EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union

UniDive

WORKING GROUP 4

QUANTIFYING AND PROMOTING DIVERSITY

Low Resourced Languages. State of the Art

Sub-Task 2

Low resourced and endangered languages

Lucía Amorós-Poveda, lamoros@um.es
Didactic and School Organization Department (UM, ISEN & UNIR)
Center for Studies on Educational Memory (CEME)

UNIVERSIDAD DE MURCIA · FACULTAD DE EDUCACIÓN · CeMe Universidad de Murcia · unir LA UNIVERSIDAD EN INTERNET

Friday 9 February, 2024
University of Naples "L'Orientale", ITALY
15:15-16:30 Parallel working sessions. Room 1.5

UNIVERSITÀ DI NAPOLI L'ORIENTALE

A brief summary of what our subgroup has done

PART 1

5 minutes approx.

STEP 1 Stakeolders
STEP 2 Sub-task and UniDive (MoU)
ACTIVE PART 1: Stakeholder participation

PART 2

15 minutes approx.

STEP 3 Synthesis (Joshi et al., 2020)
STEP 4 Open document "State-of-the-Art in LRL"
https://docs.google.com
STEP 5 Discussion Brazilian languages, by André V. Lopes Coneglian and other languages .
ACTIVE PART 2: Paper

PART 1 5 minutes approx.

STEP 1 Stakeolders and their language interests
1 minute approx.

STEP 2 Sub-task's aims and comparison with the UniDive objectives grid and sub-task relationship (MoU)
2 minutes approx.

ACTIVE PART 1
Each member's position on them to future actions on LRL* or LREL** according to this grid
2 minutes approx.

*LRL= Low resourced languages
**LREL = Low-resourced and endangered languages

STEP 1

Stakeholders and their language interests

CO-LEADER
UniDive

Dan Zeman (CZECH REPUBLIC)
Institute of Formal and Applied Linguistics (ÚFAL)
Computer Science School
Faculty of Mathematics and Physics, Charles University
https://ufal.mff.cuni.cz/daniel-zeman

Marie-Catherine de Marneffe (BELGIUM)
Computational Linguistic
FNRS - UCLouvain - The Ohio State University
https://cental.uclouvain.be/team/mcdm/

Abigail Walsh (IRELAND)
Dublin City University
DCU · School of Computing

CO-LEADERS WORKING GROUP 4

QUANTIFYING AND PROMOTING DIVERSITY

Martin Benjamin (USA/Switzerland)
Executive Director, Kamusi Project International
YouTube Channel “The Pirate Professor”
https://www.youtube.com/@pirateprofessor

Roberto A. Díaz Hernández, PhD (SPAIN)
Junior Professor and Head of *Nile in Contact*
Faculty of Humanities , University of Jaén

André V. Lopes Coneglian (BRASIL)
Professor da Faculdade de Letras (FALE)
Universidade Federal de Minas Gerais (UFMG)
http://www.letras.ufmg.br/profs/andreconeglian/

Irina Lobzhanidze (GEORGIA)
Professor of Linguistics
Ilia State University
http://irinalobzhanidze.com/

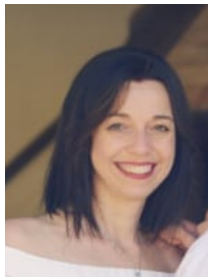
Rusudan Makhachashvili (UCRAINE)
Borys Grinchenko Kyiv University
Institute of Philology

Giedre Valunaite Oleskeviciene (LITHUANIA)
Mykolas Romeris University
https://scholar.google.com/citations?user=0xOwWMEAAAAJ&hl=en

Hiwa Asadpour (GERMANY)
Department of English and American Studies
Goethe-University Frankfurt
https://www.uni-frankfurt.de/109653793/Associate_researcher

Johnatan Bonilla Huerfano (COLOMBIA)
PhD Candidate, Master Degree on Linguistics
Ghent University (Belgium)
https://research.flw.ugent.be/nl/johnatan.bonillahuerfano

Nikolett Mus (HUNGARY)
Hungarian Research Centre for Linguistics
https://sites.google.com/view/nikolett-mus



Lucía Amorós-Poveda(SPAIN)
Department of Didactic and School Organization
Faculty of Education / ISEN. University of Murcia
International University of La Rioja

CO-LEADERS WG 4

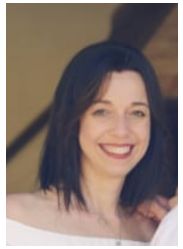
UD

Irish Language

CO-LEADER UniDive

UD

WG 3, 4
Educational Technology
Inclusive Education / Diversity



Spanish Language

NexusLinguarum
Attitudinal discourse

WG 4

African Languages

WG1

Egiptian Language

WG1

Brazilian Languages

WG 1,2

Kartvelian Languages

WG 1,2, 4

Uralic Language spoken (Tundra Nenets)

WG 1, 2, 3

Western Asia Languages

WG 2, 3

Romance and Germanic

WG 1, 2,4

AGREEMENTS

Claudia Soria
claudia.soria@ilc.cnr.it
ILC – Istituto di linguistica computazionale "Antonio Zampolli" to give us knowledge about the *Digital Language Survival Kit* (2018)

