

Challenges and Strategies in Cross-Cultural NLP

Daniel Hershcovich, Stella Frank,
Heather Lent, Miryam de Lhoneux,
Mostafa Abdou, Stephanie Brandl,
Emanuele Bugliarello, Laura Cabello Piqueras,
Ilias Chalkidis, Ruixiang Cui, Constanza Fierro,
Katerina Margatina, Phillip Rust
and Anders Søgaard

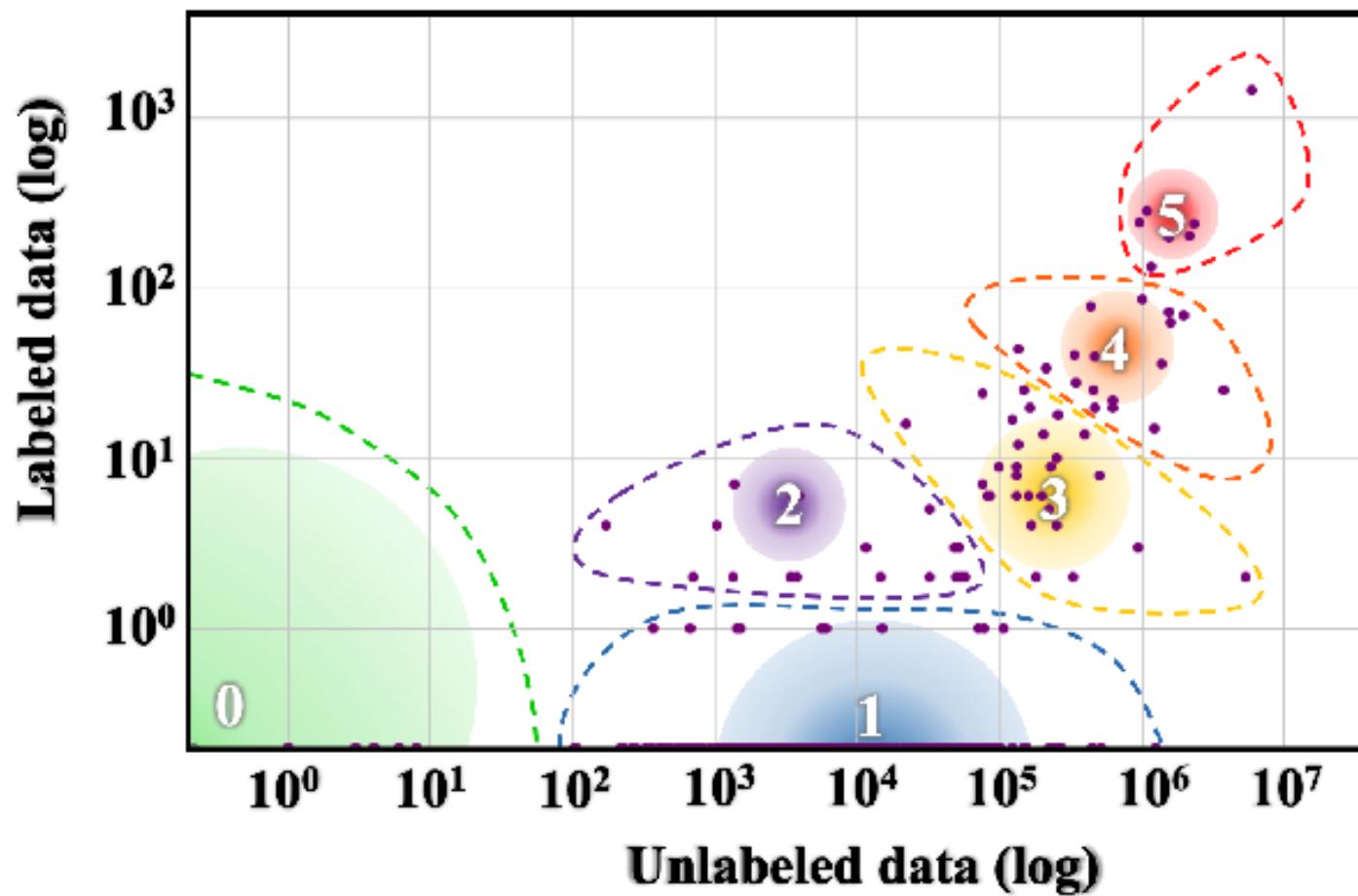
Published at ACL 2022

LIIR Journal Club, KU Leuven
21 April 2022

UNIVERSITY OF COPENHAGEN



Resource disparity for languages



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

Social factors

NLP is for people (not just languages)



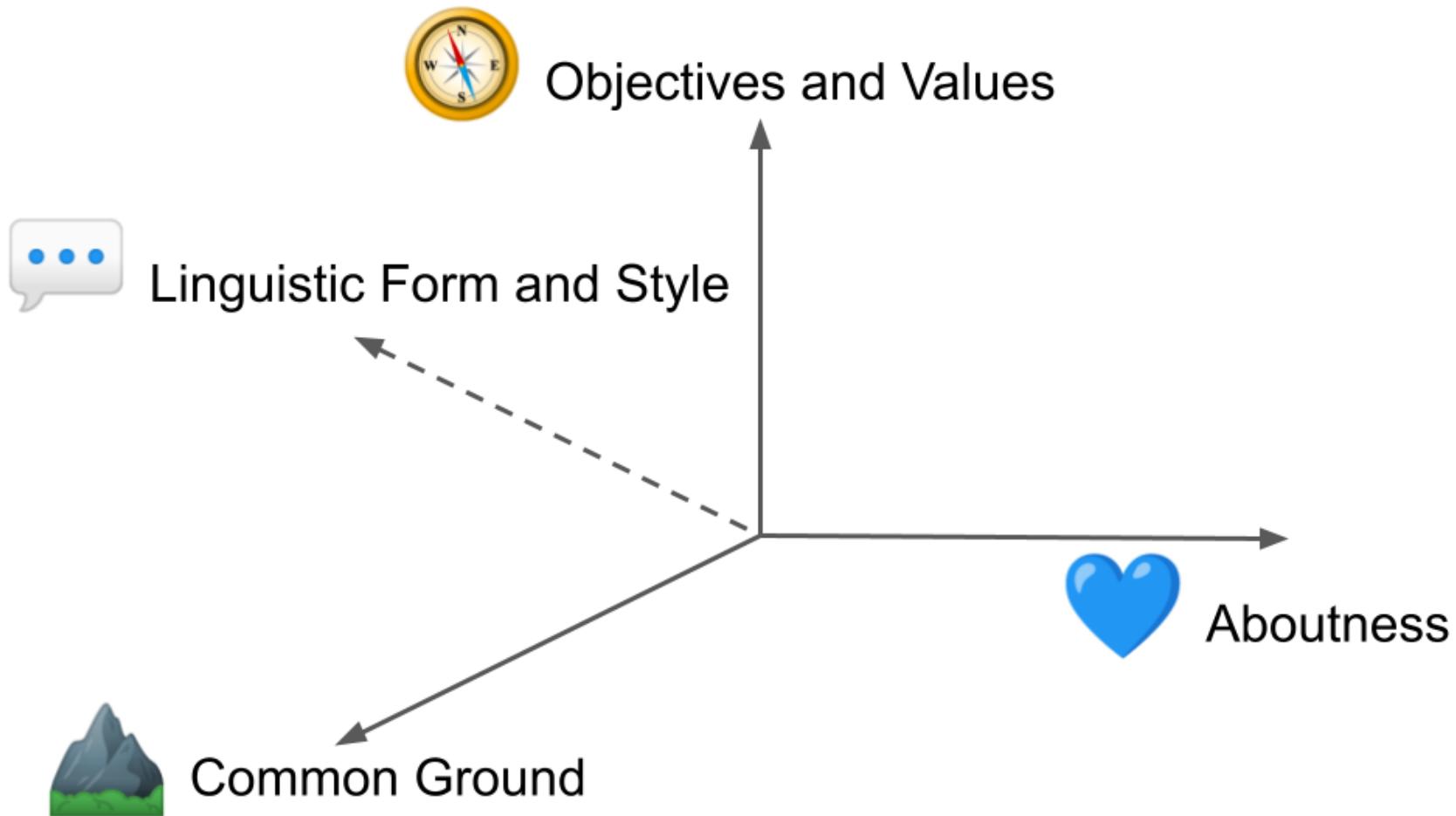
The Importance of Modeling Social Factors of Language: Theory and Practice
(Hovy & Yang, NAACL 2021)

