Challenges and Strategies in Cross-Cultural NLP

NLP Workshop at ITU
15 March 2022

Daniel Hershcovich

UNIVERSITY OF COPENHAGEN

This is a group effort

Challenges and Strategies in Cross-Cultural NLP

Daniel Hershcovich[1] **Stella Frank**[2] **Heather Lent**[1] **Miryam de Lhoneux**[1,3,4]
Mostafa Abdou[1] **Stephanie Brandl**[1] **Emanuele Bugliarello**[1] **Laura Cabello Piqueras**[1]
Ilias Chalkidis[1] **Ruixiang Cui**[1] **Constanza Fierro**[1] **Katerina Margatina**[5]
Phillip Rust[1] **Anders Søgaard**[1]
[1]University of Copenhagen [2]University of Trento [3]Uppsala University
[4]KU Leuven [5]University of Sheffield
dh@di.ku.dk

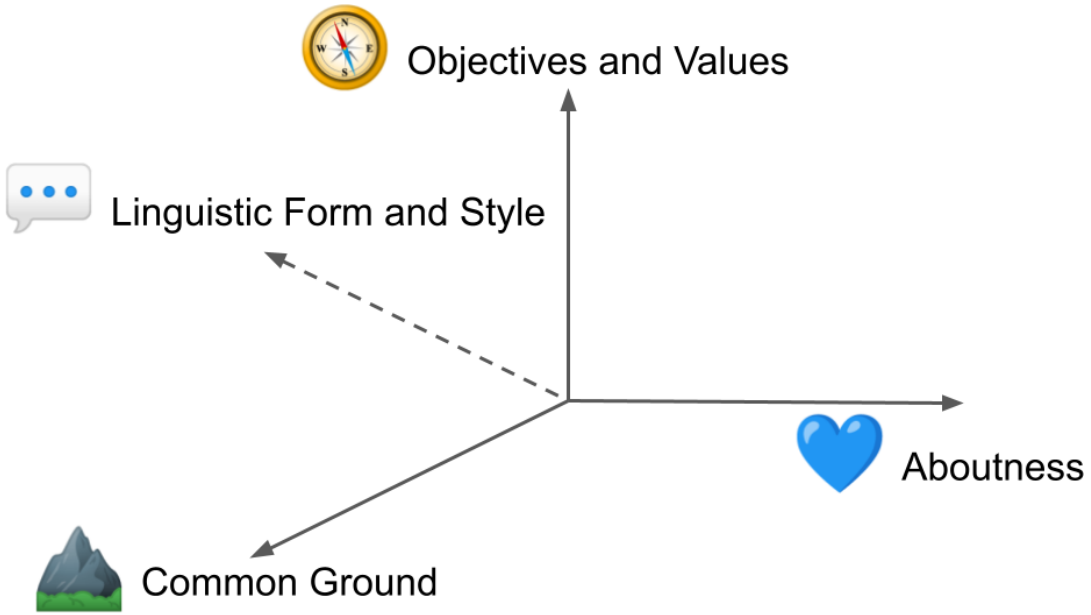
coAStaL

ACL 2022
22ND – 27TH MAY | 60TH MEETING | DUBLIN

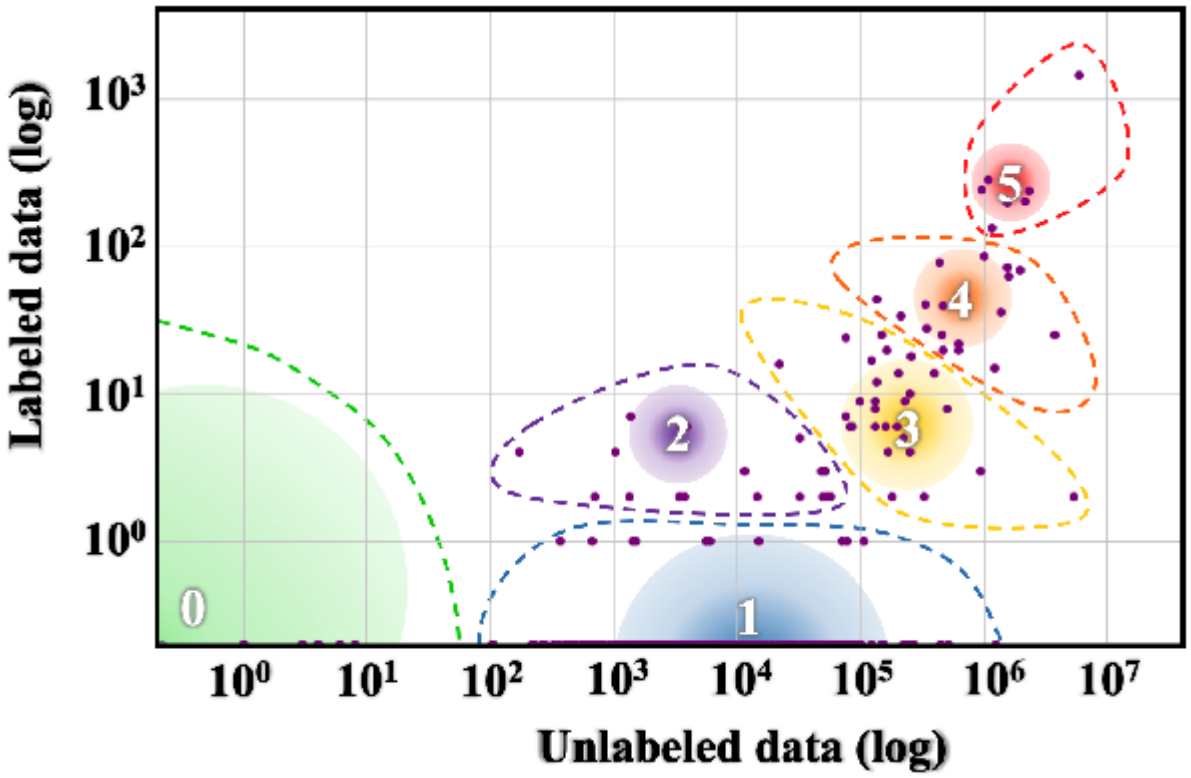
Theme: "Language Diversity: from Low-Resource to Endangered Languages"

