European Language Resource Coordination
Connecting Europe Facility

ELRC Workshop in Belgium

ELRC in Belgium

Veronique Hoste
Ghent University

First ELRC workshop in Belgium

English (en)

Home Discover Resources Services Events Anchor Points News Helpdesk

European Language Resource Coordination

Nederlandse inleiding

Introduction en français

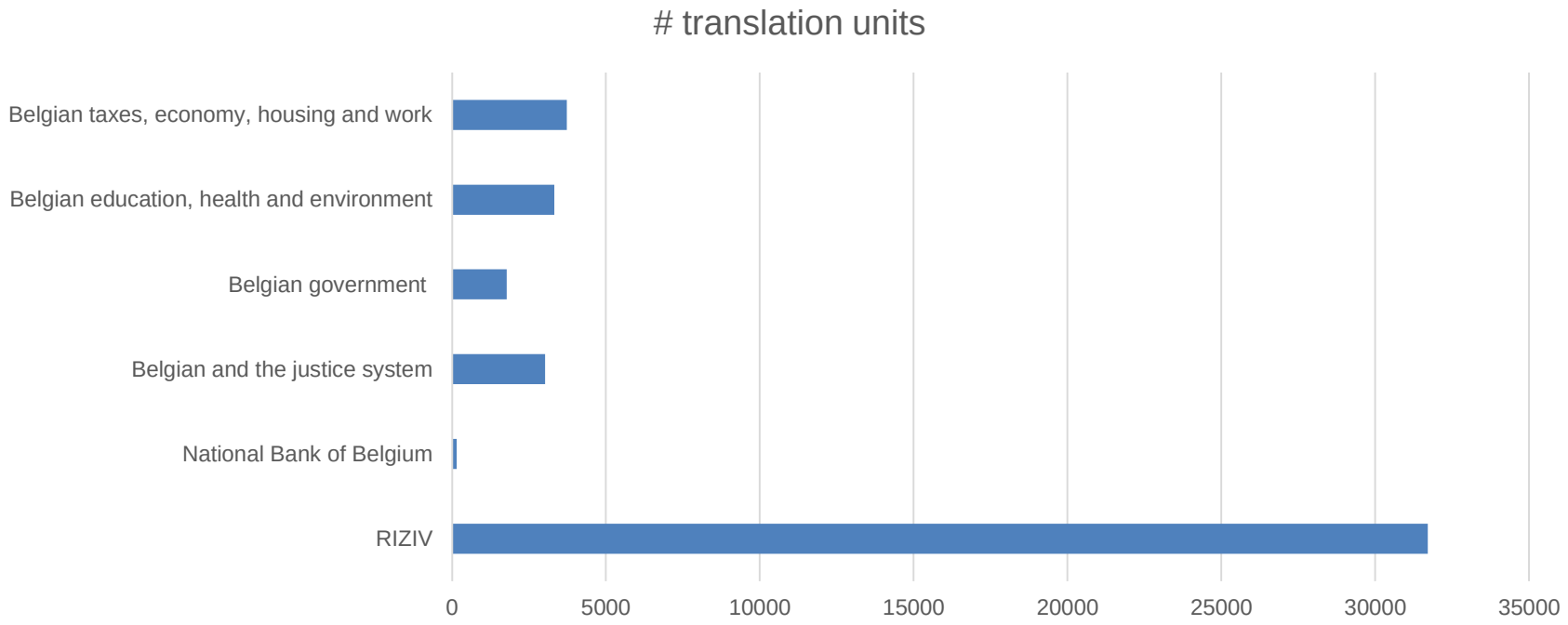
Deutsche Einleitung

ELRC BELGIUM Workshop

Data collected for Belgium

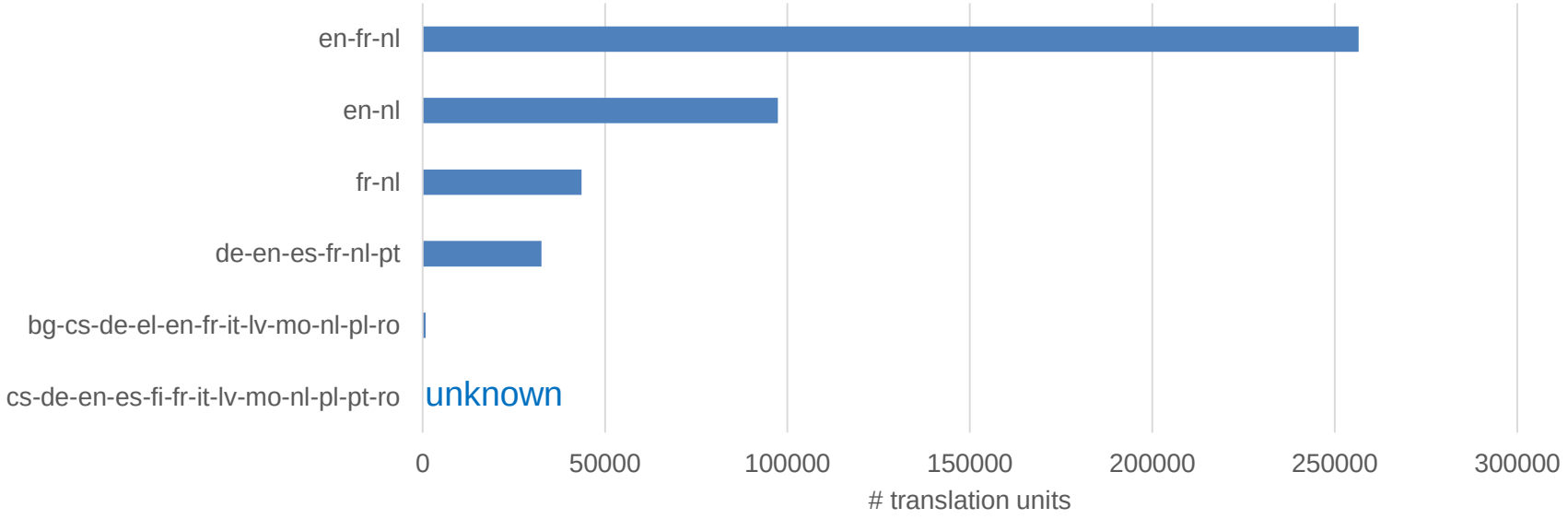
Already processed:

- **term bank** (National Bank of Belgium); 145 entries; de-en-fr-nl
- **translation memory** (RIZIV); fr-nl
- **parallel, aligned corpora** (Belgian government); fr-nl



Data collected for Dutch

- **4 monolingual corpora** (including SoNaR and DAESO)
- **1 term base** (de-en-fr-nl) from the National Bank of Belgium
- **9 bilingual parallel corpora** (en-nl and fr-nl)
  - Belgian and Dutch government and museum websites
- **5 multilingual parallel corpora** (nl + in 14 languages)
  - Incl. Dutch Parallel Corpus, ParaCrawl, multilingual subtitles (2BDutch)



European Language Resource Coordination
Connecting Europe Facility

Data collected for French

- **1 monolingual corpus**
- **12 multilingual term bases** (fr + 18 languages)
- **20 bilingual parallel corpora** (fr + de-el-en-es-nl-pt)
- **24 multilingual parallel corpora** mostly legal, (fr + 18 languages)

Difficult to compare, since size is sometimes expressed in translation units and other times in tokens.

Approximately:

2995 tokens in monolingual corpus
+250k translation units in bilingual corpora
+700k translation units in multilingual corpora
+51k terms in term bases

European Language Resource Coordination
Connecting Europe Facility

Data collected for German

- **no monolingual corpus**
- **17 term bases**
- **28 bilingual corpora** (de + cs-en-es-fr-it-pt)
- **13 multilingual corpora** (de + 19 languages

Difficult to compare, since size is sometimes expressed in translation units and other times in tokens.