Social bias in language models

Models	Demographics Alignment															
bert-base-cased	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
bert-base-uncased	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
bert-base-multilingual-cased	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
bert-large-cased	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
bert-large-uncased	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
distilbert-base-uncased	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
albert-base-v2	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
albert-large-v2	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
albert-xxlarge-v2	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
roberta-base	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
roberta-large	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
google/electra-large-generator	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
google/electra-small-generator	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
gpt2	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
gpt2-medium	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
gpt2-large	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
gpt2-xl	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
Group	以人为中心的	中性的	种族主义的	性别平等的	中性的	中性的	中性的	中性的	中性的							
Mean Rank	3.1	3.4	4.0	6.1	6.1	8.1	8.1	9.2	9.8	9.9	10.3	10.3	10.8	11.1	12.0	13.8

Sociolectal Analysis of Pretrained Language Models
(Zhang et al., EMNLP 2021)

Dimensions of culture



Form

*How we express
ourselves in
language*

Morphosyntax

Word choice

Style

Style

Stylistic aspects of
linguistic form:

Directness

Formality

Politeness

Emotional expression

Levels of granularity

Linguistic and cultural variation within groups



Idiolect	Sociolect, dialect	Standardised language	Language, language family
Individual, personality	Social group or region, sub-culture	Country, national culture	International cultures

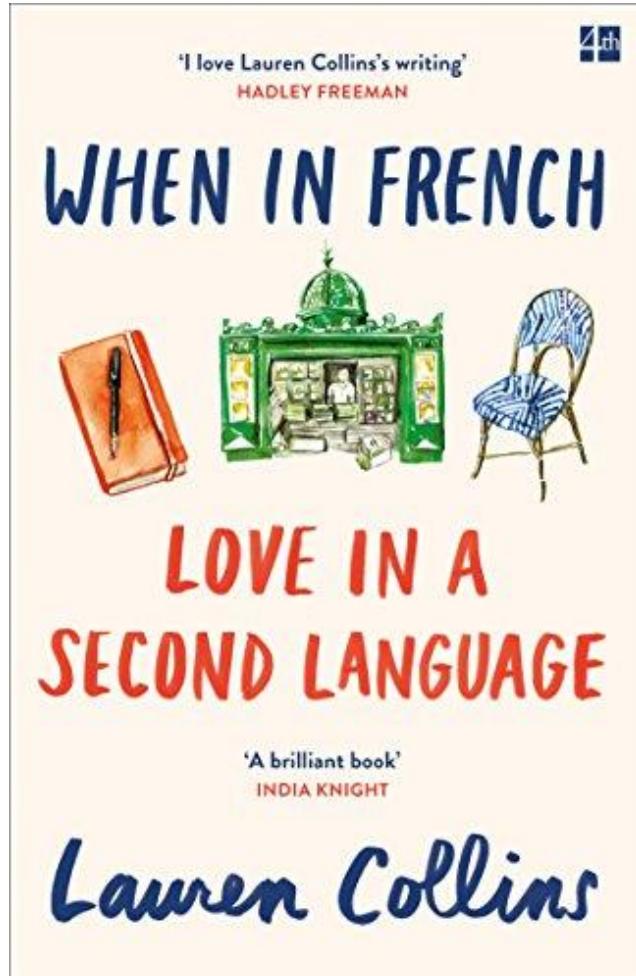
Common ground

Shared
knowledge based
on which people
reason and
communicate

Conceptualisation

Commonsense

Common ground 🏘️



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

Conceptualisation

Objects

Colours

Kinship

Space

Time

Events

Objects

Swahili
leso



Visually Grounded Reasoning across Languages and Cultures
(Liu et al., EMNLP 2021)

Events and rituals



Tamil
ஜல்லிக்கட்டு
jallikattu

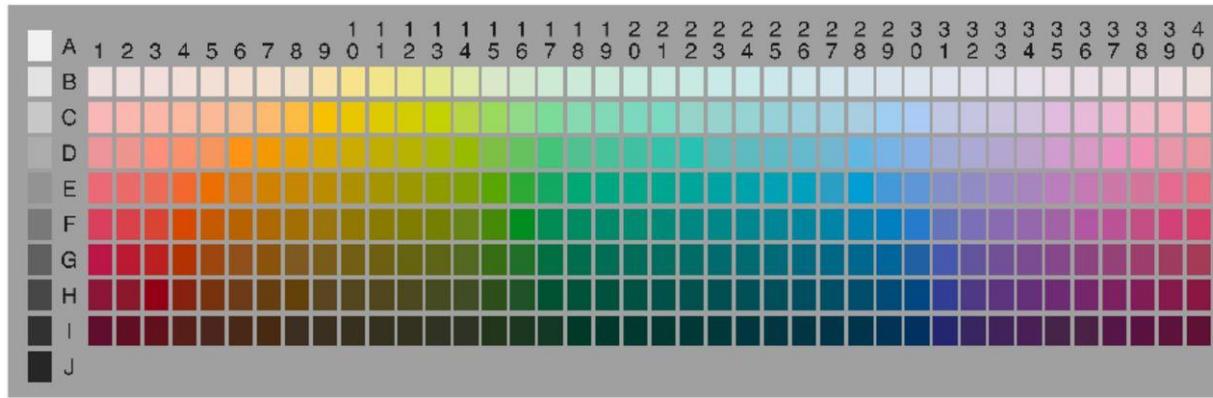
Visually Grounded Reasoning across Languages and Cultures