STEP 2

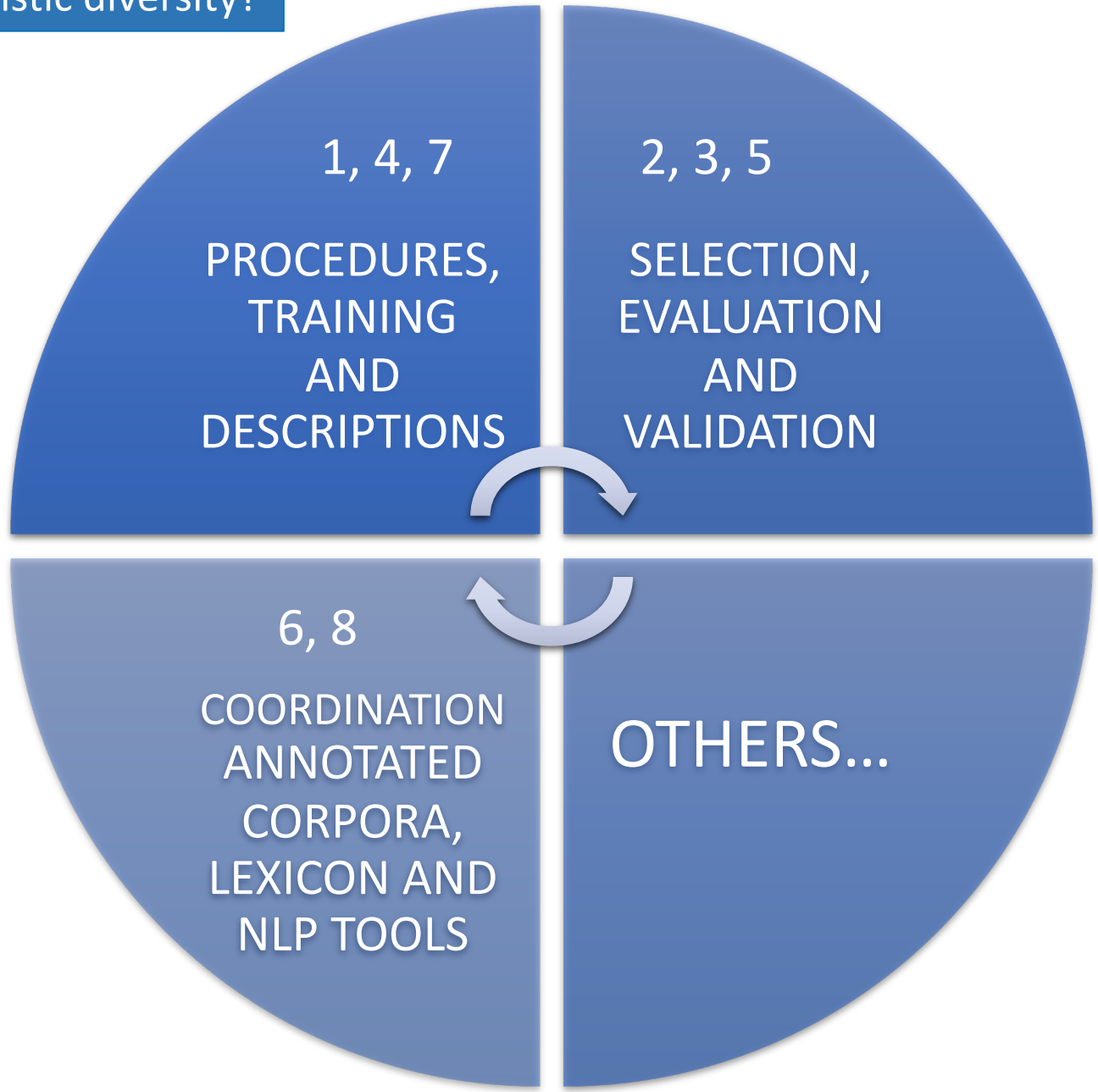
Sub-task’s aims and comparison with the UniDive objectives grid and sub-task relationship (MoU)

How the UniDive serves inter- and intra- linguistic diversity?

PROMOTING DIVERSITY

Objectives of this sub-task

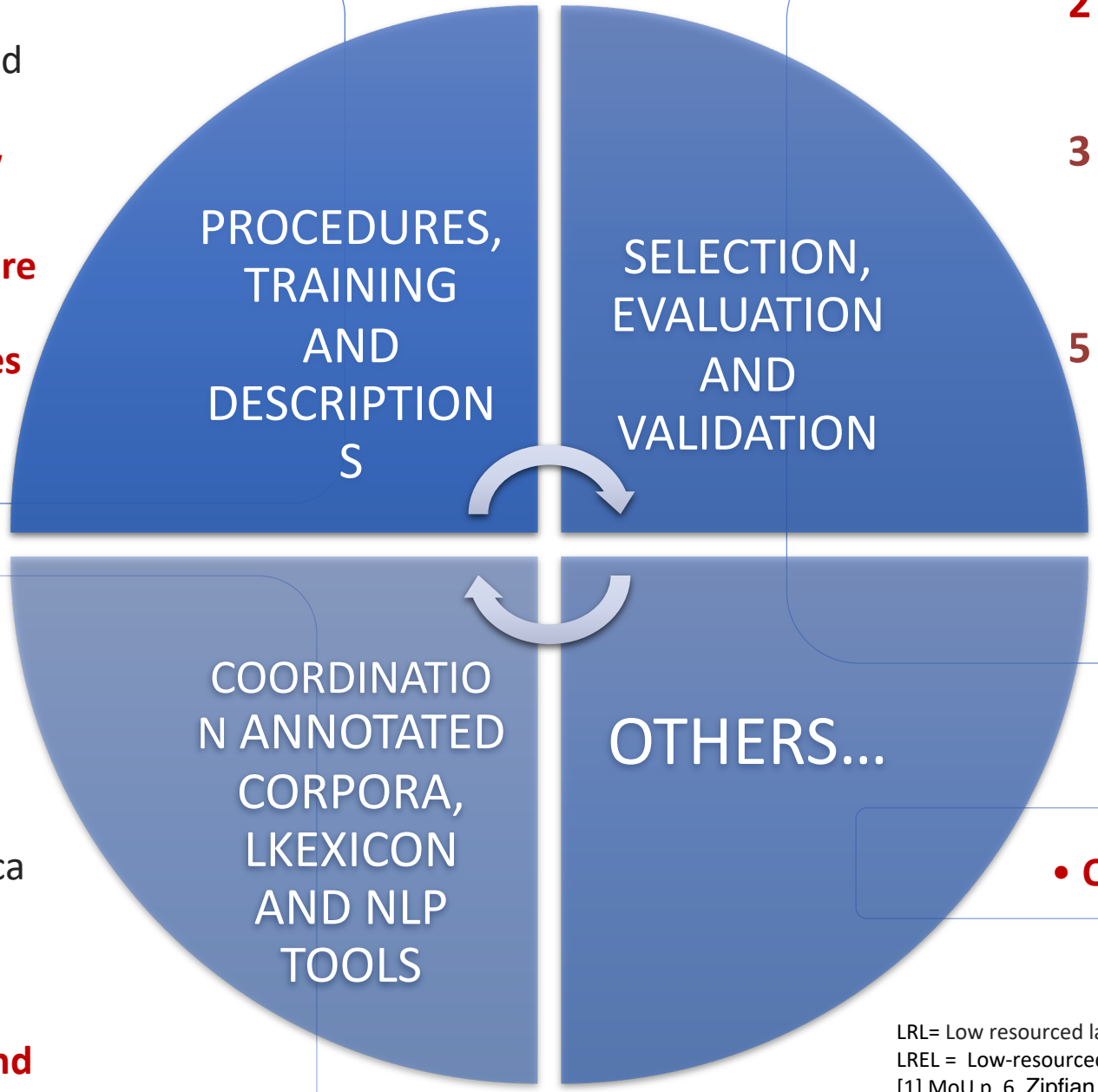
MOU, PP. 19-20



1 Procedures for better use of the existing **resources**, based on their estimated diversity

4 Integrating and training **new experts** dedicated to LREL

7 Discovering and analysing **rare linguistic phenomena**, and describing them in **resources and tools**



2 Selecting new data **to be annotated**, so as to favour intra-linguistic diversity [1]

3 Designing evaluation scenarios which favour **tools performing well** on rare and diverse phenomena and LRL

5 Validating the unified annotation guidelines (WG1) and lexicon formats (WG2) against newly included languages and **defining new language-specific categories and extensions**, if needed

6 Coordinating the **creation and enhancement** of annotated corpora and lexica for LRL

8 Coordination of the development of **NLP tools** (WG3) **for low-resourced and endangered languages**

- **Others**

LRL= Low resourced languages
LREL = Low-resourced and endangered languages
[1] MoU p. 6, Zipfian distribution, sensitivity of NLP to data, diversity within a language, unbounded dependencies, MWE, idiosyncrasies

Three NLP initiatives have recently engaged in **universality**: Universal Dependencies (UD, Nivre et al. 2020), PARSEME (Savary et al. 2018, Ramisch et al., 2020) and UniMorph (Kirov et al. 2018). […]

The contributions of **universality-driven initiatives to preserving and promoting diversity** are manifold. Unified methodologies promote **inclusiveness** because they lead to the construction of shared frameworks in which new experts are easily integrated and can contribute in a grassroots manner.

