

# Tecnologias da Linguagem para a Língua Portuguesa

---

*Workshop sobre Inteligência Artificial e Processamento da Linguagem Natural para a Administração Pública e as PME*  
22 de junho de 2021, online

**António Branco**

Diretor Geral



Infraestrutura de Investigação para a Ciência e Tecnologia da Linguagem

# A Língua Portuguesa...

- 4ª língua com mais falantes no mundo (em 7000)
  - chinês, castelhano, inglês
  - 2021, Camões IP
- 5ª língua mais utilizadores na internet
  - inglês, chinês, castelhano e árabe
  - 31/05/2020, internet world stats
- 6ª língua no Twitter
  - inglês, japonês, castelhano, coreano, árabe
  - 12/03/2018, vicinitas

# A Língua Portuguesa na Era Digital

- **LIVRO BRANCO**

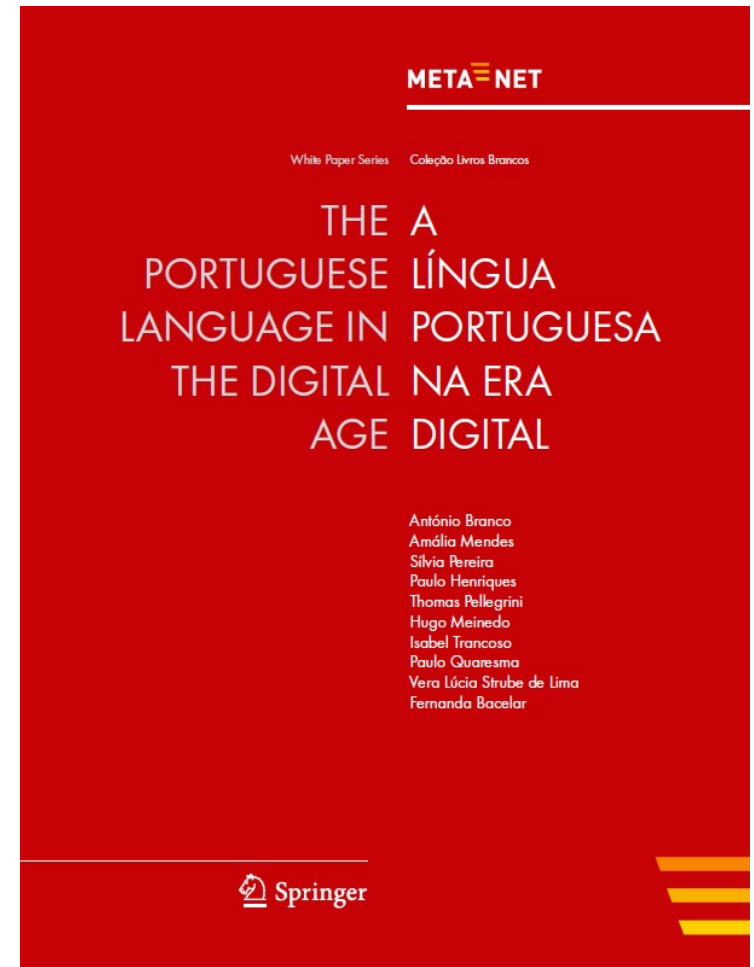
- público em geral, media, decisores, empreendedores