(Liu et al., EMNLP 2021)

Visual concepts include culture-specific activities that cannot be mapped across cultures.

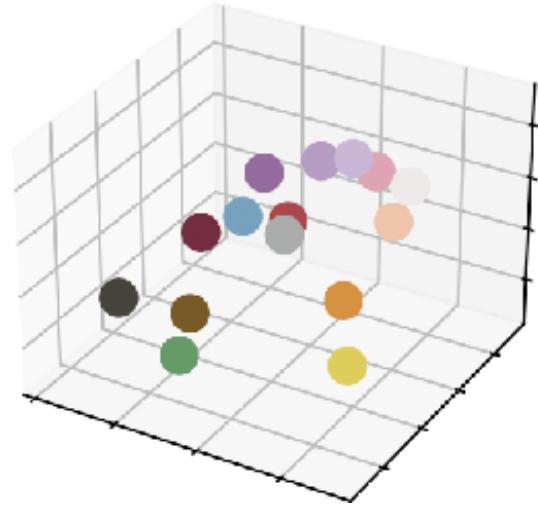
Colour

World Colour Survey

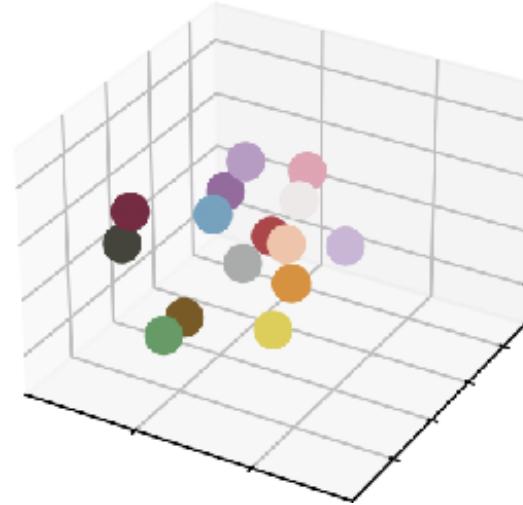


Probing colour

CIELAB



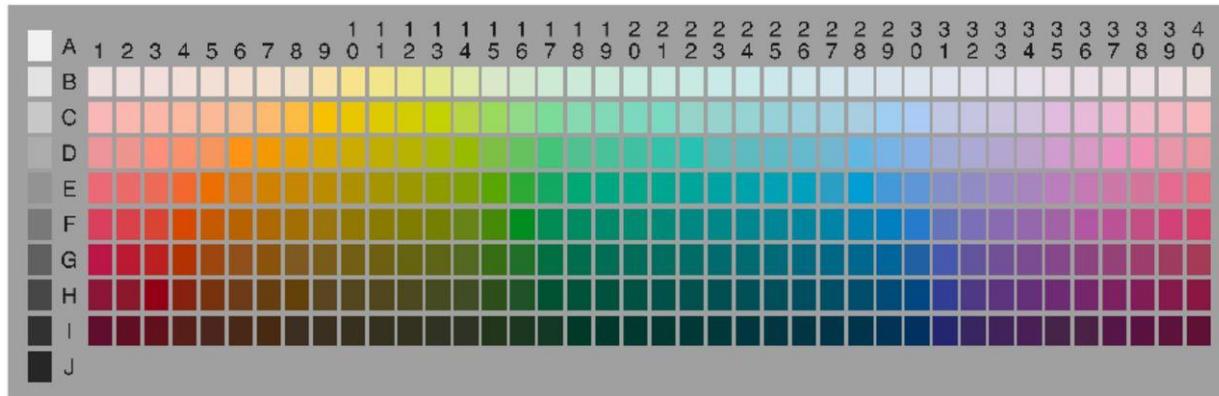
BERT, controlled context



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

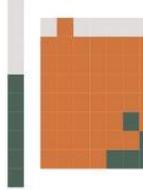
[Can Language Models Encode Perceptual Structure Without Grounding? A Case Study in Color](#)
(Abdou et al., CoNLL 2021)

Differences in colour grounding

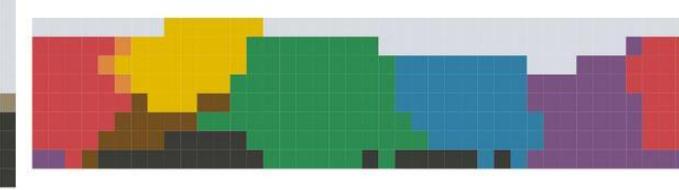


Nafaanra, a language of Ghana and Côte d'Ivoire

A. 1978 system



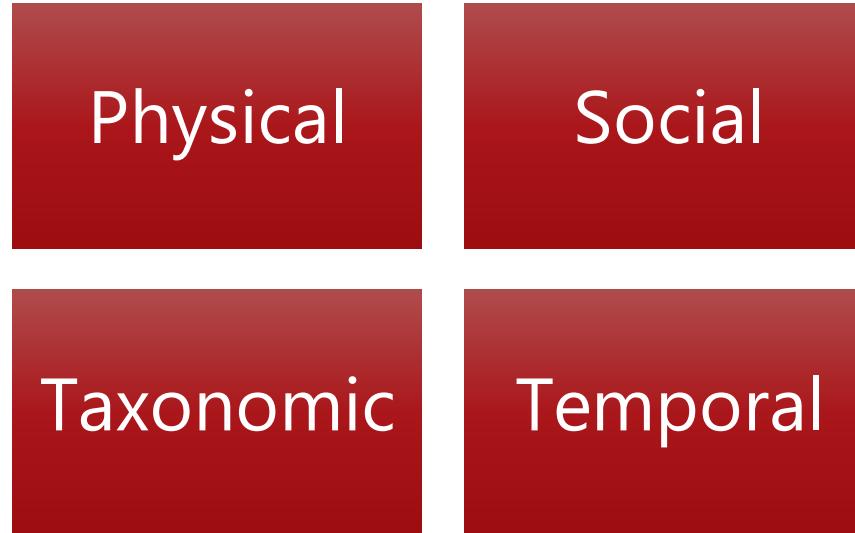
B. 2018 system



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

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

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)