TL;DR

- NLP is for people (not just languages)
- Culture is multidimensional
- Objectives are conflicting
- Generalisation-representation trade-off



Resource disparity for languages

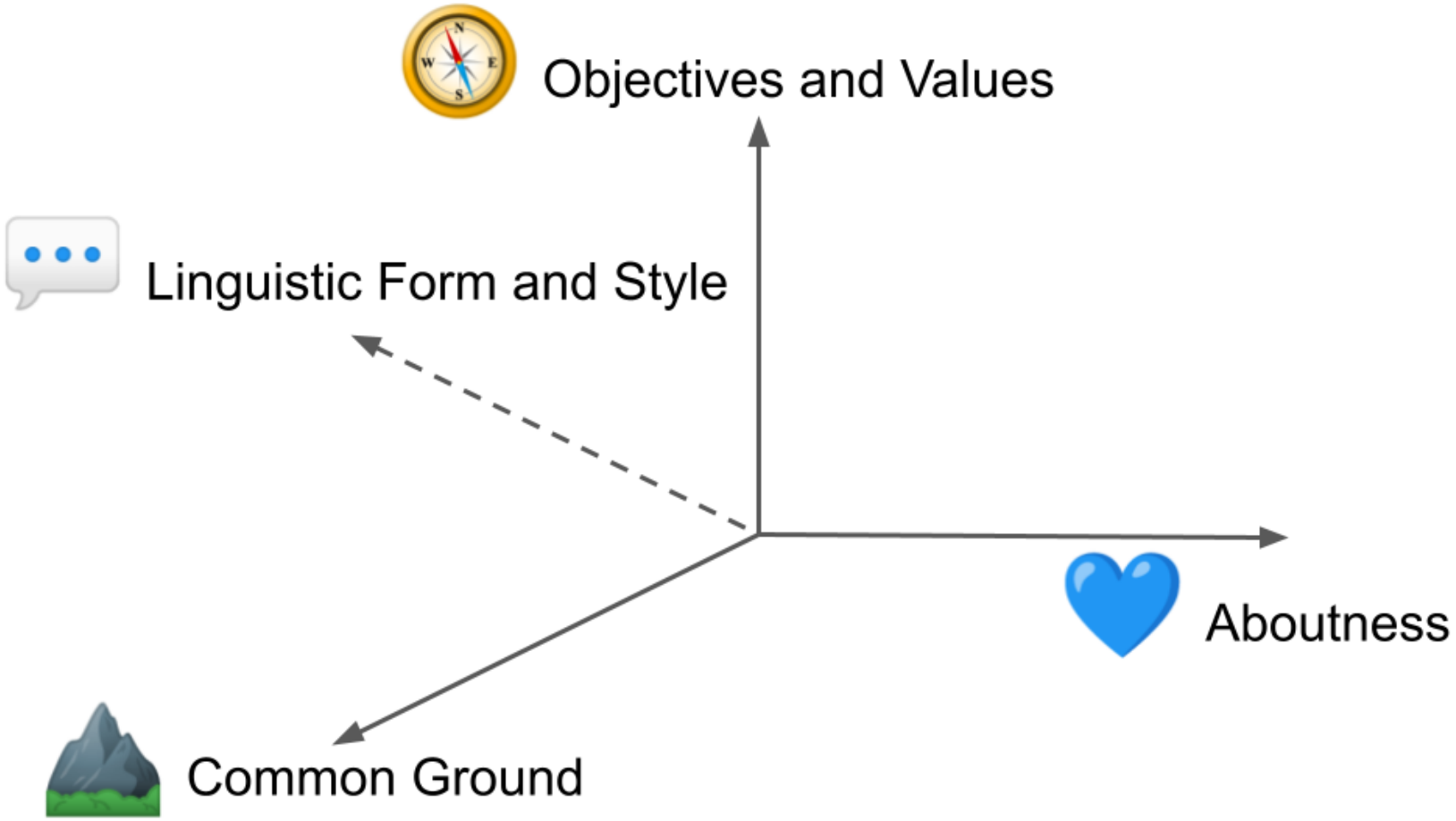


The State and Fate of Linguistic Diversity and Inclusion in the NLP World
(Joshi et al., ACL 2020)

Social factors are important

The Importance of Modeling Social Factors of Language: Theory and Practice
(Hovy & Yang, NAACL 2021)

What is culture?