[…] **UD** and **PARSEME** offer centralized collaborative infrastructures for the development of unified annotation guidelines and annotated corpora.

MoU, p. 7

Universality also benefits from recent deep learning models, which are trained on large quantities of multilingual raw (i.e., **non-annotated**) texts in language- and task-agnostic ways (Devlin et al. 2019). Such models are expected to implicitly encode knowledge about language in general (including statistical universals).

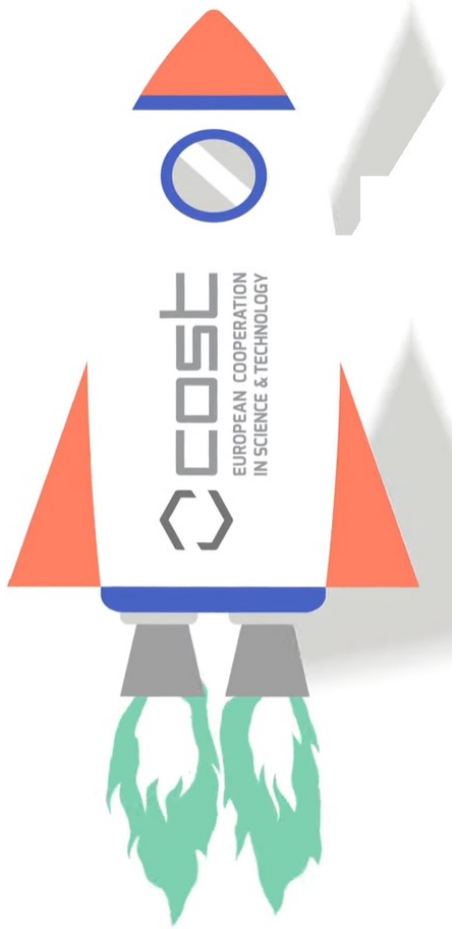
These models can then be fine-tuned **for a particular language** and task with small quantities of annotated data, and even used for languages with no annotated data (Pires et al., 2019).

Thus**, low-resourced languages benefit from large multilingual data**. Also, the intra-linguistic diversity is higher since raw data are by several orders of magnitude larger than annotated data.