Culture-dependent

Visual commonsense



Bola basket (Indonesian)



Mpira wa kikapu (Swahili)



篮球 (Chinese)



Basketbol (Turkish)



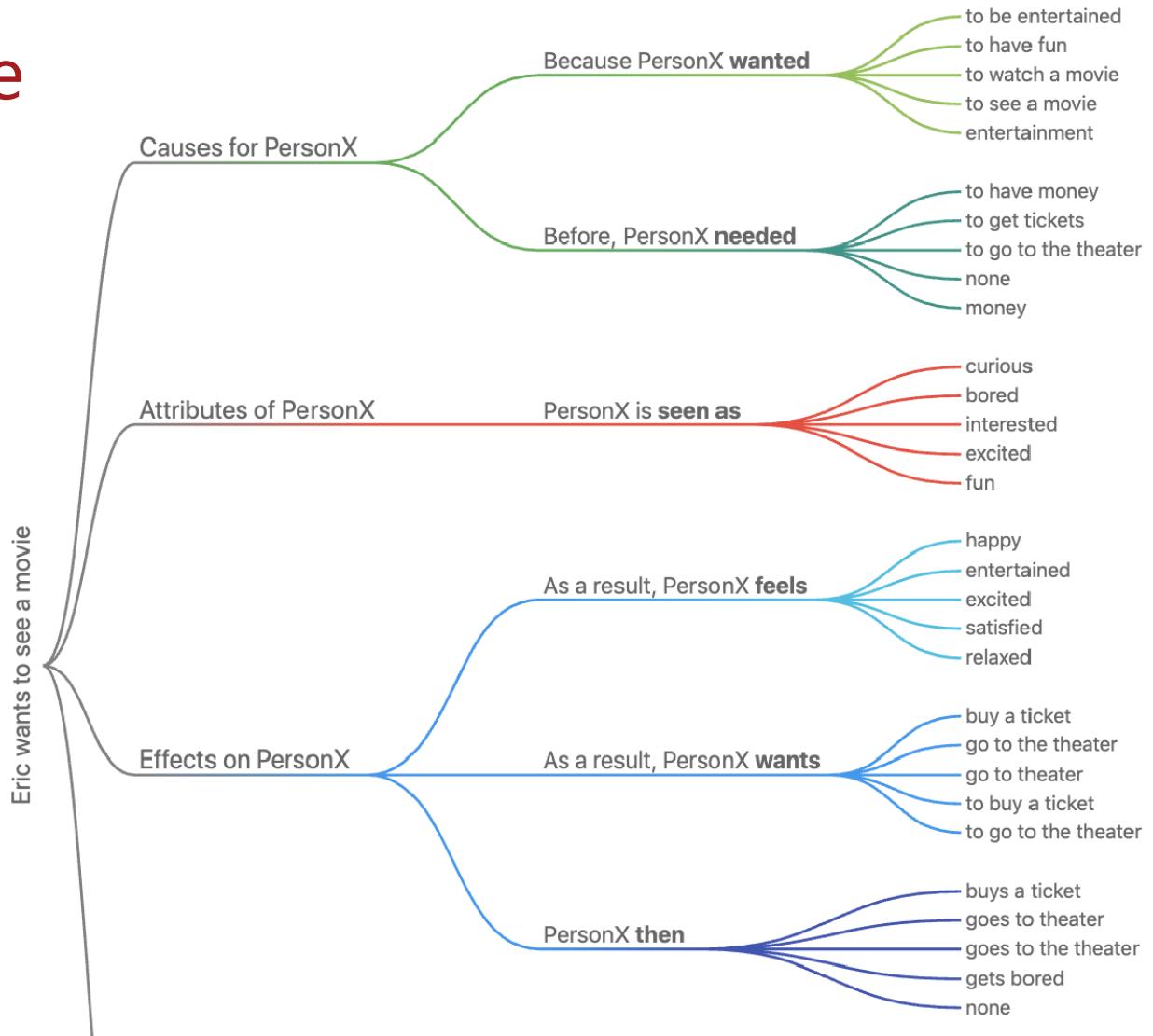
கூடைப்பந்தாட்டம் (Tamil)

Visually Grounded Reasoning across Languages and Cultures

(Liu et al., EMNLP 2021)

Commonsense

Some knowledge is "universal", other culture-specific



COMET: Commonsense Transformers for Automatic Knowledge Graph Construction
(Bosselut et al., ACL 2019)

Commonsense

Before a wedding,
the bride...



... plans the wedding



... gets to know groom's family



... buys a dress

A funeral usually
takes place...



... in church or a funeral home



... at cremation / funeral grounds



... at home

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

Knowledge bias in language models

“[X] was created in [Y]”

en

Japan (170), Italy (56)

de

Deutschland (217), Japan (70)

nl

Nederland (172), Italië (50)

it

Italia (167), Giappone (92)

The language of prompting affects the model's answer to prompts

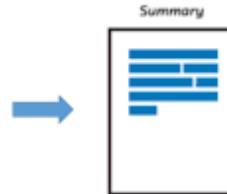


Multilingual LAMA: Investigating Knowledge in Multilingual Pretrained Language Models (Kassner et al., EACL 2021)

Aboutness ❤️

What content do people *care about*?

