a expert written corpus:

- Correct and real **style**
- Guarantees of correct translations (**semantics**)
- Use of expert **terminology** (creation of new terms)
- They start to create **habits**



# Resources



# Evaluation

We will reserve 20.000 words from the corpus:

- Manually validated by clinical professionals
- One sentence, many validations



# Postedition

Professionals will also correct the translations

The MT system will learn from those corrections







# Diolch!

**Questions?**

- [olatz.perezdevinaspre@ehu.eus](mailto:olatz.perezdevinaspre@ehu.eus)



# Credits

- Slides template: [SlidesCarnival](#)
- Pictures: [Unsplash](#)