MoU, p. 8

PROGRESS BEYOND THE STATE OF THE ART

Increasing inter-language diversity via a better NLP support for **LRL**



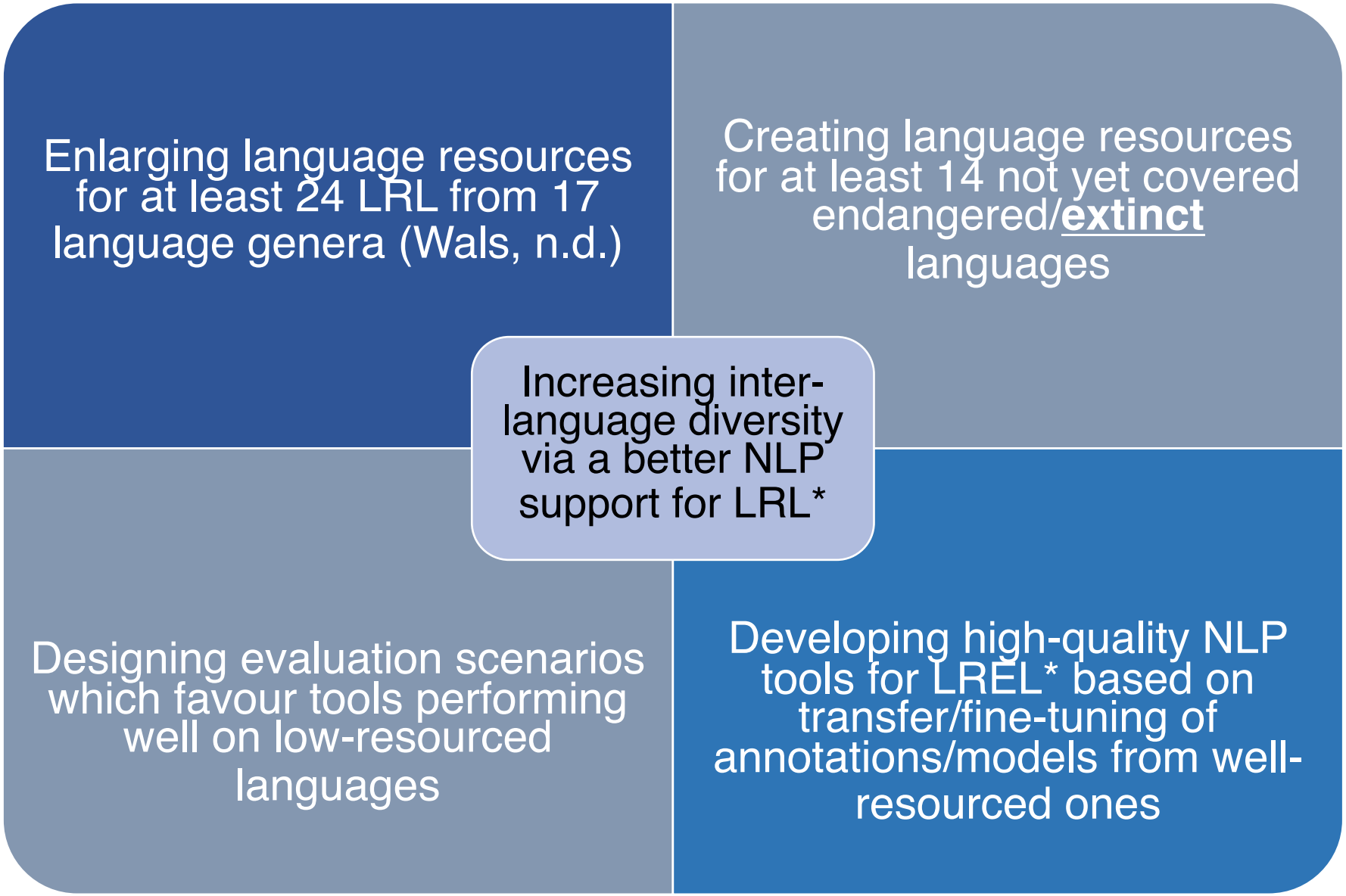
TASK = Aims of this sub-task

Comparison with the UniDive objectives grids

Sub-task relationship

Approach to the challenge

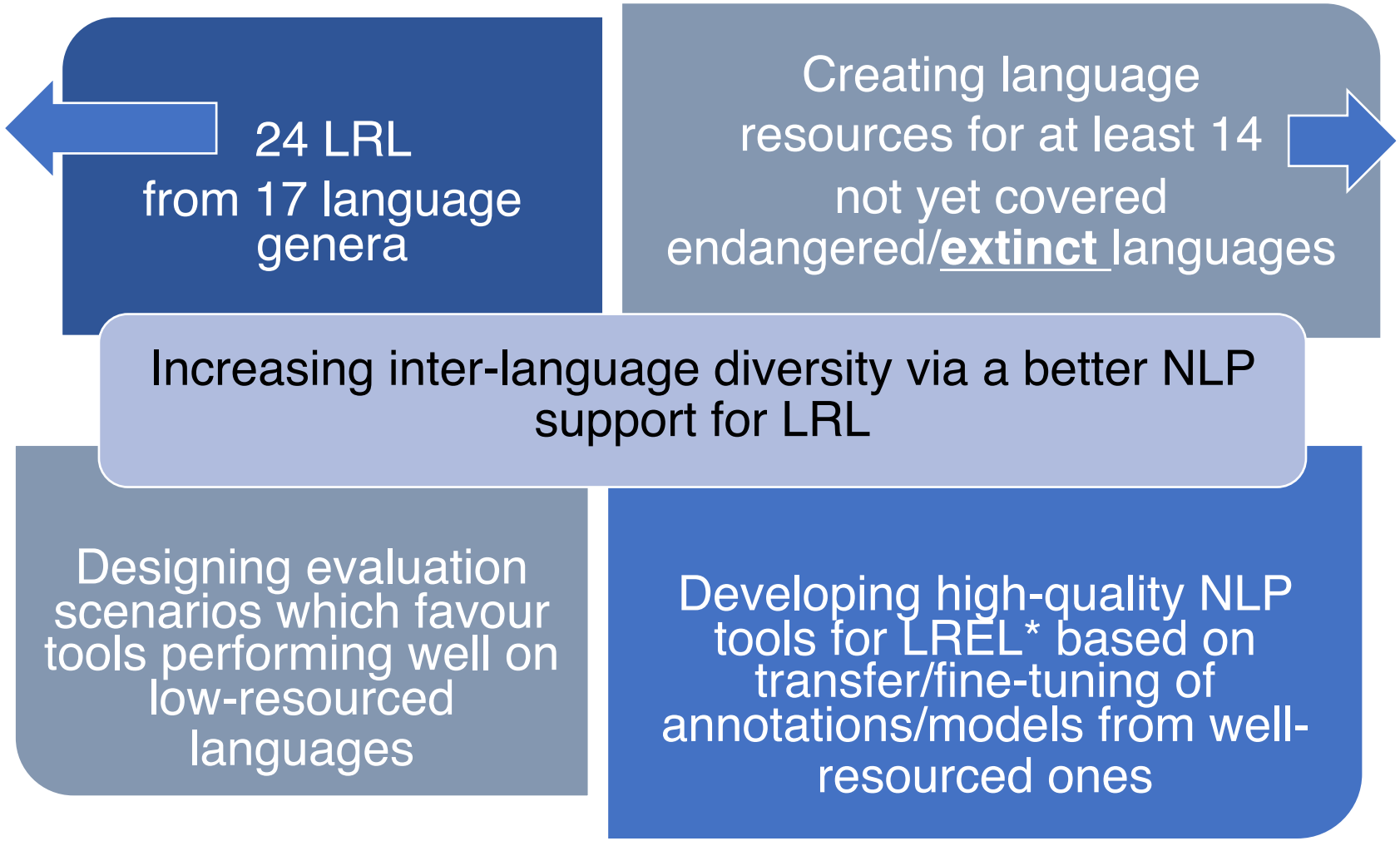
MoU, p. 9



MoU, p. 9

*LRL= Low resourced languages *LREL = Low-resourced and endangered languages

Akuntsu
Chukchi
Basque
Bulgarian
Frisian
Greek
Guajajara
Hindi
Hungarian
Irish
Javanese
Ka'apor
Lithuanian
Makurap
Maltese
Manx
Mundurukú
Odia
Romanian
Serbian
Slovene
Swedish
Tagalog
Turkish



Abaza
Cusco
Quechua
Georgian
Hittite
Kabyle
Karo
Ligurian
Maghrebi Arabic
Neapolitan
Occitan
Old Irish
Laz
Xibe
Yakut

MoU, p. 9

1.2.2. OBJECTIVES

1.2.2.1 Research Coordination Objectives

[RCO4] To coordinate efforts towards a better coverage of inter-/intra-linguistic diversity in NLP tools.
[RCO5] To raise awareness of the international community about the importance of diversity preservation in language technology.

1.2.2.2 Capacity-building Objectives

[CBO1] To create a network of experts in a large number of languages working on modelling and processing of morphological, syntactic and semantic phenomena within a common framework.
[CBO2] To foster the capacities of Young Researchers and Innovators (YRIs), with special focus on COST ITC participants.