- Related to topic/domain



Entities



Experiences



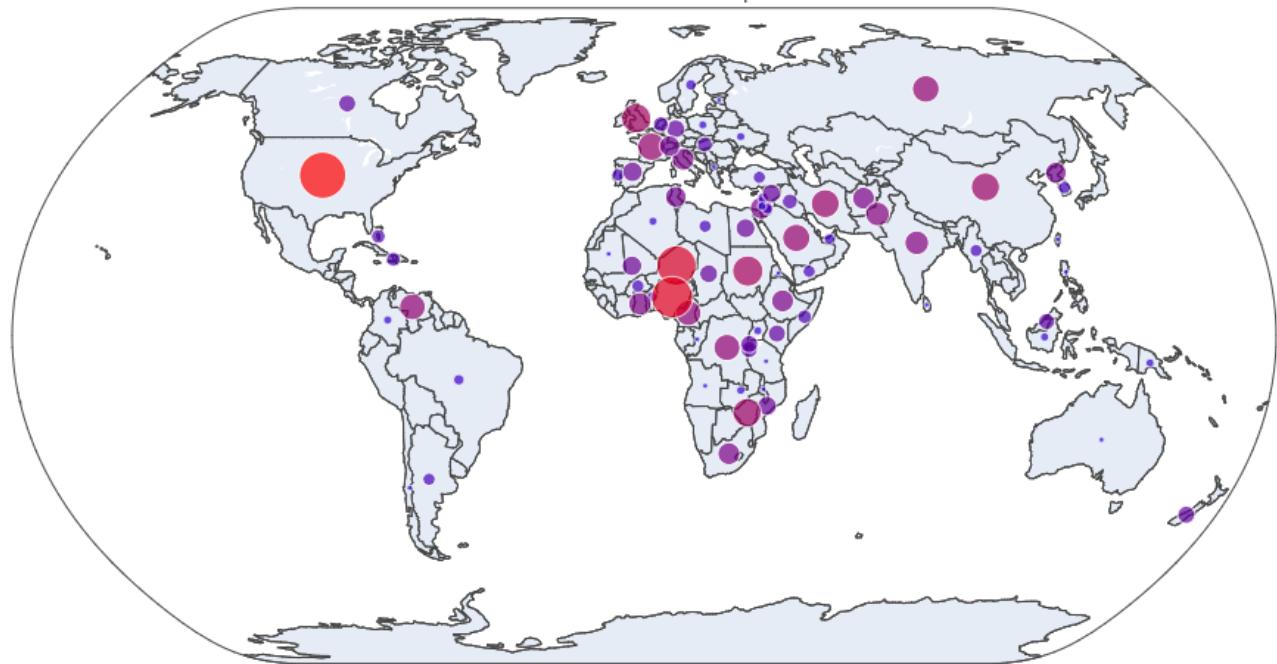
Aspects

Entities

USA & Europe
are over-
represented in
datasets across
languages

Dataset Map: Masakhaner hausa

Dataset Entities Map

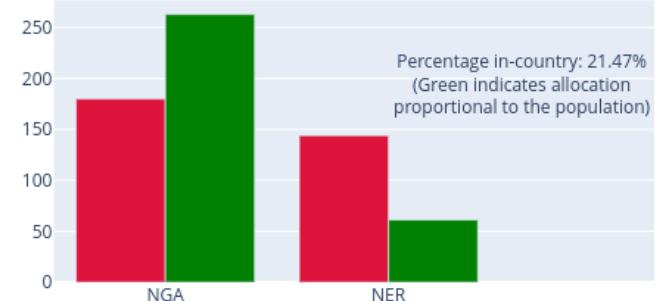


Top-10 Represented Countries

Countries Missing: 143 of 243 (58.85%)



Main Countries where language is spoken.

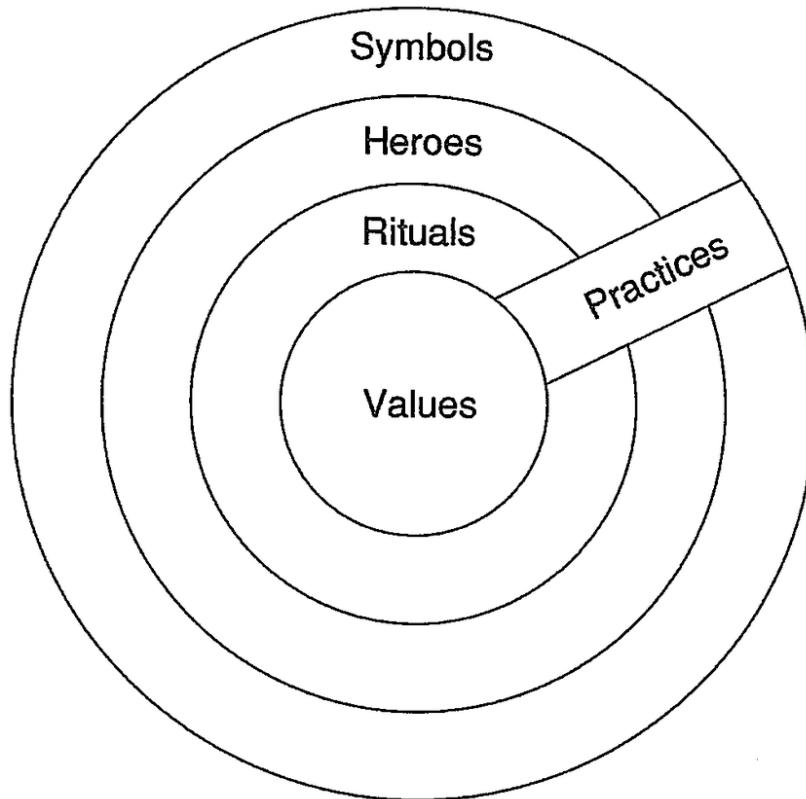


Dataset Geography: Mapping Language Data to Language Users
(Faisal et al., ACL 2022)

Values

Objectives and
goals
people strive for

- What is considered
desired or desirable



Cultures and Organizations: Software of the Mind
(Hofstede, 1991)

