

ITALY: COUNTRY PROFILE

Claudia Foti

Italian Public Services National Anchor Point

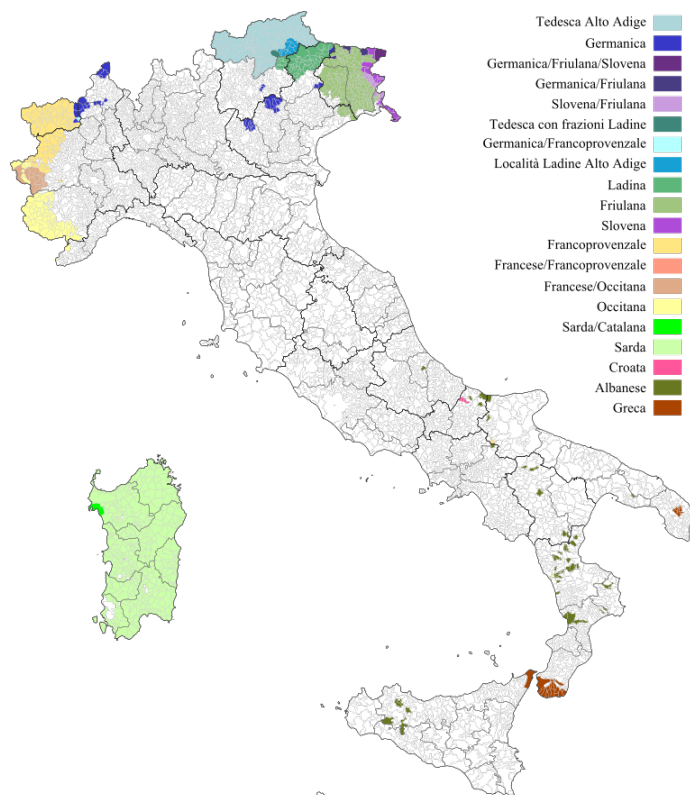
Simonetta Montemagni

Italian Technological National Anchor Point



Connecting
Europe
Facility

- Background – Languages of Italy
- ELRC in Italy
- Public services in the era of Artificial Intelligence: the Italian situation
- Status quo: translation in Italian public institutions
- Future goals



Comunità di minoranza
L. 482/99 e S.A. Prov. BZ/Sudtirolo

- Complex situation
- Italian is the official national language
 - second L1 in Europe (13%)
- Before 1999, four linguistic minorities enjoyed some kind of acknowledgment and protection
- National Law N.482/1999 Norme in materia di tutela delle minoranze linguistiche storiche
 - 12 co-official languages, including local languages spoken by minority populations and regional border languages
 - Albanian, Catalan, **German**, Greek, **Slovene**, Croatian, **French**, Franco-Provençal, Friulian, **Ladin**, Occitan, Sardinian