Form

- *How* we express ourselves in language
- Morphosyntax, word choice…
- Stylistic aspects of linguistic form:

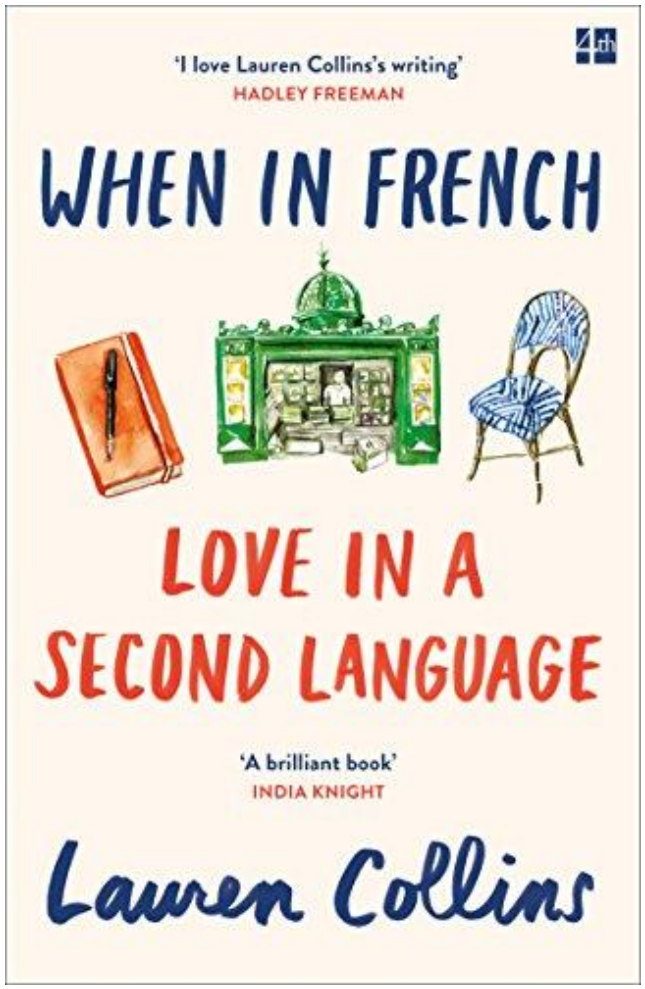
Directness

Formality

Politeness

Emotional expression

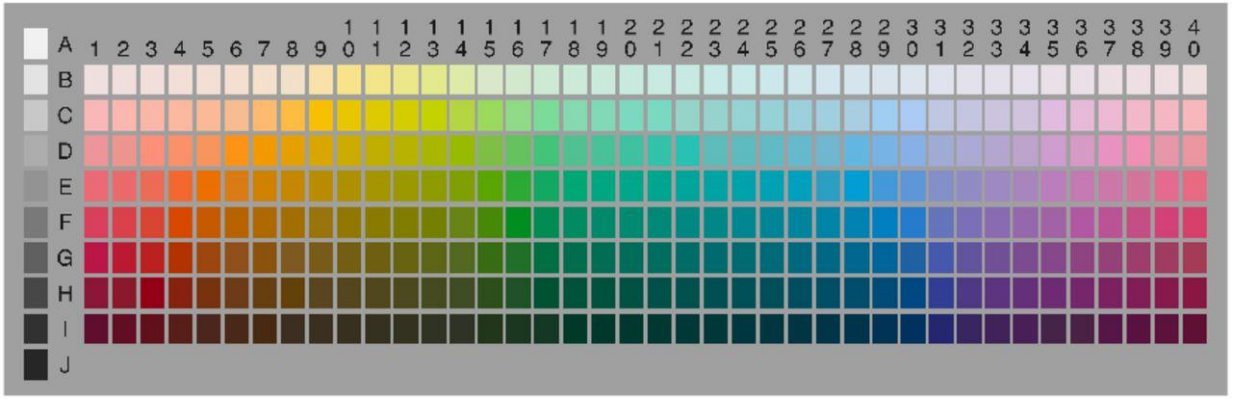
Common ground



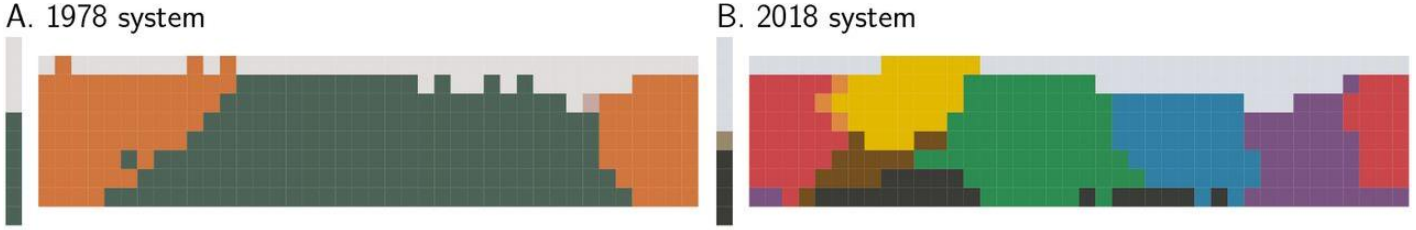
- Conceptualisation
- Commonsense
- Stories
- Metaphors
- Clichés
- ...