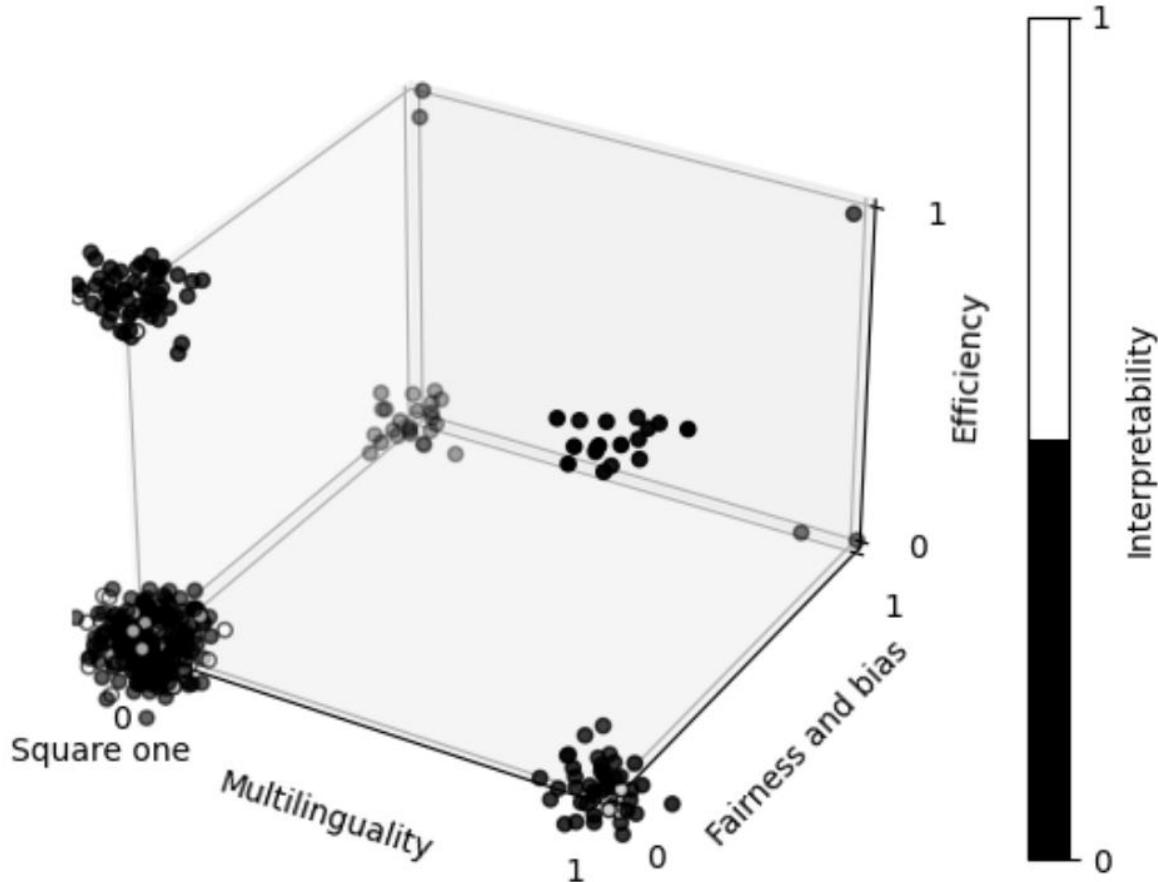
(Meta) values

Why are we doing NLP?

- Users may have different goals, often implicit

No single correct answer.
Changing the World by Changing the Data (Rogers, ACL 2021)

Common meta-objectives in NLP



Accuracy,
fairness,
etc. reflect
the values
of NLP
researchers

Square One Bias in NLP: Towards a Multi-Dimensional Exploration of the Research Manifold (Ruder et al., ACL 2022)

Conflicting objectives between stakeholders



Researchers



Practitioners



End-users



Affected communities

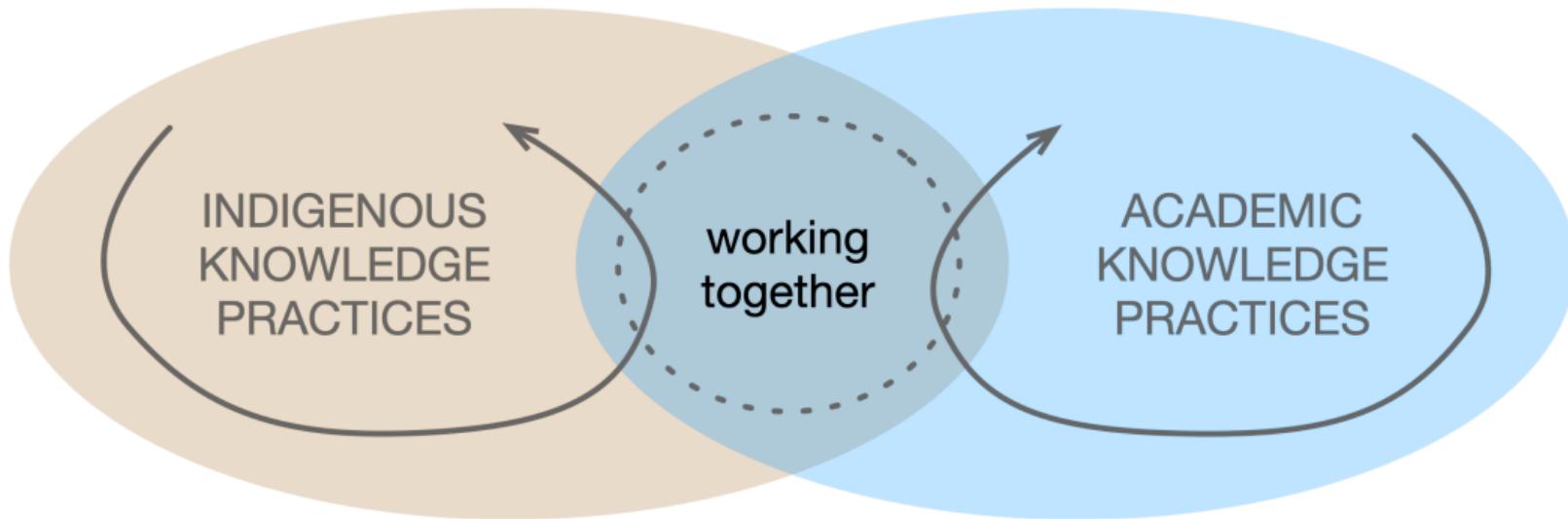


Regulators



Transparent values
facilitate adaptation
and decision making

Language technology for all (potential) users



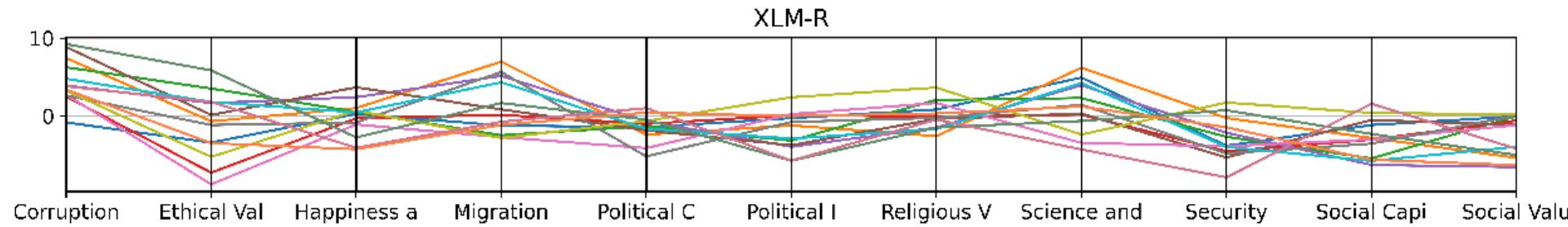
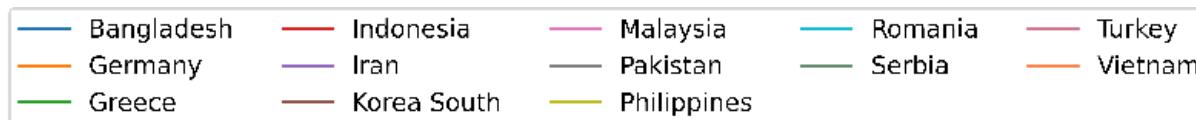
Local Languages, Third Spaces, and other High-Resource Scenarios
(Bird, ACL 2022)