MoU, p. 11

4.1.2. DESCRIPTION OF DELIVERABLES AND TIMEFRAME

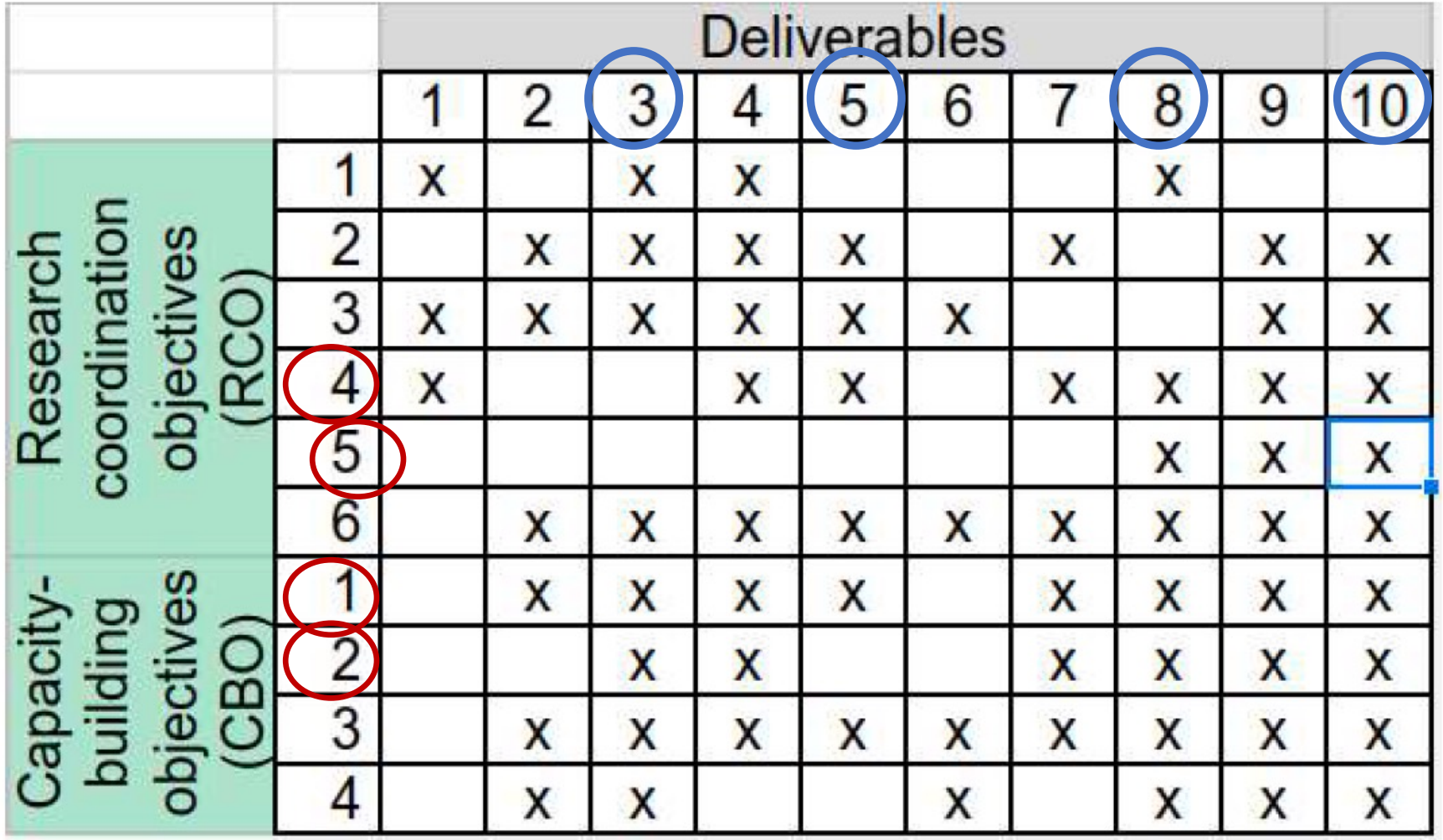
[D3] Centralized documentation of the nationally funded software infrastructures coordinated in WG1 and WG4, to support universality and diversity in language resources.

[D5] Centralized documentation of (new or enhanced) annotated corpora for at least 100 languages.

[D8] Diversity benchmarks for NLP: diversity-driven evaluation scenarios for NLP resources and tools; infrastructure for evaluation campaigns of NLP tools; evaluation results of at least 2 evaluation campaigns and focused on inter/intra-linguistic diversity in 100 languages.

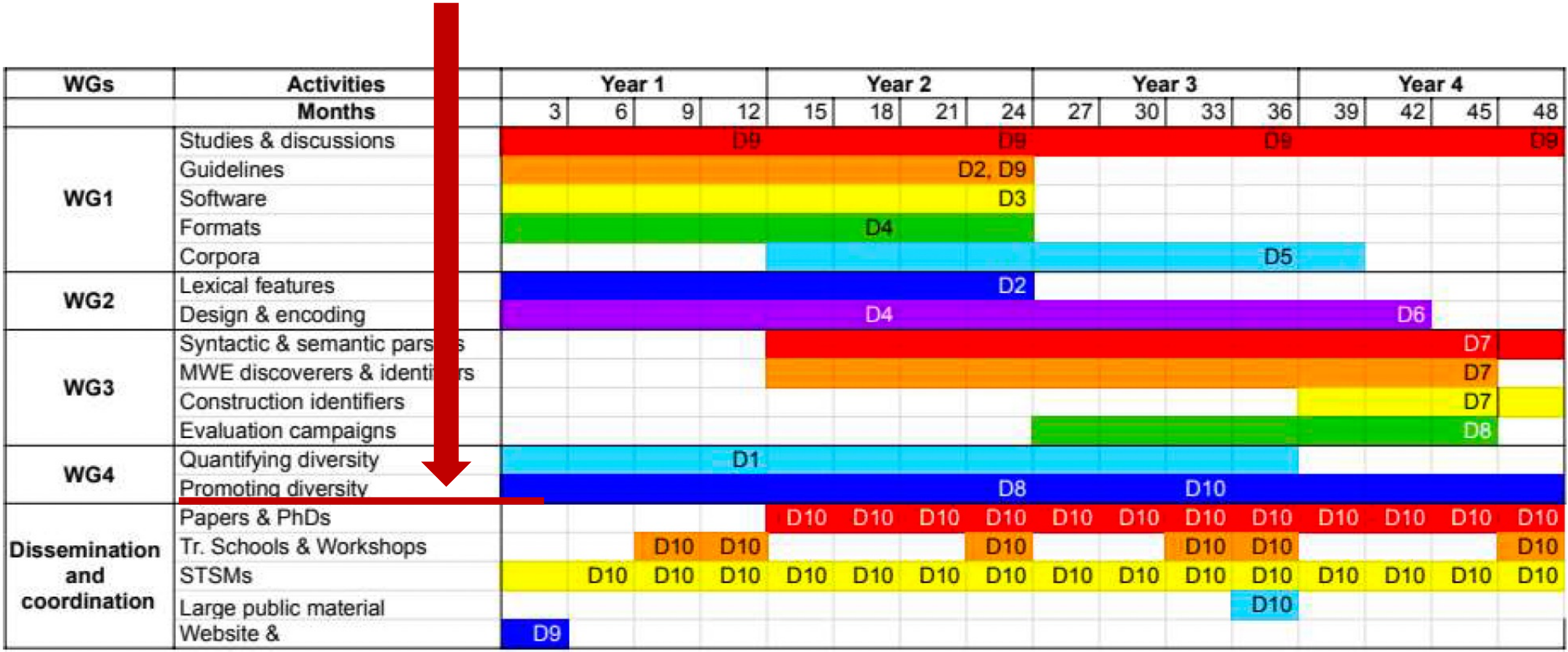
[D10] Other dissemination material: reports from STSMs; material from training schools; proceedings of workshops; joint papers in Open Access journals, conferences and books; dissemination material dedicated to a large audience (e.g., demonstrations of tools and Wikipedia entries about diversity in NLP, MWEs and idiosyncratic constructions, and interesting syntactic phenomena).
The table below shows the measurability of the objectives (§1.2.2) in terms of the deliverables