Conceptualisation

World Color Survey



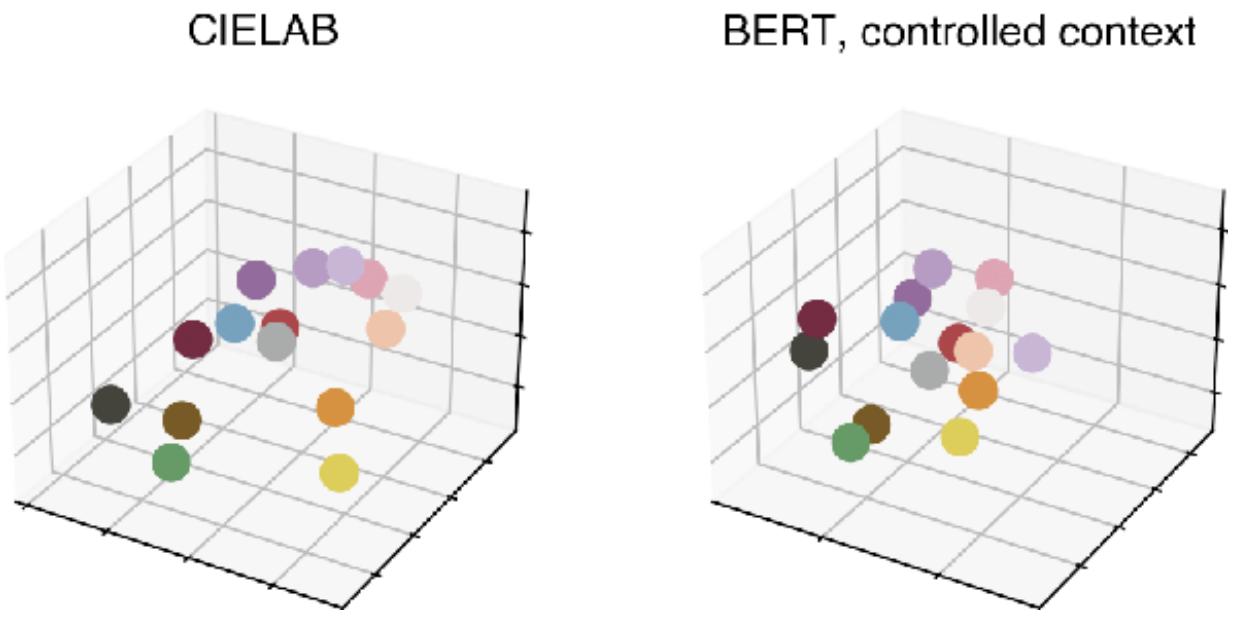
Nafaanra, a language of Ghana and Côte d'Ivoire



The evolution of color naming reflects pressure for efficiency: Evidence from the recent past

(Zaslavsky et al., Journal of Language Evolution 2022)

Probing colour



English BERT aligns with English-speaking American humans.
(What about others?)

Can Language Models Encode Perceptual Structure Without Grounding? A Case Study in Color
(Abdou et al., CoNLL 2021)

Commonsense

Bola basket (Indonesian) *Mpira wa kikapu* (Swahili) 篮球 (Chinese) *Basketbol* (Turkish) கூடைப்பந்தாட்டம் (Tamil)

Visually Grounded Reasoning across Languages and Cultures

(Liu et al., EMNLP 2021)

Commonsense

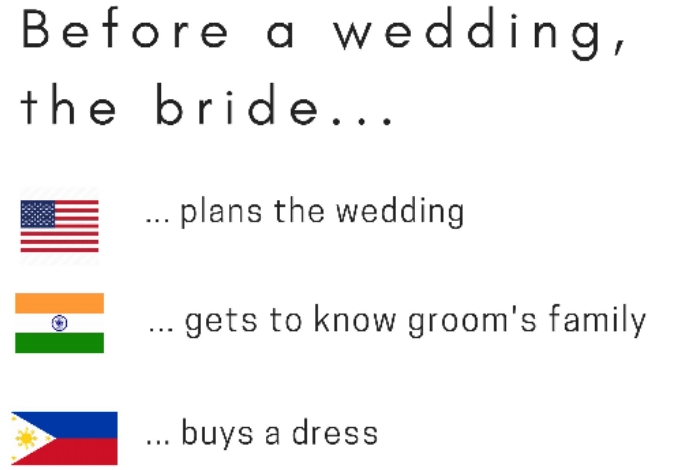
"Commonsense is the basic level of **practical knowledge** and **reasoning** concerning everyday **situations** and **events** that are **commonly** shared among **most** people."

Commonsense Reasoning for Natural Language Processing
(Sap et al., ACL 2020 Tutorial)

Stevie Wonder announces he'll be having kidney surgery during London concert

UnifiedQA: Crossing Format Boundaries with a Single QA System
(Khashabi et al., Findings 2020)

Commonsense



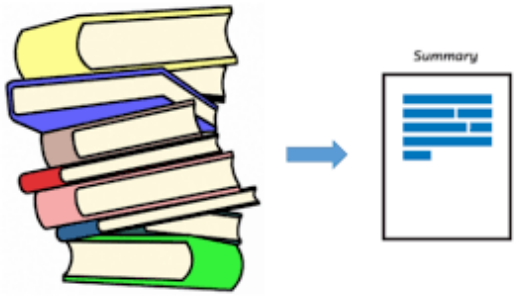
(a) Cultural differences in *wedding* ritual.

(b) Cultural differences in *funeral* ritual.