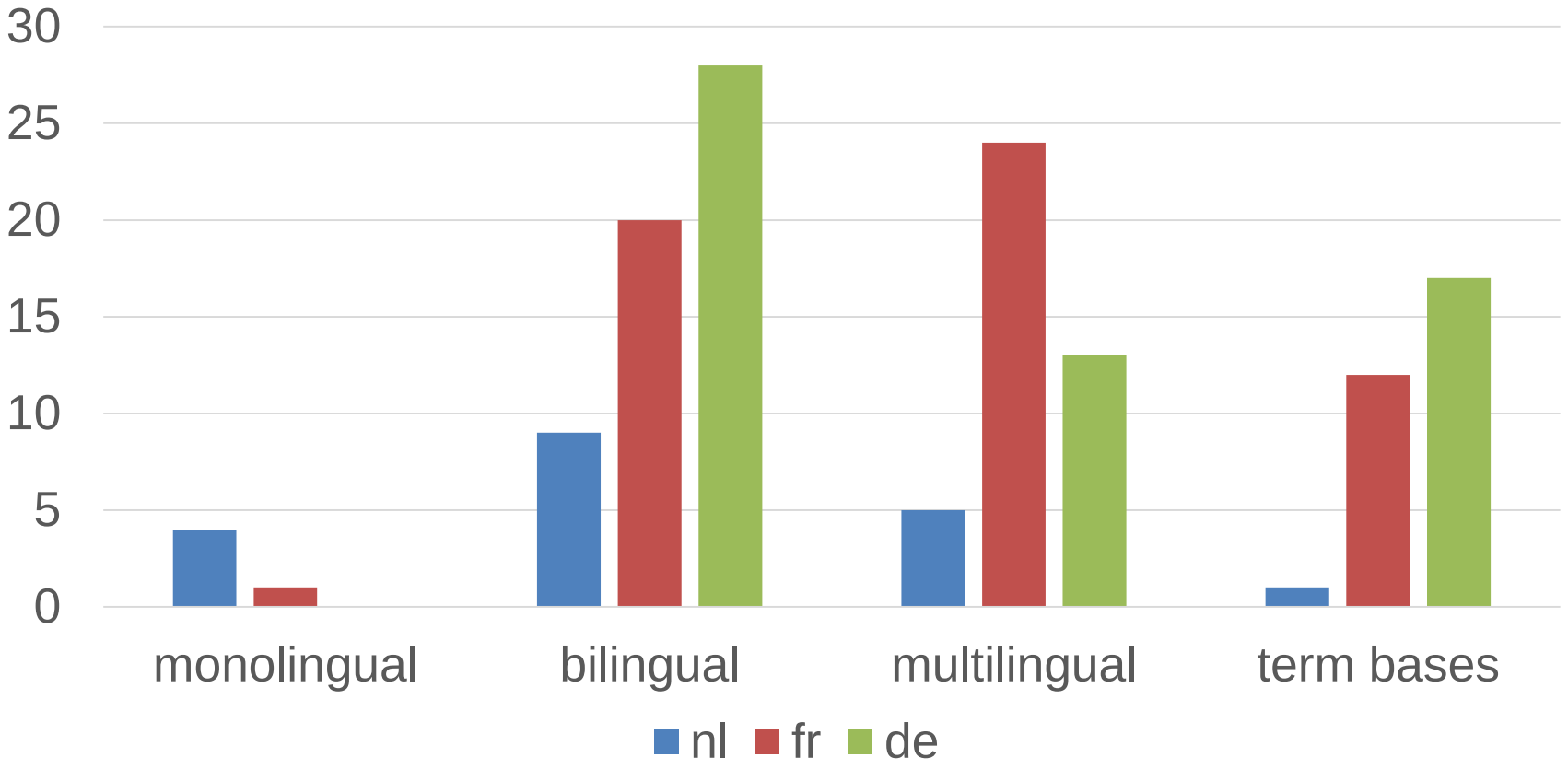
Approximately:

+63k terms
+700k translation units in bilingual corpora
±200k translation units in multilingual corpora

European Language Resource Coordination
Connecting Europe Facility

Data collected for national languages

Number of donated and processed resources, per language and per type:
(size of resources differs!)



Stakeholders involved

National Anchor Points

- Stijn Desmeytere (Chancellery of the Prime Minister)(Public Services NAP)
- Veronique Hoste & Ayla Rigouts Terryn (Ghent University)(Technology NAP)