MoU, p. 20



		Deliverables									
		1	2	3	4	5	6	7	8	9	10
Research coordination objectives (RCO)	1	x		x	x				x		
	2		x	x	x	x		x		x	x
	3	x	x	x	x	x	x			x	x
	4	x			x	x		x	x	x	x
	5								x	x	x
	6		x	x	x	x	x	x	x	x	x
Capacity-building objectives (CBO)	1		x	x	x	x		x	x	x	x
	2			x	x			x	x	x	x
	3		x	x	x	x	x	x	x	x	x
	4		x	x			x		x	x	x

MoU, p. 21


WGs
Activities
Months
Year 1
Year 2
Year 3
Year 4
3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48
WG1
Studies & discussions
Guidelines
Software
Formats
Corpora
WG2
Lexical features
Design & encoding
WG3
Syntactic & semantic pars s
MWE discoverers & identi rs
Construction identifiers
Evaluation campaigns
WG4
Quantifying diversity
Promoting diversity
Dissemination and coordination
Papers & PhDs
Tr. Schools & Workshops
STSMs
Large public material
Website &
D9
D2, D9
D3
D4
D5
D2
D6
D7
D8
D1
D10

MoU, p. 22

ACTIVE PART 1

Each member's position on them to future actions on LRL* or LREL** according to this grid

2 minutes approx.

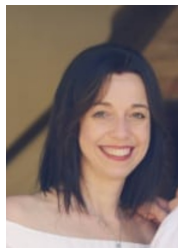
*LRL= Low resourced languages

**LREL = Low-resourced and endangered languages

CO-LEADERS WG 4

CO-LEADER UNIDIVE UD

1 to 8



WG1

WG1

WG 1,2

2 to 5, 7

WG 1, 2,3

1 to 8

WG 2,3

WG 1, 2,4

1

PART 2 15 minutes approx.

STEP 3 Synthesis of "The State and Fate of Linguistic Diversity and Inclusion in the NLP World" (Joshi et al., 2020). 2 minutes approx.

STEP 4 Introduction to the open document "State-of-the-Art in Low resourced languages" with https://docs.google.com (sent to you by email). 1 minutes approx.

STEP 5 Discussion Brazilian languages, by André V. Lopes Coneglian and other languages . 5 minutes approx.

ACTIVE PART 2. Structure of the paper. 7 minutes approx.

STEP 3

Synthesis of "The State and Fate of Linguistic Diversity and Inclusion in the NLP World" (Joshi et al., 2020)

2 minutes approx.

THE LANGUAGE TAXONOMY

0 The Left-Behinds Languages that have been and are still ignored in the aspect of language technologies. Exceptionally limited resources. Needs: a monumentous, probably impossible effort to lift them up in the digital space / Unsupervised pre-training methods only make the 'poor poorer'

1 The Scraping-Bys Some amount of unlabeled data. Needs: a solid, organized movement that increases awareness, strong effort to collect labelled datasets

2 The Hopefuls A small set of labeled datasets has been collected / there are researchers and language support communities which strive to keep them alive in the digital world / Promising NLP tools can be created

3 The Rising Stars Positive effect with unsupervised pre-training /strong web presence / Insufficient efforts in labeled data collection

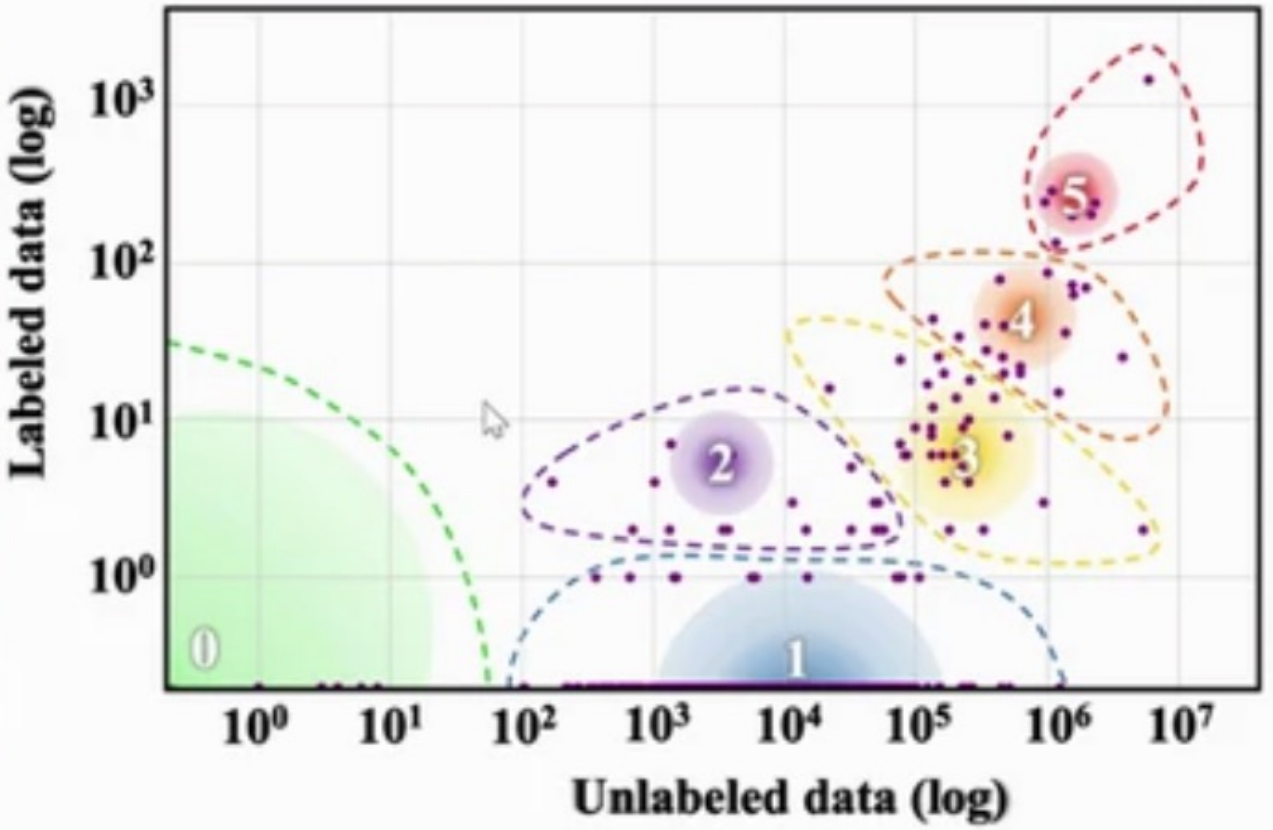
4 The Underdogs large amount of unlabeled data. Whith dedicated NLP communities conducting research on these languages, they have the potential to become winners

5 The Winners dominant online presence, there have been massive industrial and government investments

Joshi et al., 2020, pp. 6284-6285

THE LANGUAGE TAXONOMY - VISUALIZATION

- Class 0 (The left-behinds) - Gondi, Mundari
- Class 1 (The Scraping-Bys) - Bhojpuri, Assamese
- Class 2 (The Hopefuls) - Konkani, Wolof
- Class 3 (The Rising Stars) - Tamil, Marathi
- Class 4 (The Underdogs) - Bengali, Hindi
- Class 5 (The Winners) - English, French



Language Resource Distribution

Size of the gradient circle = number of languages in the class.