Towards an Atlas of Cultural Commonsense for Machine Reasoning
(Acharya et al., CSKGs 2021)

Aboutness

- What do we *care about*?
- Related to topic/domain



Visual concepts

Beer reviews

News generation

Values

- *Why* are we doing this?
- Users may have different goals, often implicit
- Common meta-objectives in *NLP research culture*

Accuracy

Fairness

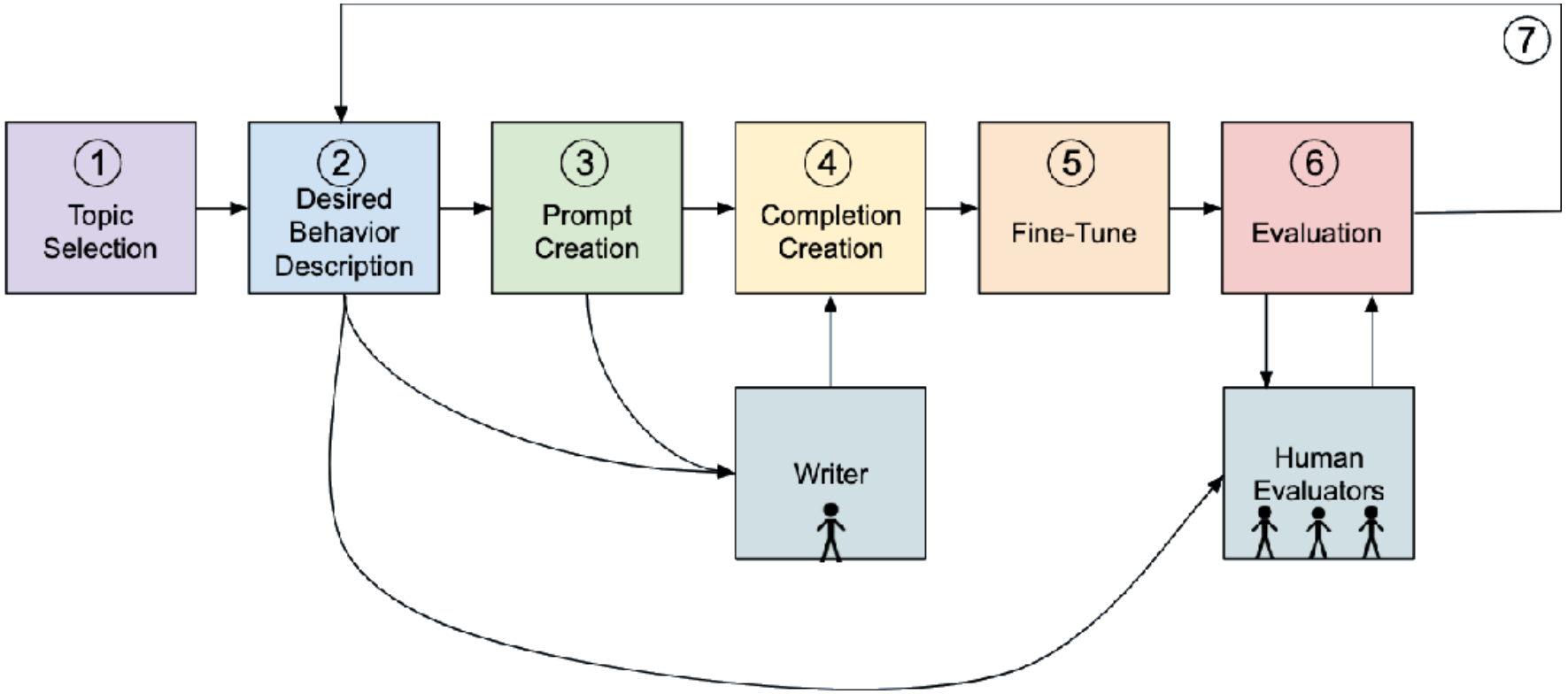
Robustness

Interpretability

Conflicting objectives?

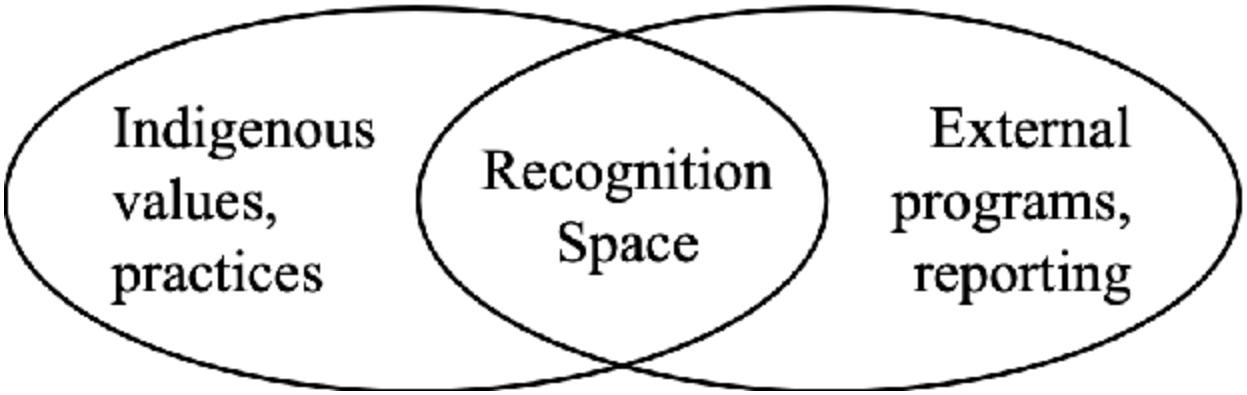
- Researchers
- Practitioners
- End-users
- Regulators