Benefit to all requires finding the intersection,
particularly for local languages

Value bias in language models

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



Probing Pre-Trained Language Models for Cross-Cultural Differences in Values
 (Arora et al., 2022)

Value bias in language models



Die allermeisten von uns kennen den Zustand völliger Erschöpfung auf der Flucht, verbunden mit Angst um das eigene Leben oder das Leben der Kinder oder der Partner, zum Glück nicht. Menschen, die sich zum Beispiel aus Eritrea, aus Syrien oder dem Nordirak auf den Weg machen, müssen oft Situationen überwinden oder Ängste aushalten, die uns wahrscheinlich schlichtweg zusammenbrechen ließen. Deshalb müssen wir beim Umgang mit Menschen, die jetzt zu uns kommen, einige klare Grundsätze gelten lassen. Diese Grundsätze entstammen nicht mehr und nicht weniger als unserem Grundgesetz, unserer Verfassung.

Values are altered
to reflect US culture

(translation)

GPT-3 →
summarise

"1. I am in favor of
limiting immigration.
2. I am in favor of
limiting immigration for
humanitarian reasons.
3. I am in favor of
limiting immigration for
economic reasons."



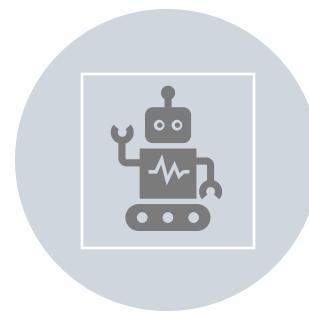
The Ghost in the Machine has an American accent: value conflict in GPT-3
(Johnson et al., 2022)