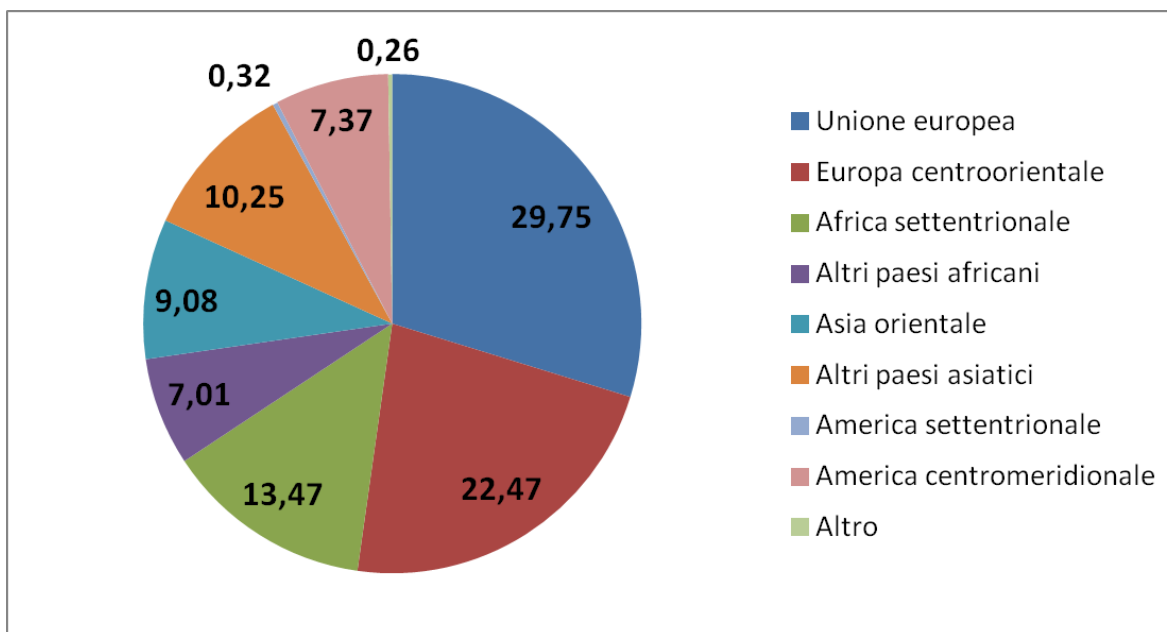
- Immigration countries of recently established linguistic minorities

- Romania 22,6%
- Albania 9,8%
- Morocco 9,0%
- China 5,3%
- Ukraine 4,5%
- Philippines 3,4%
- India 2,9%
- Moldova 2,9%

- Languages involved in main economic and commercial relations

- German, French, English, Chinese, Dutch, Russian, Spanish, Arabic, Polish

Distribution of migrants by origin country



ELRC 1 (2015-2017)

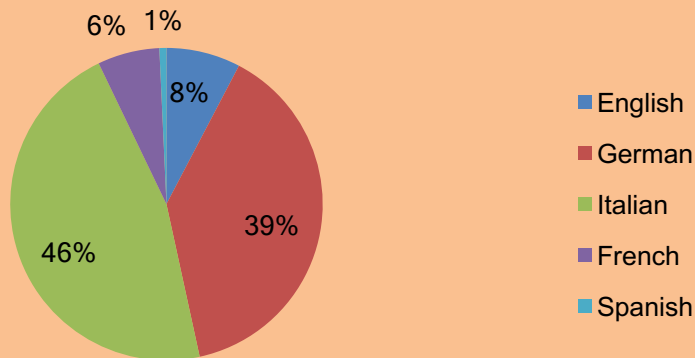
- Identification of potential stakeholders
- First ELRC Workshop in Italy (15 March 2016)
- Questionnaire on existing linguistic resources for the specializaion of the EC machine translation system
- Identification, collection, representation and validation of selected linguistic resources

ELRC 2 (2017-2019)

- Extension of the network of potential stakeholders
- Second ELRC Workshop in Italy (27 September 2018)
- Identification, collection, representation and validation of further linguistic resources

- Among the languages targeted by ELRC, LR for Italian rank fifth in terms of all LR available in ELRC-SHARE
- Resources collected by Italian NAPs for Italian include:
 - bi- and multilingual corpora, for a total of 5,471,657 tokens
 - a linguistically annotated monolingual corpus (25,218,385 tokens)
 - 131,711 translation units
- Textual corpora
 - different genres (administrative acts, laws, manuals, web sites)
 - some of them enriched with linguistic annotation
 - multilingual corpora often include alignment (at different levels, mainly sentence and paragraph levels)
 - different languages

Bi- and multi-lingual corpora: distribution of languages



included:

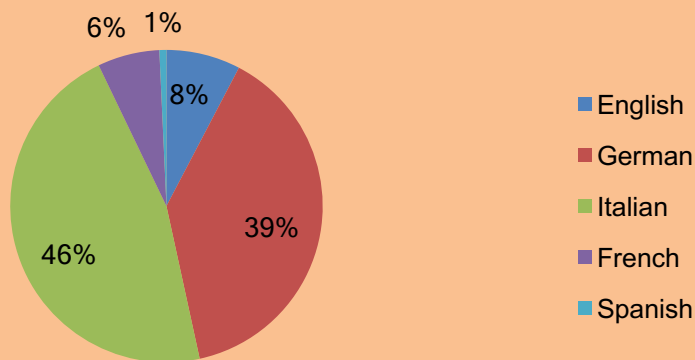
a total of 5,471,657 tokens

lingual corpus (25,218,385 tokens)

acts, laws, manuals, web sites)

- some of them enriched with linguistic annotation
- multilingual corpora often include alignment (at different levels, mainly sentence and paragraph levels)
- different languages

Bi- and multi-lingual corpora: distribution of languages



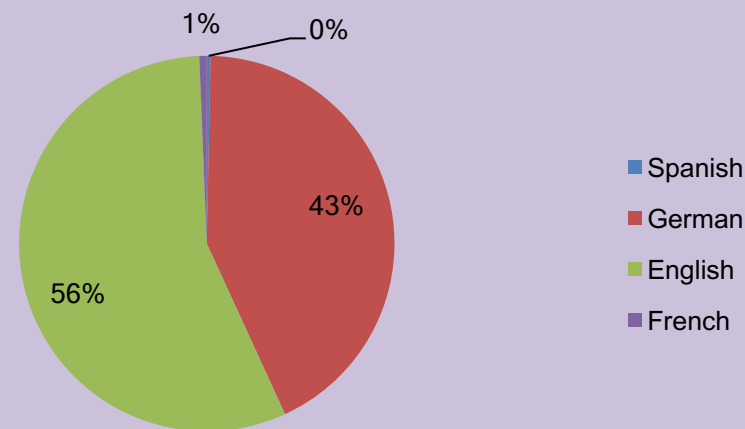
- some of them enriched with
- multilingual corpora often include
- mainly sentence and paragraph
- different languages

included:

a total of 5,471,657 tokens

lingual corpus (25,218,385 tokens)

Translation units: target languages



Contacted institutions

- Presidenza del Consiglio (Presidency of the Council of Ministers)
- Parlamento (Parliament)
- Ministeri (Interni, Giustizia, ...) (Ministries)
- Regioni e Province (regions and provinces)

Data providers so far

- Ministero degli Interni (Ministry of Interior)
- Ministero della Giustizia (Ministry of Justice)
- Provincia di Bolzano (Province of Bolzano)
- Prefettura di Firenze (Prefecture of Florence)
- AgID - Agenzia per l'Italia Digitale (Agency for Digital Italy)
- Università di Bologna e di Pisa (Universities of Bologna and Pisa)
- Institute for Specialised Communication and Multilingualism, EURAC Research

Potential data providers

- Department for European Policies, Presidency of the Council of Ministers, Ministry for Defence, Bank of Italy

- According to the Digital Economy and Society Index (DESI 2018) Italy is growing
 - E.g. the Open Data value has progressed significantly from 52% in 2017 to 81% in 2018, placing Italy in 8th position in the country ranking
- Yet there is still a gap between the supply of digital services and their actual use by the population
- Italy has developed its national strategy by converting the objectives of the European Digital Agenda into initiatives aimed at the digital transformation of Public Administration
 - three-year plan for information technology in the Public Administration for the development of four pillars:
 - digital ecosystems
 - immaterial infrastructures
 - physical infrastructures
 - cybersecurity