AI alignment



Process for Adapting Language Models to Society (PALMS) with Values-Targeted Datasets
(Solaiman & Dennison, NeurIPS 2021)

Language technology for all (potential) users



Decolonising Speech and Language Technology
(Bird, COLING 2020)

Strategies

DATA

MODELS

TASKS

Data

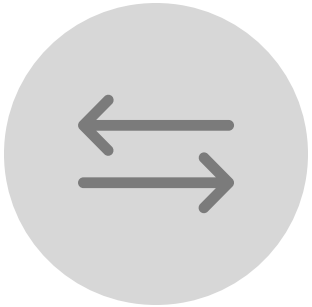
Selection

Annotation

Projection

Models

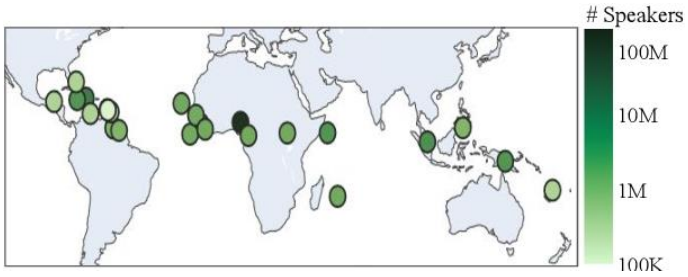
TRAINING



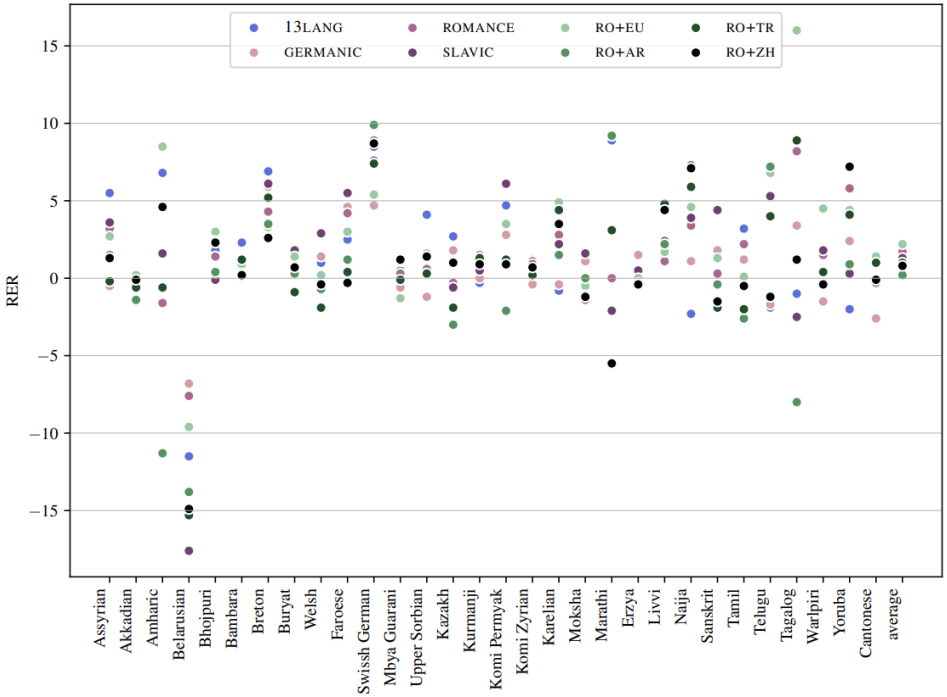
TRANSFER

PRE-TRAINED LANGUAGE MODELS

Balancing generalisation and representation



On Language Models for Creoles
(Lent et al., CoNLL 2021)



Zero-Shot Dependency Parsing with Worst-Case Aware Automated Curriculum Learning
(de Lhoneux et al., ACL 2022)

Probing

Context	Indonesia is the Germany of the Asean. So then, Malaysia is the France.
Question	What country is Indonesia similar to?
Answer	Germany
Prediction	Malaysia

Locke's Holiday: Belief Bias in Machine Reading (Søgaard, EMNLP 2021)

"Stance bias"?

(Work in progress)

For each of the following, indicate how important it is in your life. Would you say it is (*read out and code one answer for each*):

		Very important	Rather important	Not very important	Not at all important
Q1	Family	1	2	3	4
Q2	Friends	1	2	3	4
Q3	Leisure time	1	2	3	4
Q4	Politics	1	2	3	4
Q5	Work	1	2	3	4
Q6	Religion	1	2	3	4

Cross-cultural translation

Bridging between cultures
as a task

"I saw Merkel eating a Berliner from Dietsch on the ICE"

I saw Biden eating a Boston Cream from Dunkin' Donuts on the Acela

Adapting Entities across Languages and Cultures
(Peskov et al., Findings 2021)