Color spectrum = total speaker population size from low to high

Bounding curves = covered points by that language class

Joshi et al., 2020, p. 6284

THE LANGUAGE TAXONOMY

Number of languages, number of speakers, and percentage of total languages for each language class

Class	5 Example Languages	#Langs	#Speakers	% of Total Langs
0	Dahalo, Warlpiri, Popoloca, Wallisian, Bora	2191	1.2B	88.38%
1	Cherokee, Fijian, Greenlandic, Bhojpuri, Navajo	222	30M	5.49%
2	Zulu, Konkani, Lao, Maltese, Irish	19	5.7M	0.36%
3	Indonesian, Ukranian, Cebuano, Afrikaans, Hebrew	28	1.8B	4.42%
4	Russian, Hungarian, Vietnamese, Dutch, Korean	18	2.2B	1.07%
5	English, Spanish, German, Japanese, French	7	2.5B	0.28%

Joshi et al., 2020, p. 6285

Sugestions…

- Jointly learn the representations of Conferences, Authors and Languages, collectivelly termed as Entities
- Typologies considerations
- Focused communities
- Inclusive Conferencies

Sugestions...

Towards Inclusive Conferencies

A way to promote change could be the addition of D&I (Diversity and Inclusion) clauses involving language-related questions in the submission and reviewer forms:

- Do your methods and experiments apply (or scale) to a range of languages?
- Are your findings and contributions contributing to the inclusivity of various languages?

Joshi et al., 2020

STEP 4

Introduction to the open document

Goal

Share ideas on the topic *State-of-the-Art in LRL*

https://docs.google.com (sent to you by email)

Need - Research question

To look for solutions - Why the EU needs to care about LREL?

1 minutes approx.

Collaborative Scientific Research Report

January 5, 2024

Free Collaborative document

https://docs.google.com/document/d/1-5laKC5cEa6v30sndi1VA4tmphc5YVKMgdnO-QwGYIk/edit?usp=sharing

Low Resourced Languages.
State of the Art

Collaborative Scientific Research Report
January 5, 2024

Please: Include your Surname, Name. Affiliation (Country), e-mail

Amorós-Poveda, Lucía.
University of Murcia (Spain)
International University of La Rioja (Spain)
lamoros@um.es

Lobzhanidze, Irina
School of Arts and Science - ILIA State University (Georgia)
irina_lobzhanidze@iliauni.edu.ge

INDEX

1. INTRODUCTION

Why attend Low Resourced Languages? Why focus on the State of the Art?

> NOTION A) To consider languages with fewer economic resources, such as Swahili, Hindi, and Brazilian local environmental languages, alongside extinct languages like Egyptian, Latin, Greek, Akkadian, and Classical Chinese.
>
> Chat with Roberto Díaz (University of Jaen)

Why is necessary that no language is left behind?
What are the benefits of do it?

How we attempt to convince the ACL community to prioritize the resolution of the reasons highlighted here? How we attempt to convince the European Union for the same? What is the situation about Low resourced languages?

1.1. Context. What is Uni-Dive? What kind of things we can do?
1.2. Related Terms: Low resourced languages Vs Endangered-Extincted Languages / Massive Multilinguality / Multi-lingual, coss-lingual / Language Diversity (Louis)

INDEX

2. METHODS

State of the Art model (documental research)
Aims:
Specific aims:

3. ANALYSIS

4. RESULTS

INDEX

2. METHODS

State of the Art model (documental research)
Aims:
Specific aims:

3. ANALYSIS

4. RESULTS

INDEX

5. DISCUSSION

TO DISCUSS

SENT BY André Coneglian on December, 1st, 2023
An example to discuss – Brazilian native languages

INDEX

REFERENCES

PRINCIPAL PAPER TOPIC:

https://aclanthology.org/2020.acl-main.560.pdf

The State and Fate of Linguistic Diversity and Inclusion in the NLP World

By Pratik Joshi, et al.

Choudhury Microsoft Research, India

VIDEO Duration 12:00 aprox.
https://slideslive.com/38929069/the-state-and-fate-of-linguistic-diversity-in-the-nlp-world

SLIDES 55

REFERENCES, others documentation shared:

1 About a deep study towards the minor languages in Africa, https://www.zotero.org/malangali

2 About a list of multilingual shared tasks
https://docs.google.com/document/d/1GsZR44JTWEngHrZC0alwtK9VJqBzffvve4haAcuJoTk/edit#heading=h.q35b4tdr8139

3 About CoNLL 2018 Shared Task, https://universaldependencies.org/conll18/results.html

4 about XTREME, a massively multilingual multi-task benchmark for evaluating thinking in a cross-lingual generalization by Aditya Siddhant, Junjie Hu, Melvin Johnson, Orhan Firat and Sebastian Ruder done in ICML 2020 (2020). This is an introduction into the Transfer Gap for zero-shot transfer. It is negligible for languages with better pre-training resources, but indeed it goes down for lesser resources ones, https://research.google/pubs/pub49271/

5 About a hard number thinking in language exclusion in computational linguistics and NLP by Martin Benjamin, Kamusi Project. Proceedings of the *11th International Conference on Language Resources and Evaluation*. The document explains Kamusi Project International. For more info here is a paper. **Martin wrote for thinking it was LREC, a few years ago, that uses jobs in NLP as the dataset for what languages are able to get research attention.**
http://lrec-conf.org/workshops/lrec2018/W26/pdf/23_W26.pdf

6 Looking for a methodology for differences between languages. We can check (not free document) the volume 2 from *Complex Constructions* (2nd edition)
Chopin, Tmiothy. (2007). Language Typology and Syntactic Description. Vol 2. Complex constructions. Cambridge University Press.
https://doi.org/10.1017/CBO9780511619434
Nenad Ivanović has sent us an abstract about this book.

7 An open point is the intervention from Loïc Grobol: https://scholar.google.com/citations?user=I5AsU5oAAAAJ&hl=en
Associate Professor, MoDyCo, Université Paris Nanterre

8 On Hodge-Riemann Cohomology Classes (2021) by Julius Ross & Matei Toma https://doi.org/10.48550/arXiv.2106.11285**,** sent by Branislav Gerazov

STEP 5

Discussion of the document on Brazilian languages sent by André V. Lopes Coneglian and other languages according to the interest and expectations of the members

5 minutes approx.

