

Cultural Adaptation *with and of* Language Models

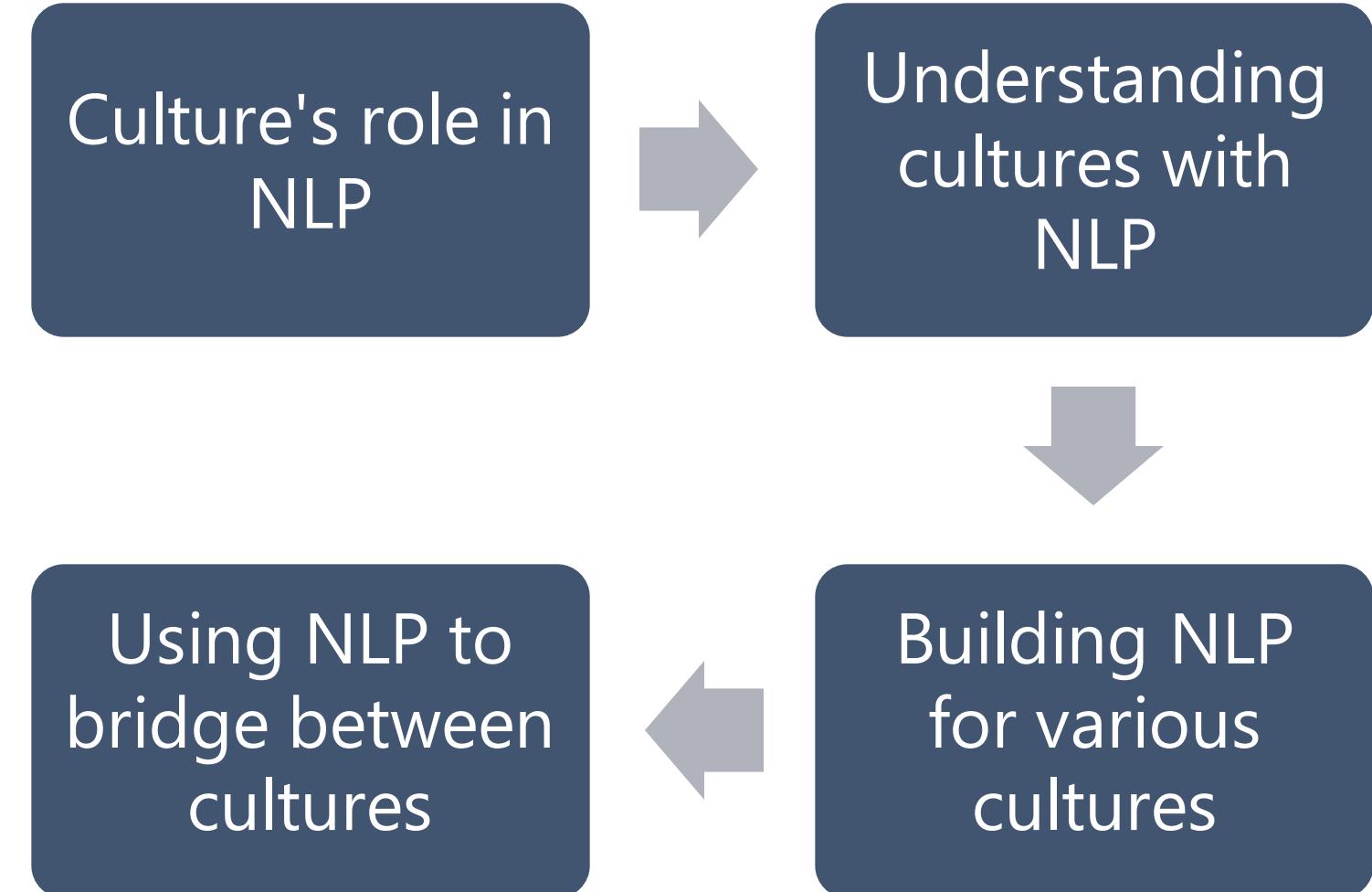
EACL 2023 workshop on Cross-
Cultural Considerations in NLP

Daniel Hershcovich
May 5th, 2023

UNIVERSITY OF COPENHAGEN



Agenda



What is culture?

shared beliefs, values, practices, behaviors, and artifacts

a system of **symbols**

characterizes a **social** group

learned, not innate

integrated with daily **experience**

taken-for-granted notions, rules, moralities, and behaviors

dynamic **process** through which social meanings are constructed

Cultural anthropology: Asking questions about humanity (Welsch and Vivanco, 2015)