Style transfer

Entity adaptation

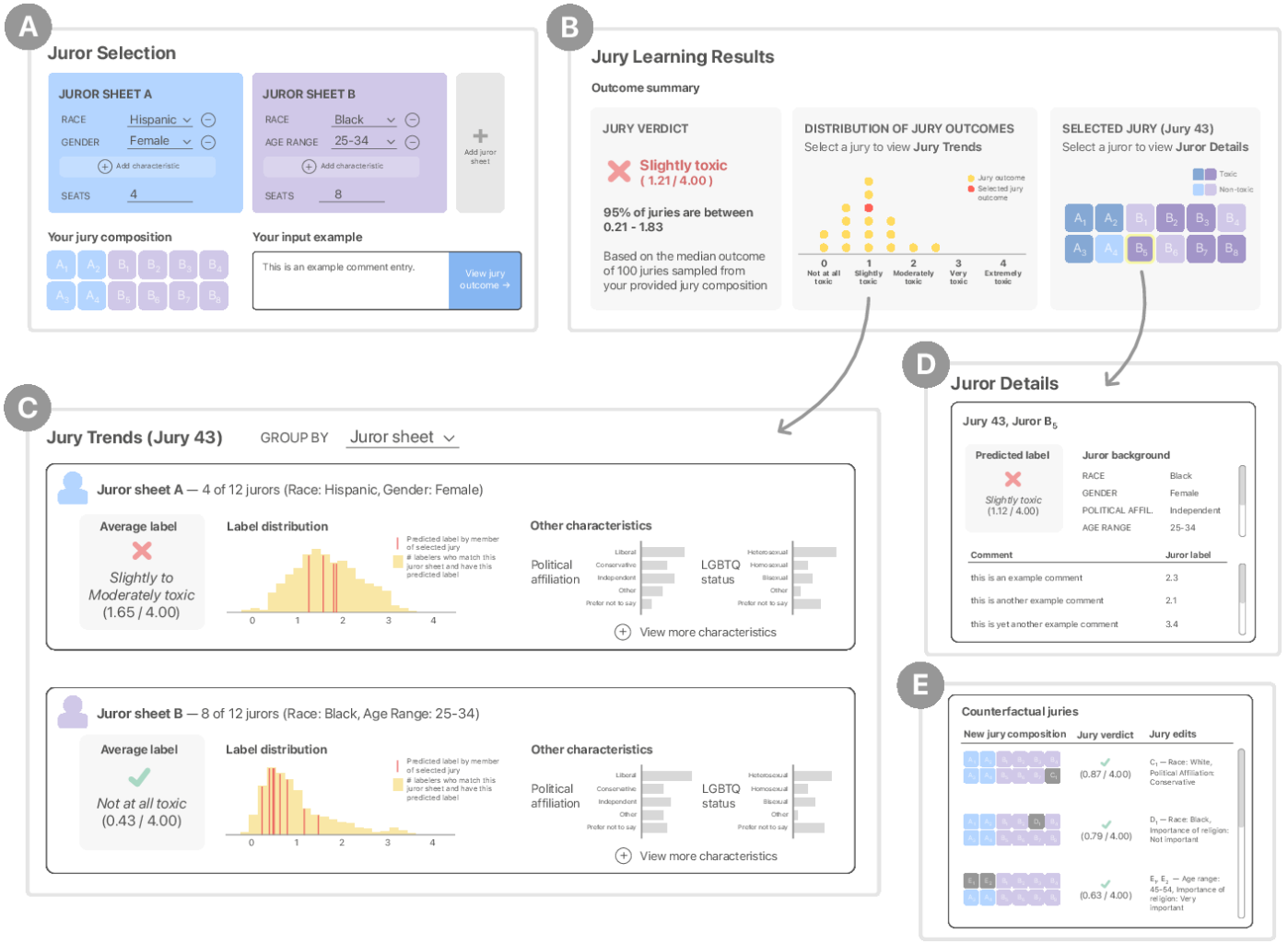
Explanation by analogy

Levels of granularity

- Linguistic variation within a "language"
- Also applies to cultures

Idiolect	**Sociolect, dialect**	**Standardised language**	**Language, language family**
Individual	Social group or region	Country	

Multi-granularity adaptation



Jury Learning: Integrating Dissenting Voices into Machine Learning Models
(Gordon et al., CHI 2022)

Join us!

- Hiring 3-year postdoc with NorS at UCPH
- Linguistic and societal change in 19-century Danish and Norwegian novels