- **Coleção**

- Livros brancos para 30 idiomas por 200+ peritos

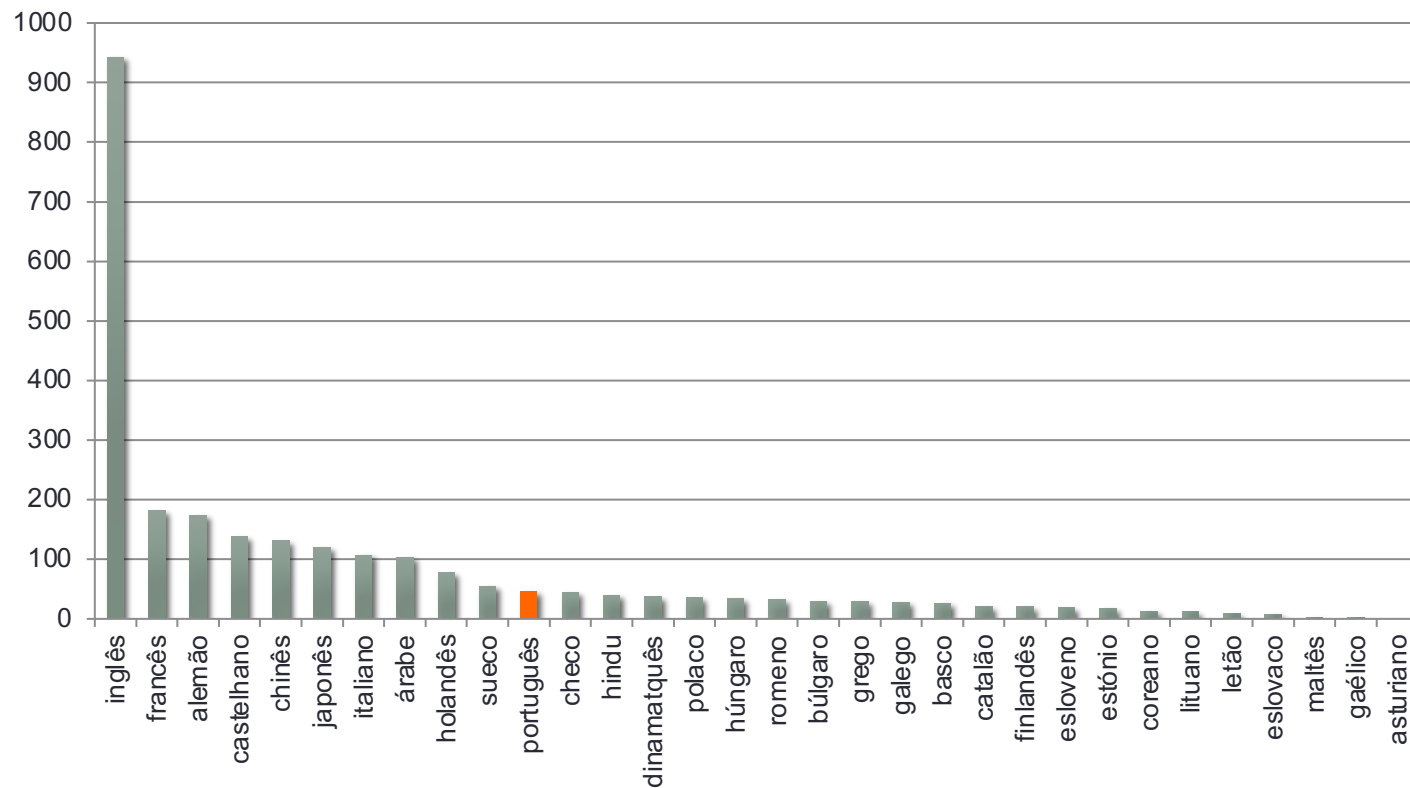
- **Objetivo**

- divulgar o estado de preparação tecnológica da língua portuguesa
- avançar recomendações



# Preparação tecnológica

- Número de referências a recursos linguísticos por língua nas 10 conferências de topo 2010-2012
- Mariani and Frankopoulo, 2012, "Language Matrices and the Language Resource Impact Factor", *PAROLE Seminar*



Falantes do português: 4% PIB mundial  
 Falantes do inglês: ca. 1/3

# Investigação e inovação

- **Comunidade de investigadores**
  - conferências internacionais PROPOR
- **Spinoffs**
  - DefinedCrowd
  - Unbabel
  - ...
- **Infraestrutura de investigação**
  - PORTULAN CLARIN



<https://portulanclarin.net>

**PORTULAN CLARIN**

Infraestrutura de Investigação para a Ciência e Tecnologia da Linguagem

Repositório Bancada Apoio Alcance

pt

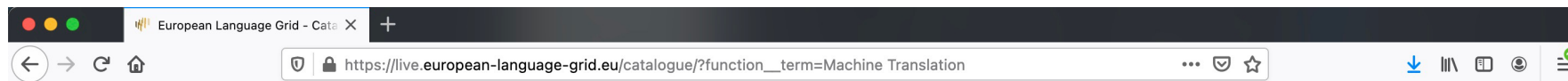
# DISPONIBILIZAR REUTILIZAR IMPULSIONAR

Serviços e dados para investigadores, inovadores, estudantes e profissionais da **linguagem**.