Data contributors so far

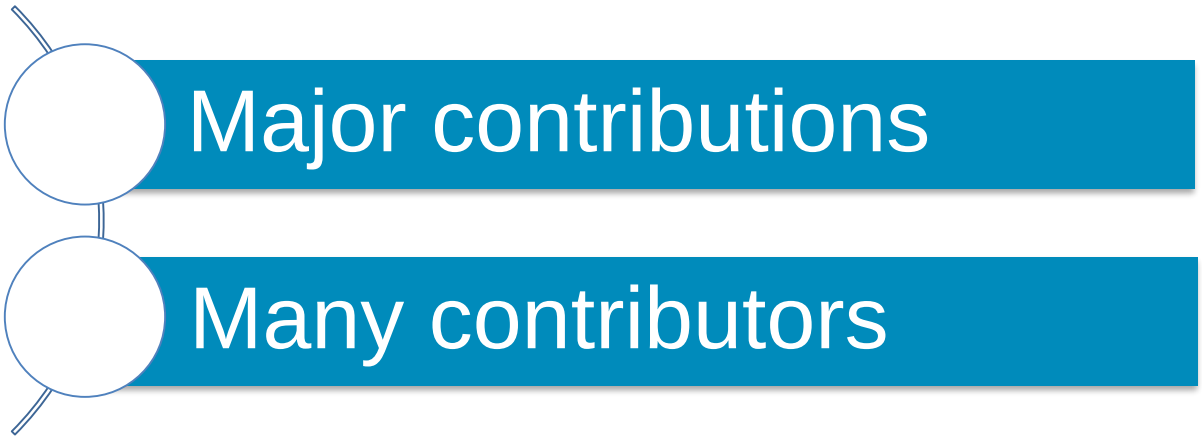
- different federal and regional public services

Potential data contributors / others

- Present institutions

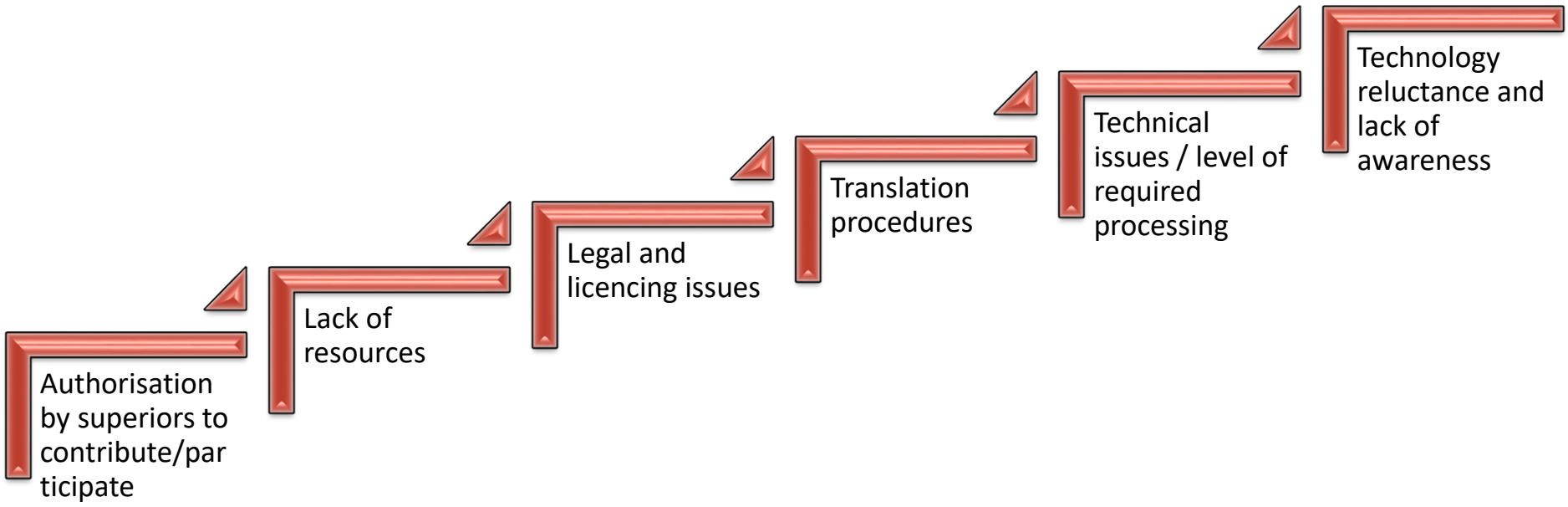
Obstacles and Achievements

Major achievements



Obstacles and Achievements

Main obstacles



Highlights and best practice

Highlights and best practice - what worked well?

Key role / central player(s) and their support provided

Modifications / improvements in translation procedures to facilitate data sharing

Legal support

European Language Resource Coordination
Connecting Europe Facility

Thank you for your attention!

Email: info@lr-coordination.eu

Website: www.lr-coordination.eu