SENDED BY André Coneglian on December, 1st, 2023

An example to discuss – Brazilian native languages

Language	#	Source of data	Other relevant info
Akuntsu (Tupian, Tupari)	1449 tokens 1468 words 343 sentences	Grammar	Text in English Indicates source of the sentence
Apurina (Arawakan)	938 tokens 941 words 152 sentences	Grammar	Text in English and Portuguese Gloss in Portuguese
Bororo (Bororoan)	1905 tokens 1905 words 371 sentences	grammar examples, myths, and other sources.	Text in Portuguese
Guajajara (Tupian, Maweti-Guarani)	8870 tokens 9160 words 1182 sentences	descriptions of the language, short stories, dictionaries and translations from the New Testament	Text in English, Portuguese and Spanish Indicates source of the sentence Source of sentence not indicated in the list of sources
Kaapor (Tupian, Maweti-Guarani)	366 tokens 366 words 83 sentences	Grammar and papers	Text in English Indicates source of the sentence Annotation is not uniform: some sentences have features, others do not.
Karo (Tupian, Ramarama)	2319 tokens 2319 words 674 sentences	Grammar and dictionaries	Text in English Indicates source of the sentence
Madi (Arawan)	114 tokens 115 sentences 20 sentences	grammar examples, oral stories, didatic material, and dictionary examples	Text in Portuguese / and some in English Annotation is not uniform: some sentences have features, others do not.
Makurap (Tupian, Tupai)	178 tokens 178 words 37 sentences	Grammar	Text in English Indicates source of the sentence
Munduruku (Tupian, Munduruku)	1016 tokens 1022 words 158 sentences	Grammar, particular descriptions, other texts	Text in English, Portuguese
Nheengatu (Tupian, Maweti-Guarani)	12621 tokens 12743 words 1239 sentences	grammatical descriptions, fables, myths, coursebooks, and dictionaries	Only original text Indicates source of the sentence
Tupinamba (Tupian, Maweti-Guarani)	4397 tokens 4508 words 581 sentences	cathecisms, letters, poems, theater plays, and grammars (sixteenth and seventeenth century)	Text in English Indicates source of the sentence
Xavante (Macro-jê)	1589 tokens 1597 words 148 sentences	Grammar	Text in English, Portuguese

Issues to discuss

1) **One fact, one problem**:
   In terms of language description/documentation, UD is **semasiologically** organized – typically how reference grammars are organized. Previous experience with standardization in description/documentation: *Lingua questionnaire* by Comrie and Smith (1977), which resulted in the Routledge Reference Grammar series. However, Gast (2009) makes two points regarding the use of a standard questionnaire for documentation: (i) not all questions will be relevant to all languages; (ii) the information elicited from the questionnaire cannot come close to a full coverage of a language.
   UD provide, so to speak, a standard questionnaire for language 'documentation' and Gast's first problem is a significant one for the framework, especially if we wish to come up with guidelines for one to start a treebank from scratch.
   Three questions:
   Question 1: Can we reframe UD categories – mainly UPOS and *deprels* (the *features* require a separate discussion) – so they have a more "onomasiological" (or functional) flavor? Much in line with what Croft (2017) and Croft et al (2017) have already proposed. – This question should be addressed in conjunction with Task 1.3.
   But why is this question relevant for Task 1.1? (BTW, this is not the second question just yet.)
   Question 2: If the starting point of UD is "form", do *all* categories (UPOS and *deprels*) apply to *all* languages? They certainly do not – this question is in the spirit of traditional cross-language investigation: we are basically trying to see if some cross-linguistic category can be found in all languages (see Dryer, 1997).
   Question 3: If this is not the case, then, can we select a few categories that will undoubtably figure in most languages and star drafting guidelines from there?
   A potential solution:
   Instead of writing up one unified "recipe", we write up a sort of a flow chart, with implicational questions which can be designed by a combination of both functional and formal properties of constructions, in order not only to avoid (European) biases but also to allow the analyst some space in the documentation and annotation.

2) Some practical consequences of coming up with "guidelines for beginners"
   a) Try to involve, in some capacity, field workers and grammarians – show some potential benefits of working together with NLP people.
   b) With respect to "How to create a (UD) treebank" section of the website: be stricter with the criteria, particularly morphemic glossing – otherwise, typological work will become increasingly difficult.
      - A good opportunity to integrate UD and UniMorph.
      - Use automatic "morphologizers" in combination with human evaluation (e.g.Cyrino & Mattos, 2020) .
   c) With respect to "Guidelines for language-specific documentation": reframe these guidelines so that language-specific documentation is more of a typological sketch for NLP (UD) purposes. These documentation pages are still pretty much uneven.

References

COMRIE, B.; SMITH, N. (1977) Lingua descriptive series: questionnaire. *Lingua* 42 (1): 1-72.

CROFT, W. (2017) Using typology to develop guidelines for Universal Dependencies. Invited Talk.

CROFT, W. et al. (2017) Linguistic typology meets Universal Dependencies. *Proceedings of the 15th TLT*, p. 63-75.

CYRINO, J. P. L.; MATTOS, E. B. (2020) An exploratory study into morpheme classification through hierarchical clustering for typological comparison. *Revista do GELNE* 22 (2): 395-407.

DRYER, M. (1997) Are grammatical relations universal?, in Bybee, J. et al. (eds.) *Essays on language function and language type.* Amsterdam: John Benjamins, p. 115-144.

GAST, V. (2009) A contribution of 'two-dimensional' language description: the typological database of intensifiers and reflexives, in Everaert, M. et al (eds.) *The use of databases in cross-linguistic studies.* Berlin: Mouton de Gruyter, p. 209-234.

ACTIVE PART 2: Structure of the paper

7 minutes approx.

Paper

Title: _____
Index:
(topics to be developed and the members interested in each one)
Abstract: 150-200 words
Identify the best journal: _____
Dates for writing – deadline: June-2024?
Dates for sending – deadline July-2024?

THANK YOU

Lucía Amorós-Poveda, lamoros@um.es
Didactic and School Organization Department (UM, ISEN & UNIR)
Center for Studies on Educational Memory (CEME)

Friday 9 February, 2024
University of Naples "L'Orientale", ITALY
15:15-16:30 Parallel working sessions. Room 1.5

SELECTED MoU's REFERENCES

McDonald, Petrov, Hall. 2011. Multi-Source Transfer of Delexicalized Dependency Parsers. Proc. EMNLP

Nivre, de Marneffe, Ginter, Goldberg, Hajič, Manning, Pyysalo, Schuster, Tyers, Zema (2020): Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. Proc. LREC 2020

Nivre, Rimell, McDonald, Gomez-Rodríguez (2010): Evaluation of Dependency Parsers on Unbounded Dependencies, Proc. COLING 2010

Phillips (2014): How Diversity Makes Us Smarter. Scientific American, October

Pires, Schlinger, Garrette (2019): How Multilingual is Multilingual BERT? Proc. ACL-2019

Ponti, Reichart, Korhonen, Vulić (2018): Isomorphic Transfer of Syntactic Structures in Cross-Lingual Natural Language Processing. Proc. ACL 2018.