# European Language Grid



- projeto europeu de 3 anos
- consórcio
  - German Research Center for Artificial Intelligence
  - Institute for Language and Speech Processing
  - University of Sheffield
  - Charles University
  - ELDA European Language Distribution Agency
  - Tilde
  - Hensoldt analytics
  - Expert System Iberia
  - University of Edinburgh
- A ELG irá desenvolver e implantar uma plataforma de nuvem escalonável, fornecendo, de uma forma fácil de integrar, acesso a centenas de tecnologias linguísticas comerciais e não comerciais para todas as línguas europeias, incluindo ferramentas e serviços em execução, bem como conjuntos de dados e recursos



RELEASE 2

Technologies Resources Community Events Documentation About ELG

Search


[Clear all filters](#)

Language resources & technologies

Service functions



- Machine Translation (59)

Languages

Licences

Conditions of use for data

59 search results

[Machine Translation](#)


### ILSP Machine Translation Service for English to modern Greek

0.4.0

122 views

172 times used

Functional service

A machine translation (MT) tool for English-to-Greek based on the Marian machine translation framework. The translation model is a basic transformer model trained over a Greek-English corpus of 10.7M sentence pairs using

Keywords: Machine Translation · modern Greek · English · Neural machine translation

Languages: English · Modern Greek (1453-)

Licence: ILSP services Terms of Service



### ILSP Machine Translation Service for modern Greek to English

0.4.0

41 views

186 times used

Functional service

A machine translation (MT) tool for Greek-to-English based on the Marian machine translation framework. The translation model is a basic transformer model trained over a Greek-English corpus of 10.7M sentence pairs using

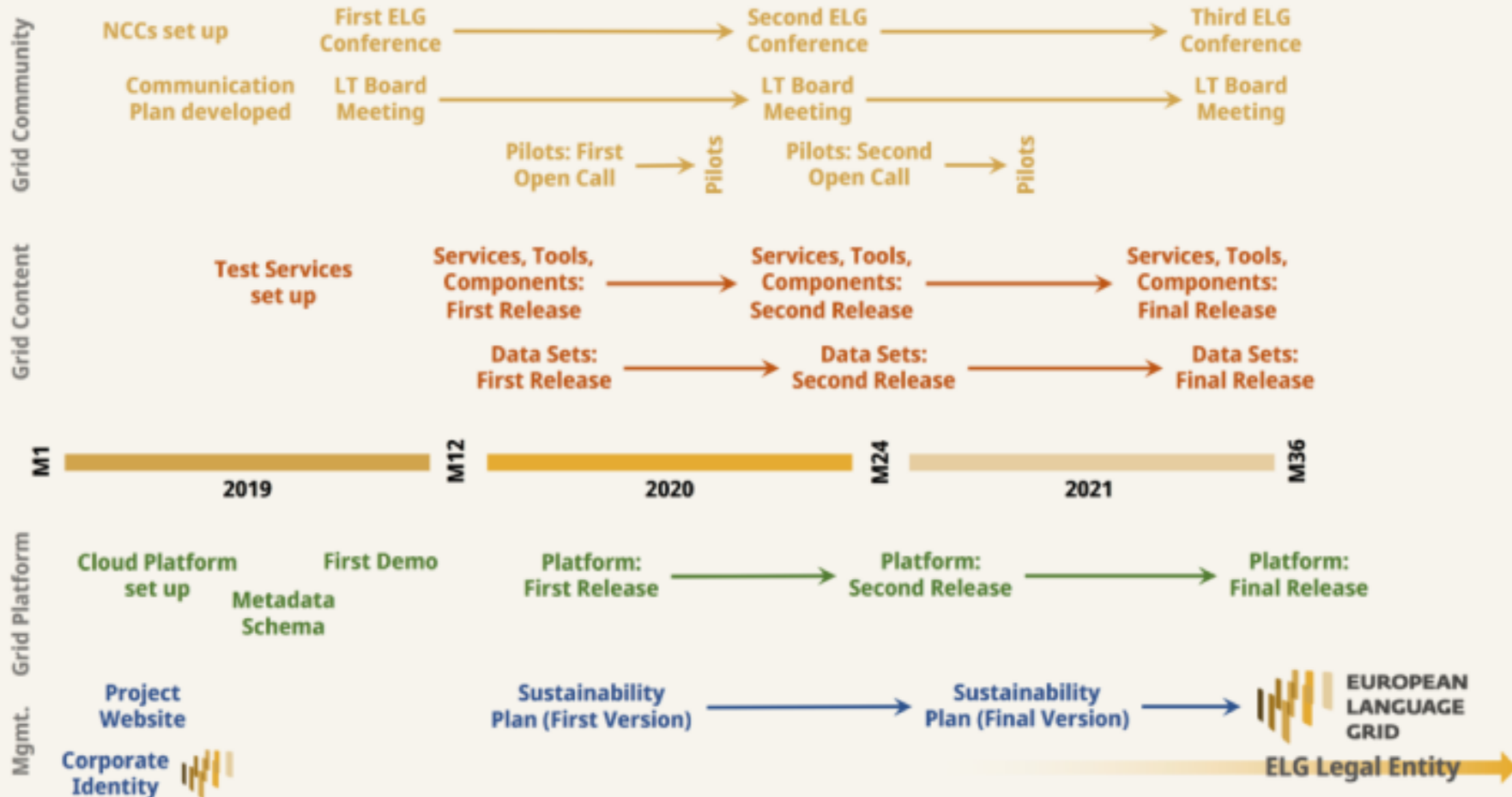
Keywords: Machine Translation · modern Greek · English · Neural machine translation



