EUROPEAN LANGUAGE EQUALITY

META-NET
META-FORUM 2023
META-NET
META-NET

GL-BLARK – A BLARK for minoritized languages in the era of deep learning: expertise from academia and industry

Sofía García (imaxin|software), Iria de-Dios-Flores (Universidade de Santiago de Compostela)
sofia.garcia@imaxin.com, iria.dedios@usc.gal

Objectives & Methodology

- **Analyze** the key LT resources in the era of deep learning

- **Create** a BLARK matrix that covers basic resources and tasks

- **Provide** LT experts with a wide-covering and flexible BLARK to evaluate the development of minoritized languages

- **Evaluate** resources and tasks using parameters: **quality, quantity, license**

BLARK layout

Cross-cutting resources	40%
Corpora	70%
NPL tools	10%
Lexical resources	10%
Language models	10%

LT task	60%
Speech synthesis	20%
Speech recognition	20%
Machine translation	20%
Other LT task	30%
Benchmarking	10%

Corpora
(70% of the cross-cutting resources category)

Speech synthesis
(20% of the LT task category)

RESOURCE	Size 30%		Quality 40%		Licence 30%	
Annotated corpora 20%	Small	20%	Small	20%	Closed	15%
	Medium	60%	Medium	60%	Restricted	30%
	Large	100%	High	100%	Open	100%
	Size 40%		**Quality 30%**		**Licence 30%**	
Reference corpus 30%	Small	20%	Small	20%		
	Medium	60%	Medium	60%		
	Large	100%	High	100%	Closed	15%
Macro corpus 50%	Small	20%	Small	20%	Restricted	30%
	Medium	60%	Medium	60%	Open	100%
	Large	100%	High	100%		

RESOURCE	Size 40%		Quality 30%		Licence 30%	
Corpus 70%	Small	20%	Small	20%	Closed	15%
	Medium	60%	Medium	60%	Restricted	30%
	Large	100%	High	100%	Open	100%
			Quality 70%		**Licence 30%**	
Model 30%			Small	20%	Closed	15%
			Medium	60%	Restricted	30%
			High	100%	Open	100%

ELE   European Language Equality

31

Conclusions & Limitations

- The BLARK needs to be **validated** with other **minoritized languages**

- The BLARK needs to be continuously **updated**

- This tool is only usable by **NLP experts**

European Language Equality

Thank you!

Sofía García (imaxin|software), Iria de-Dios-Flores (Universidade de Santiago de Compostela)
sofia.garcia@imaxin.com, iria.dedios@usc.gal

27-06-2023 META-FORUM 2023 – Digital Language Equality for a Multilingual Europe
http://european-language-equality.eu