Strategies

Existing and potential ways to address the challenges



DATA



MODELS



TASKS

Data



Selection

Annotation

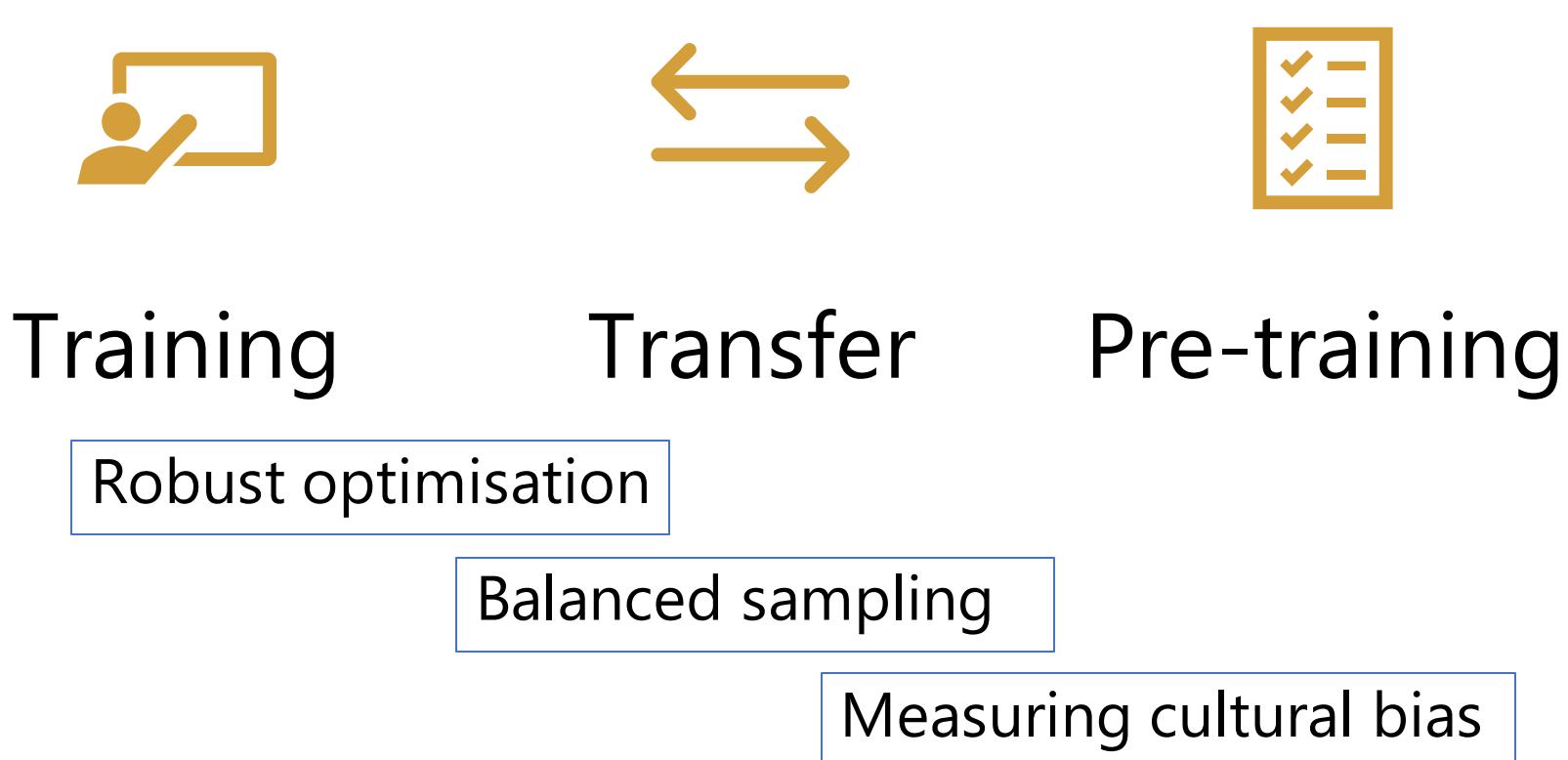
Projection

Culture-sensitive curation

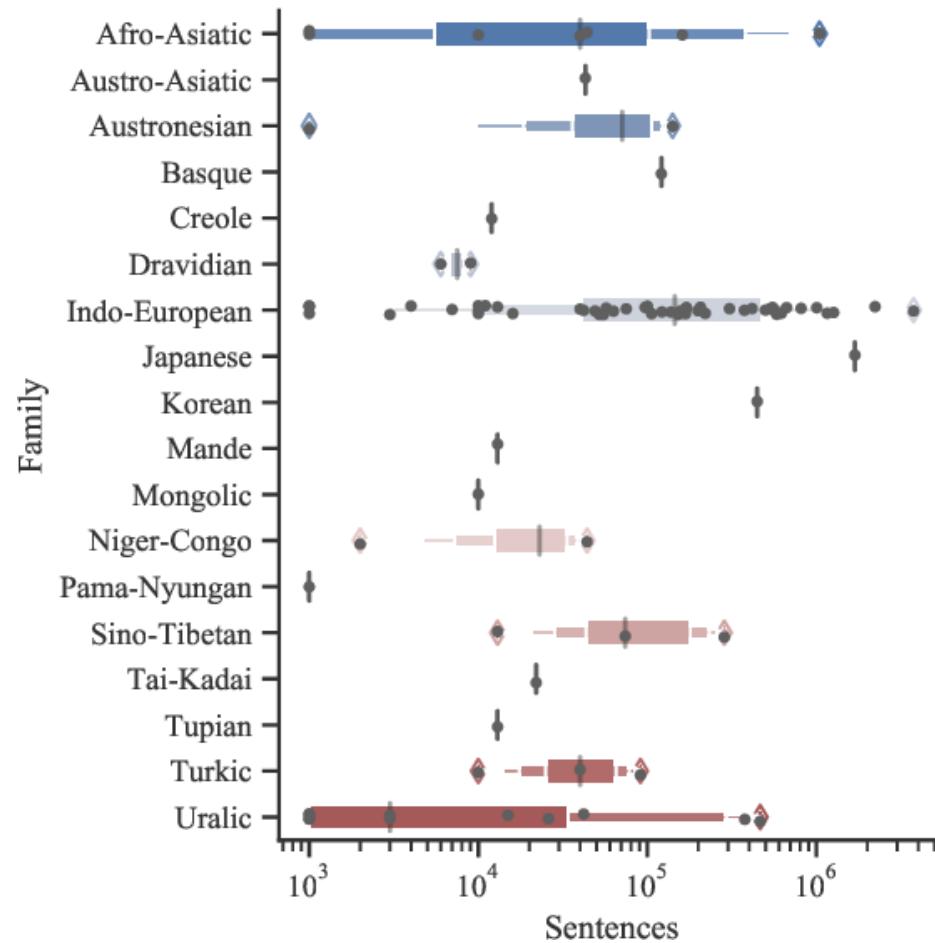
Culturally diverse collection

Native data or culturally sensitive translation

Models



Robust optimisation



Minimax and Neyman–Pearson Meta-Learning for Outlier Languages
(Ponti et al., Findings 2021)

Robust optimisation does not always work

Tamil	Mandarin(我们)	Cantonese(拍拖)	English	Malay	Eng	Malay	Hokkien/ Hakka(店)	X
Dey Hey	wǒ men ,	paktor we	always date	makan always	at eat	at kopitiam	coffee shop	<INTJ>

Standard English: “Hey, when we date we always eat at the coffee shop”

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

Tasks

Cross-cultural translation



"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

Multi-granularity adaptation

A Juror Selection

JUROR SHEET A
RACE Hispanic ✓
GENDER Female ✓
SEATS 4
Add characteristic

JUROR SHEET B
RACE Black ✓
AGE RANGE 25-34 ✓
SEATS 8
Add characteristic

+ Add juror sheet

Your jury composition: A₁, A₂, B₁, B₂, B₃, B₄, A₃, A₄, B₅, B₆, B₇, B₈

Your input example: This is an example comment entry.
View jury outcome →

B Jury Learning Results

Outcome summary: JURY VERDICT Slightly toxic (1.21 / 4.00)
95% of juries are between 0.21 - 1.83
Based on the median outcome of 100 juries sampled from your provided jury composition

DISTRIBUTION OF JURY OUTCOMES: Select a jury to view Jury Trends

Jury outcome (Yellow dots), Selected jury outcome (Red dot)

C Jury Trends (Jury 43)

GROUP BY Juror sheet ▾

Juror sheet A — 4 of 12 jurors (Race: Hispanic, Gender: Female)
Average label: Slightly to Moderately toxic (1.65 / 4.00)
Label distribution: Histogram showing predicted labels for Juror sheet A jurors.

Juror sheet B — 8 of 12 jurors (Race: Black, Age Range: 25-34)
Average label: Not at all toxic (0.43 / 4.00)
Label distribution: Histogram showing predicted labels for Juror sheet B jurors.

D Juror Details

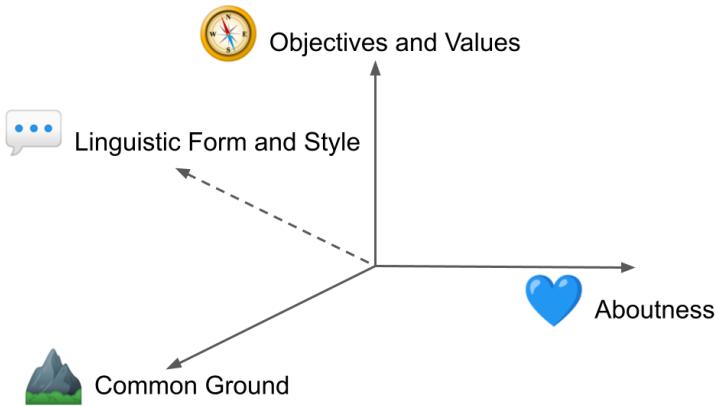
Jury 43, Juror B₅
Predicted label: Slightly toxic (1.12 / 4.00)
Juror background: RACE Black, GENDER Female, POLITICAL AFFIL. Independent, AGE RANGE 25-34
Comment: this is an example comment (2.3), this is another example comment (2.1), this is yet another example comment (3.4)

E Counterfactual juries

New jury composition	Jury verdict	Jury edits
	(0.87 / 4.00)	C ₁ — Race: White, Political Affiliation: Conservative
	(0.79 / 4.00)	D ₁ — Race: Black, Importance of religion: Not important
	(0.63 / 4.00)	E ₁ , E ₂ — Age range: 45-54, Importance of religion: Very important

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

Summary



NLP is for people (not just languages)

Culture is multidimensional

Objectives may be in conflict

Generalisation-representation trade-off

danielhers.github.io

dh@di.ku.dk

 @daniel_hers