European Language Grid – META-FORUM 2020
Bernardo Magnini (Fondazione Bruno Kessler, Italy)
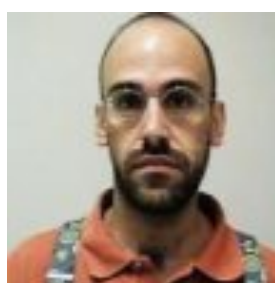
ELG Pilot Project:
European Clinical Case Corpus

Bernardo Magnini
Coordinator

Begona Altuna

Alberto Lavelli

Manuela Speranza

Roberto Zanoli

FBK – Trento, Italy
NLP Research Group

E3C European Clinical Case Corpus

A project aiming at the creation of a corpus of clinical cases in 5 European languages

Start date: July, 6 2020

Duration: 12 months

Languages: Italian, Spanish, French, English, Basque

ELG coach: Khalid Choukri

https://e3c.fbk.eu/

FBK
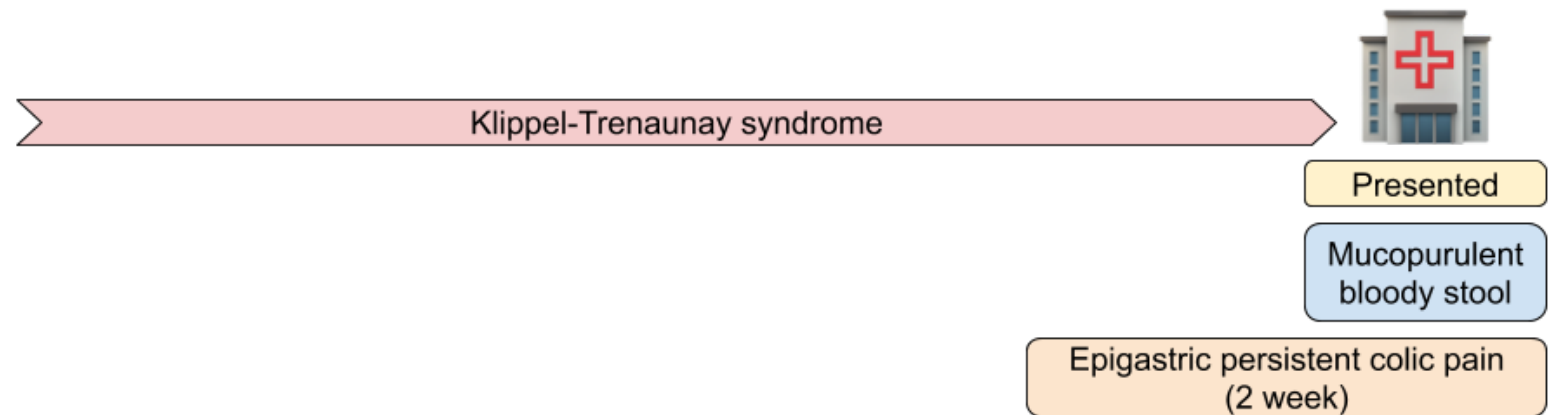FONDAZIONE BRUNO KESSLER

EUROPEAN LANGUAGE GRID

E3C European Clinical Case Corpus

A project aiming at the creation of a corpus of clinical cases in 5 European languages

Three types of annotations:

- **clinical entities**: pathologies, symptoms, etc.

- **temporal information**: events, time expressions and temporal relations;

- **factuality**: event factuality and assessment of the effect of negation, uncertainty and hedge expressions

A 25-year-old man with a history of Klippel-Trenaunay syndrome presented to the hospital with mucopurulent bloody stool and epigastric persistent colic pain for 2 wk.

Klippel-Trenaunay syndrome

Presented

Mucopurulent bloody stool

Epigastric persistent colic pain (2 week)

PubMed

ICD-10

THYME

SNOMED International — Leading healthcare terminology, worldwide

E3C European Clinical Case Corpus

A project aiming at the creation of a corpus of clinical cases in 5 European languages

Annotations Guidelines

E3C Annotation Guidelines. [Deliverable D1](#) (2020)

Data Collection

Language	Clinical cases	Tokens	Tok./doc
Italian	1,323	73K	55.1
English	9,533	928K	97.2
French	1615	548K	339.1
Spanish	1,400	531K	379.27
Basque	122	26K	214.2

First release (CC-BY-NC)

The E3C Corpus: Data collection, planned January 2021

Publications

The E3C Project: Collection and Annotation of a Multilingual Corpus of Clinical Cases" to appear at [CLiC-it 2020](#), the Seventh Italian Conference on Computational Linguistics, Bologna, 1-3 March 2021