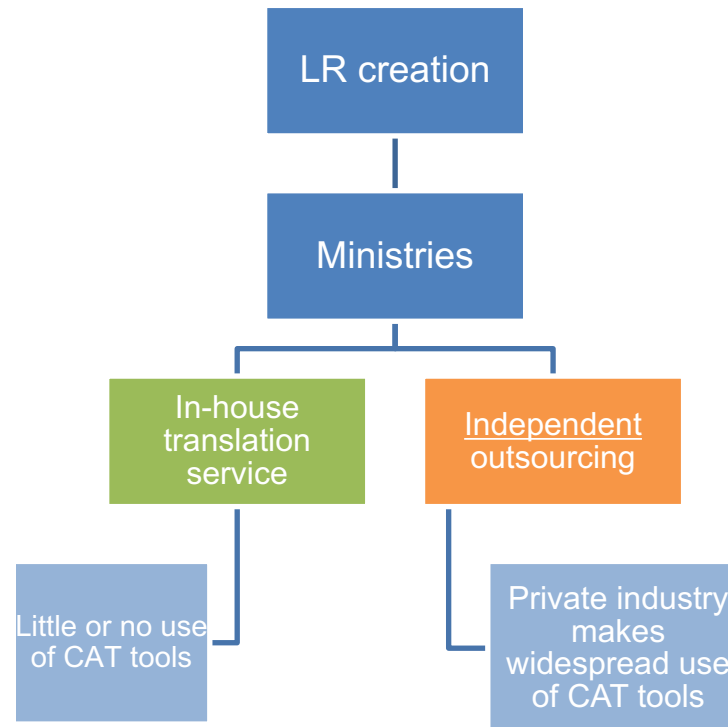
AI at the service of the citizen

- AI methods and tools applied to services is one of the sectors in which Italy is trying to achieve the objectives of the Three-Year Plan for Information Technology in Public Administration
- The White Paper (March 2018), edited by the AI Task Force promoted by the Agency for Digital Italy, is the result of a consultation, synthesis and analysis process that has involved hundreds of public and private subjects dealing with AI
 - ***Machine translation among the technologies playing a key role to meet the challenges of AI at the service of citizens***



<https://libro-bianco-ia.readthedocs.io/en/latest/index.html>

Infrastructure of language data on national level in Italy



- All Ministries and Public Bodies need translation of documents and websites
 - Some provide online cross-border services (ADR, e-justice, social security, data protection authority, IMI, SOLVIT) for which e-Translation is used
- Larger Ministries have in-house translation services but outsourcing is largely resorted to in case of excessive workload or languages not covered
 - But TM ownership is not retained

- Little awareness of the importance of language resources
- Little willingness to share translations
- No centralised terminology management
- Widespread use of paper and pdf (poor quality scans)
- Average age of civil servants in Ministries in 2016: 54.46 (source: General Accounts Office)
- Cost of investing in CAT Tools (Trados, Wordfast)

- Data Ownership (Art. 11 of Italian Copyright Law)
- GDPR – Confidential / Personal Data for many Ministries (Justice/Interior)
- How to find the person who knows if data can be shared?
- Lack of language data management flows and policies
- Language issues relevant only because they are a cost
- Lack of centralised translation services
 - When centralised, head of office is not a linguist
- When outsourced, translations need to be anonymised and sensitive data deleted