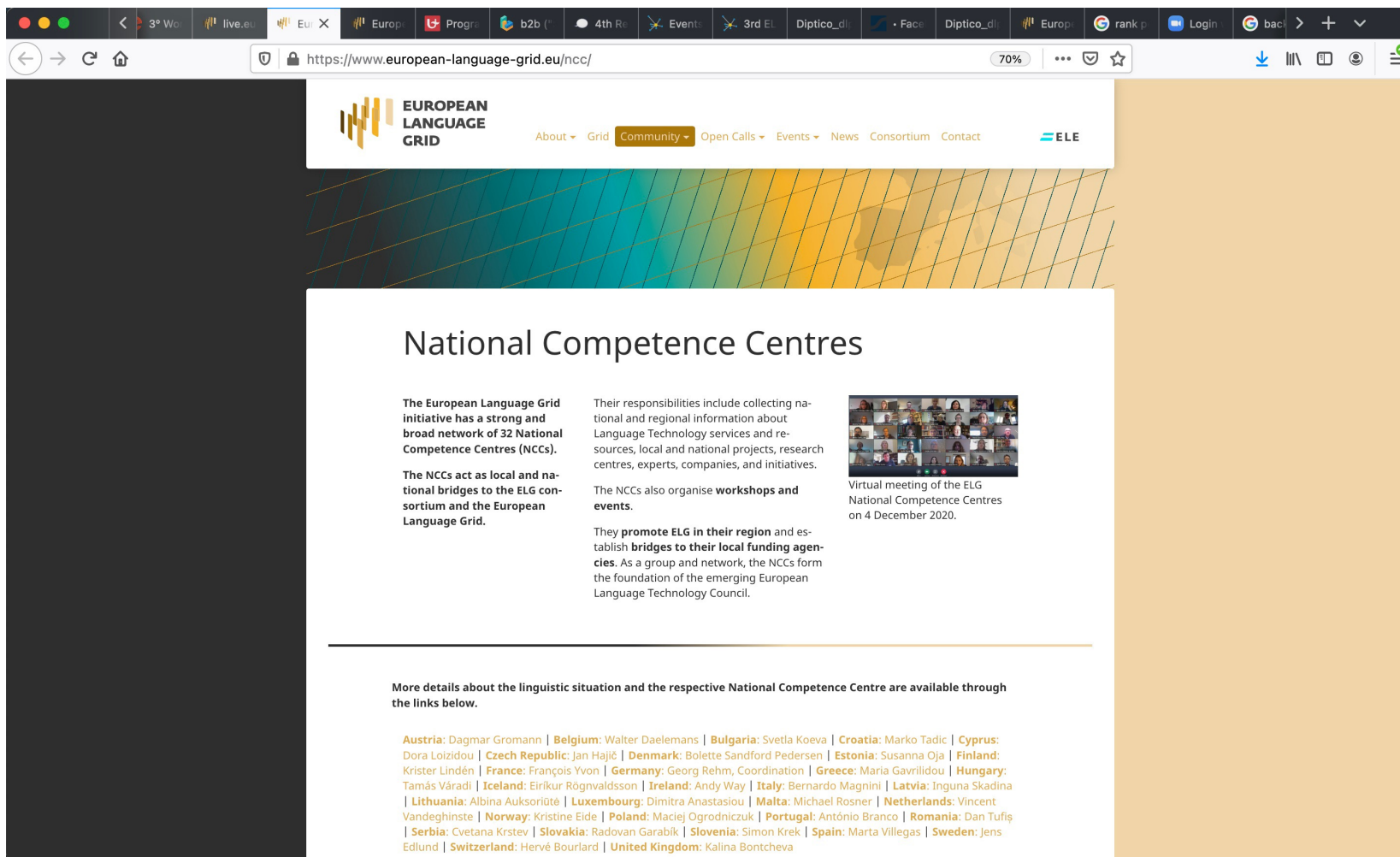
# Plataforma



- Versão 1 lançada no início de maio de 2020.
  - A funcionalidade e o design atuais da interface são limitados; no entanto, a plataforma está sendo constantemente aprimorada e ampliada.
- Versão 2 será disponibilizada no início de 2021
- Versão 3 será lançada no final de 2021.
  - Incluirá mais de 800 serviços LT funcionais e mais de 3500 conjuntos de dados LT, corpora, recursos e modelos.



# Centros nacionais de competência



The screenshot shows a web browser window displaying the website <https://www.european-language-grid.eu/ncc/>. The page features the European Language Grid logo and navigation menu. The main content area is titled "National Competence Centres" and contains several paragraphs of text and a video thumbnail.

**EUROPEAN LANGUAGE GRID** | About | Grid | **Community** | Open Calls | Events | News | Consortium | Contact | ELE

## National Competence Centres


**The European Language Grid initiative has a strong and broad network of 32 National Competence Centres (NCCs).**

**The NCCs act as local and national bridges to the ELG consortium and the European Language Grid.**

Their responsibilities include collecting national and regional information about Language Technology services and resources, local and national projects, research centres, experts, companies, and initiatives.

The NCCs also organise **workshops and events**.

They **promote ELG in their region** and establish **bridges to their local funding agencies**. As a group and network, the NCCs form the foundation of the emerging European Language Technology Council.



Virtual meeting of the ELG National Competence Centres on 4 December 2020.