https://jobportal.ku.dk/videnskabelige-stillinger/?show=155976

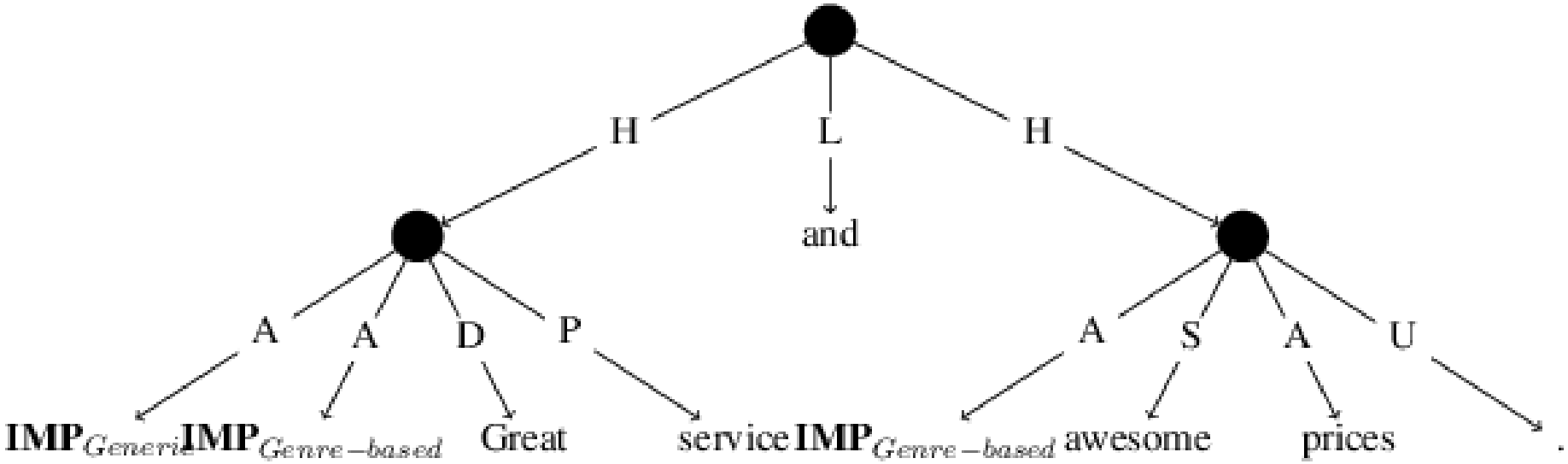
coASTaL

UNIVERSITY OF COPENHAGEN
FACULTY OF HUMANITIES

Thank you

danielhers.github.io
dh@di.ku.dk

Implicit arguments



Great service and awesome prices.

Refining Implicit Argument Annotation for UCCA
(Cui & Hershcovich, DMR 2020)

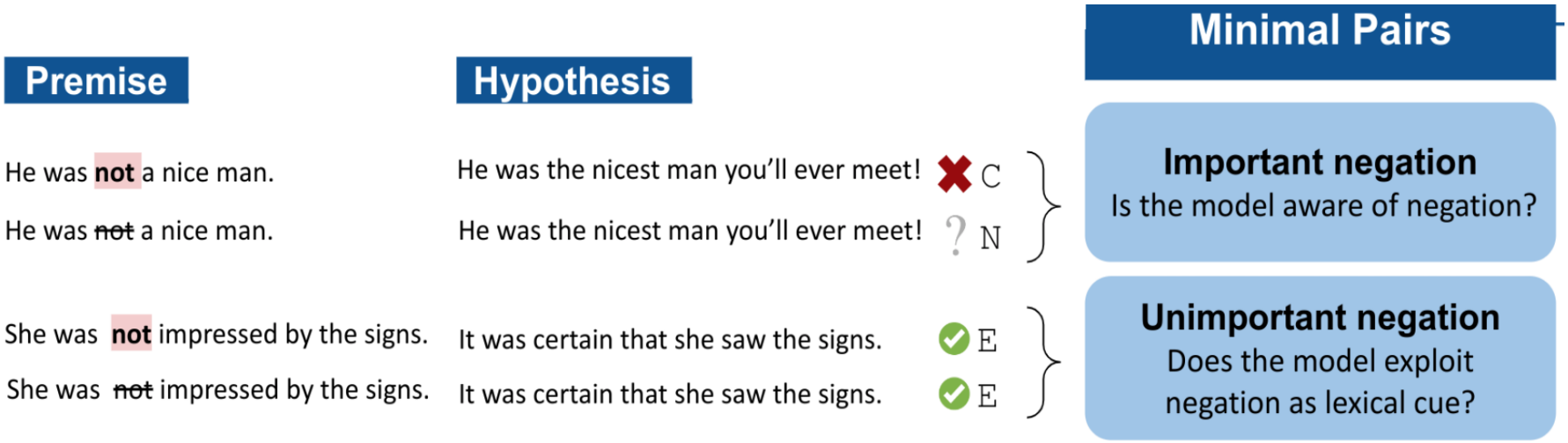
Great Service! Fine-grained Parsing of Implicit Arguments
(Cui & Hershcovich, IWPT 2021)

Multilingual compositional generalisation

Lang.	CWQ field	Content
En	questionWithBrackets	Did [Lohengrin] 's male actor marry [Margarete Joswig]
	questionPatternModEntities	Did M0 's male actor marry M2
He	questionWithBrackets	האם השחקן הגברי של [לוהנגרין] התחתן עם [מרגרט יוסוויג]
	questionPatternModEntities	האם השחקן הגברי של M0 התחתן עם M2
Kn	questionWithBrackets	[ಲೋಹೆಂಗರ್ರಿನ್] ಅವರ ಪುರುಷ ನಟ ವಿವಾಹವಾದರು [ಮಾರ್ಗರೇಟ್ ಜೋಸ್ವಿಗ್]
	questionPatternModEntities	M0 ನ ಪುರುಷ ನಟ M2 ಅನ್ನು ಮದುವೆಯಾಗಿದ್ದಾರೆ ಯ
Zh	questionWithBrackets	[Lohengrin]的男演员嫁给了[Margarete Joswig]吗
	questionPatternModEntities	M0的男演员和M2结婚吗
	sparql	`ASK WHERE { ?x0 wdt:P453 wd:Q50807639 . ?x0 wdt:P21 wd:Q6581097 . ?x0 wdt:P26 wd:Q1560129 . FILTER ( ?x0 != wd:Q1560129 )}`
	sparqlPatternModEntities	`ASK WHERE { ?x0 wdt:P453 M0 . ?x0 wdt:P21 wd:Q6581097 . ?x0 wdt:P26 M2 . FILTER ( ?x0 != M2 )}`
	recursionDepth	20
	expectedResponse	True

Multilingual Compositional Wikidata Questions
(Cui et al., arXiv)

Probing negation



Multilingual BERT for XNLI relies on lexical cues.

A Multilingual Benchmark for Probing Negation-Awareness with Minimal Pairs
(Hartmann et al., CoNLL 2021)