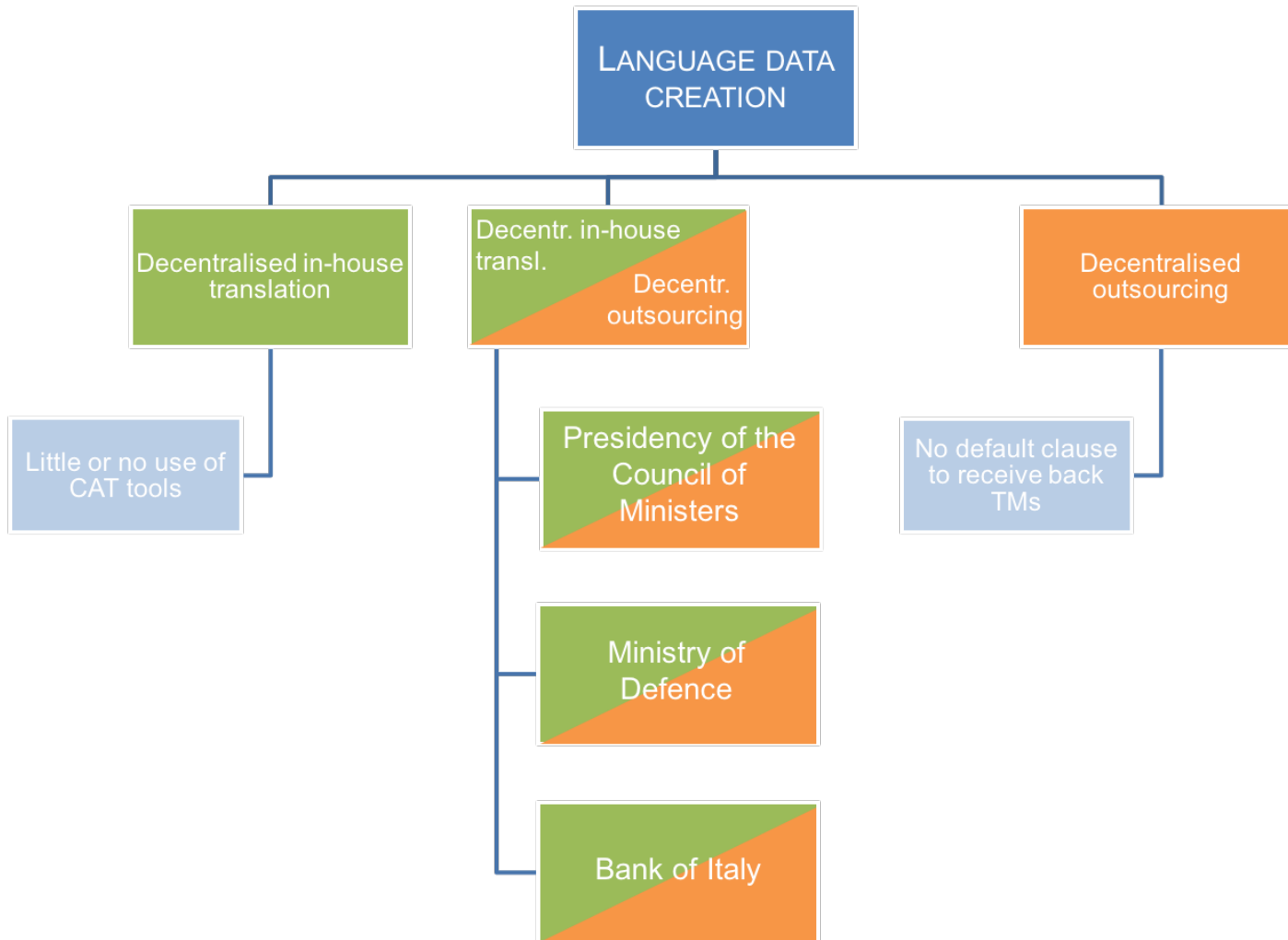
BUT....



(Unconfessed) use of Google Translate and DeepL
without considering privacy issues

Many translators say DeepL is better

Infrastructure of language data on national level in Italy v.2



- Insert clauses in contract for outsourced translations to retain ownership of TMs
- Automating the anonymisation process (support from ELRC?)
- Need to translate (often complex) websites
- Find new language resources
 - also covering migrants' languages
- Connect with AGiD, which is currently focusing on the definition of controlled vocabularies to communicate clearly and unambiguously when designing, assessing and documenting interoperable digital public services

Thanks for the attention!

claudia.foti@giustizia.it

simonetta.montemagni@ilc.cnr.it

ELRC Italy

Email: elrc.italy@ilc.cnr.it



Connecting
Europe
Facility