**More details about the linguistic situation and the respective National Competence Centre are available through the links below.**

**Austria:** Dagmar Gromann | **Belgium:** Walter Daelemans | **Bulgaria:** Svetla Koeva | **Croatia:** Marko Tadic | **Cyprus:** Dora Loizidou | **Czech Republic:** Jan Hajič | **Denmark:** Bolette Sandford Pedersen | **Estonia:** Susanna Oja | **Finland:** Krister Lindén | **France:** François Yvon | **Germany:** Georg Rehm, Coordination | **Greece:** Maria Gavrilidou | **Hungary:** Tamás Váradi | **Iceland:** Eiríkur Rögnvaldsson | **Ireland:** Andy Way | **Italy:** Bernardo Magnini | **Latvia:** Inguna Skadina | **Lithuania:** Albina Aukšoriūtė | **Luxembourg:** Dimitra Anastasiou | **Malta:** Michael Rosner | **Netherlands:** Vincent Vandeghinste | **Norway:** Kristine Eide | **Poland:** Maciej Ogrodniczuk | **Portugal:** António Branco | **Romania:** Dan Tufiş | **Serbia:** Cvetana Krstev | **Slovakia:** Radovan Garabík | **Slovenia:** Simon Krek | **Spain:** Marta Villegas | **Sweden:** Jens Edlund | **Switzerland:** Hervé Bourliard | **United Kingdom:** Kalina Bontcheva

# Preparação tecnológica da língua portuguesa

- **Desafio societal estratégico**
  - língua portuguesa é um dos pilares da **soberania** portuguesa
  - para sobreviver e prosperar na era digital precisa de ser estudada cientificamente e **preparada tecnologicamente**
  - só assim poderemos **aceder a todas as pessoas, serviços e bens** que irão ficando disponíveis na e através da sociedade da informação
  - e assegurar a nós, e à nossa cultura, a **cidadania plena na sociedade do futuro**

# Obrigado.

## António Branco

Diretor Geral



Infraestrutura de Investigação para a Ciência e Tecnologia da Linguagem