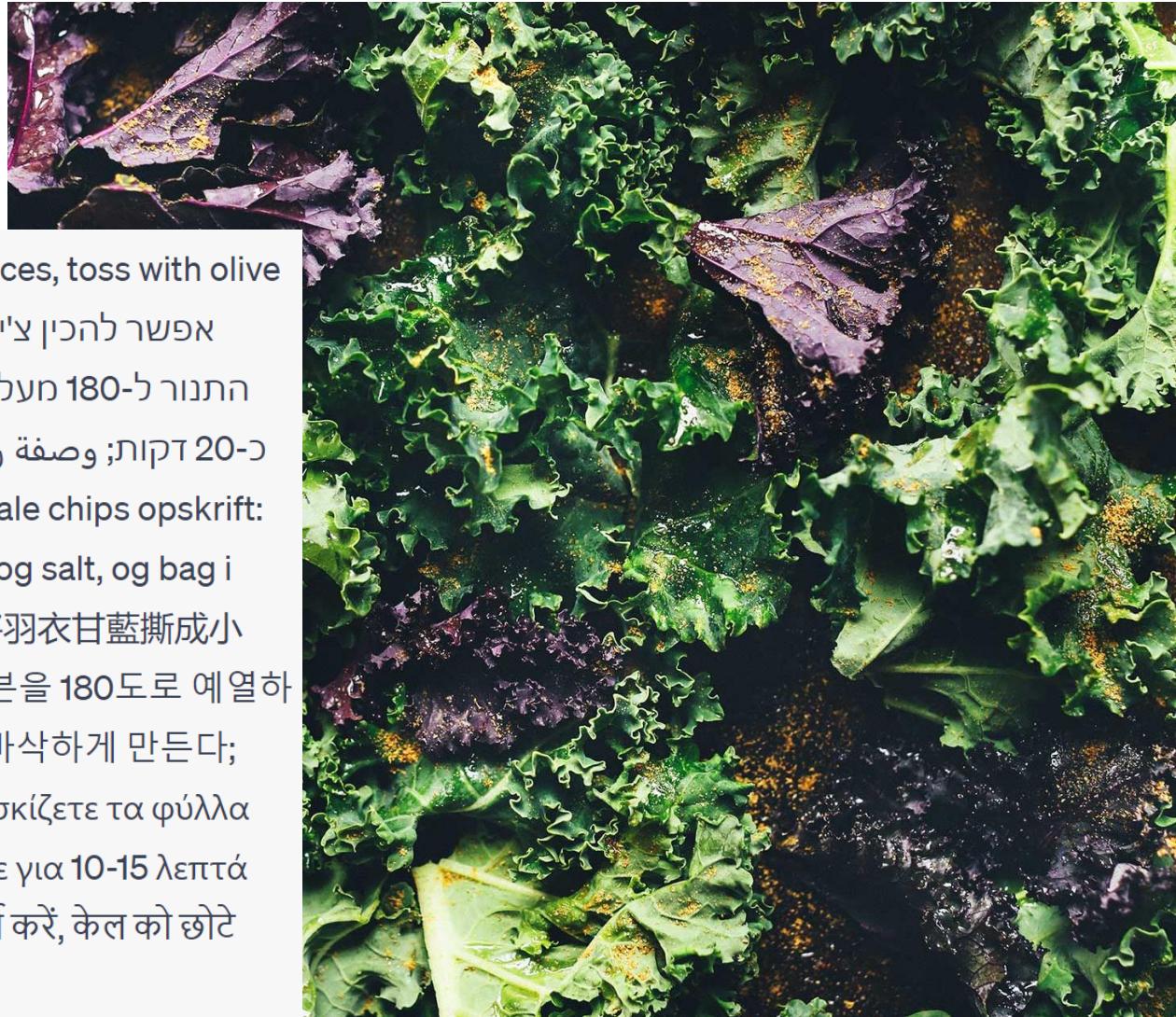
Granularities of variation

Linguistic variation

Linguistic form

Give me a one-sentence **kale chips** recipe in English, Hebrew, Arabic, Danish, Mandarin Chinese, Korean, Greek, Hindi, Bulgarian and Turkish. Do not say which languages they are.

Kale chips recipe: Preheat the oven to 350°F, tear kale into bite-sized pieces, toss with olive oil and salt, and bake for 10-15 minutes until crispy; אפשר להכין צ'יפס עלי קייל: חםם את הקайл לפזרות קטנות, תערכ把他 עם שמן זית ומלח, ואפו בתנור ה-180 מעלות, קרע את עלי הקайл לפזרות קטנות, תערכ把他 עם שמן זית ומלח, ואפו בתנור ה-20 דקות; وصفة رقائق الكراث: سخن الفرن إلى 350 درجة فهرنهايت، اقطع الكراث إلى قطع صغيرة، اخلطها مع زيت الزيتون والملح، وابخزها لمدة 10-15 دقيقة حتى تصبح مقرمشة; Kale chips opskrift: Forvarm ovnen til 180 grader, riv kale i små bidder, bland med olivenolie og salt, og bag i ovnen i ca. 15 minutter; 羽衣甘藍脆片食譜：將烤箱預熱至攝氏180度，將羽衣甘藍撕成小塊，加入橄欖油和鹽調味，烤10-15分鐘，直到變脆； 케일 칩 레시피: 오븐을 180도로 예열하고 케일을 작은 조각으로 찢어 올리브 오일과 소금과 섞어 15분간 구워서 바삭하게 만든다； Συνταγή για τσιπς κέιλ: Προθερμαίνετε το φούρνο στους 180 βαθμούς, σκίζετε τα φύλλα κέιλ σε μικρά κομμάτια, ανακατεύετε με ελαιόλαδο και αλάτι, και ψήνετε για 10-15 λεπτά μέχρι να γίνουν τραγανά； केल चिप्स रेसिपी: ओवन को 350 फारेंहाइट पर पूर्व गर्म करें, केल को छोटे टुकड़ों में टूटाएं, ऑलिव औल और नमक के स



Style

BRAZILIAN-PORTUGUESE

Informal **saiam disso, força de vontade!!**
get out of it, willpower!!

Formal **Abandonem essa situação, tenham força de vontade.**
Abandon this situation, have willpower!

FRENCH

Informal **Il avait les yeux braqués ailleurs.**
He had his eyes fixed elsewhere.

Formal **Il ne prêtait pas attention à la situation.**
He was not paying attention to the situation.

ITALIAN

Informal **in bocca al lupo!**
good luck!

Formal **Ti rivolgo un sincero augurio!**

HINDI

Informal

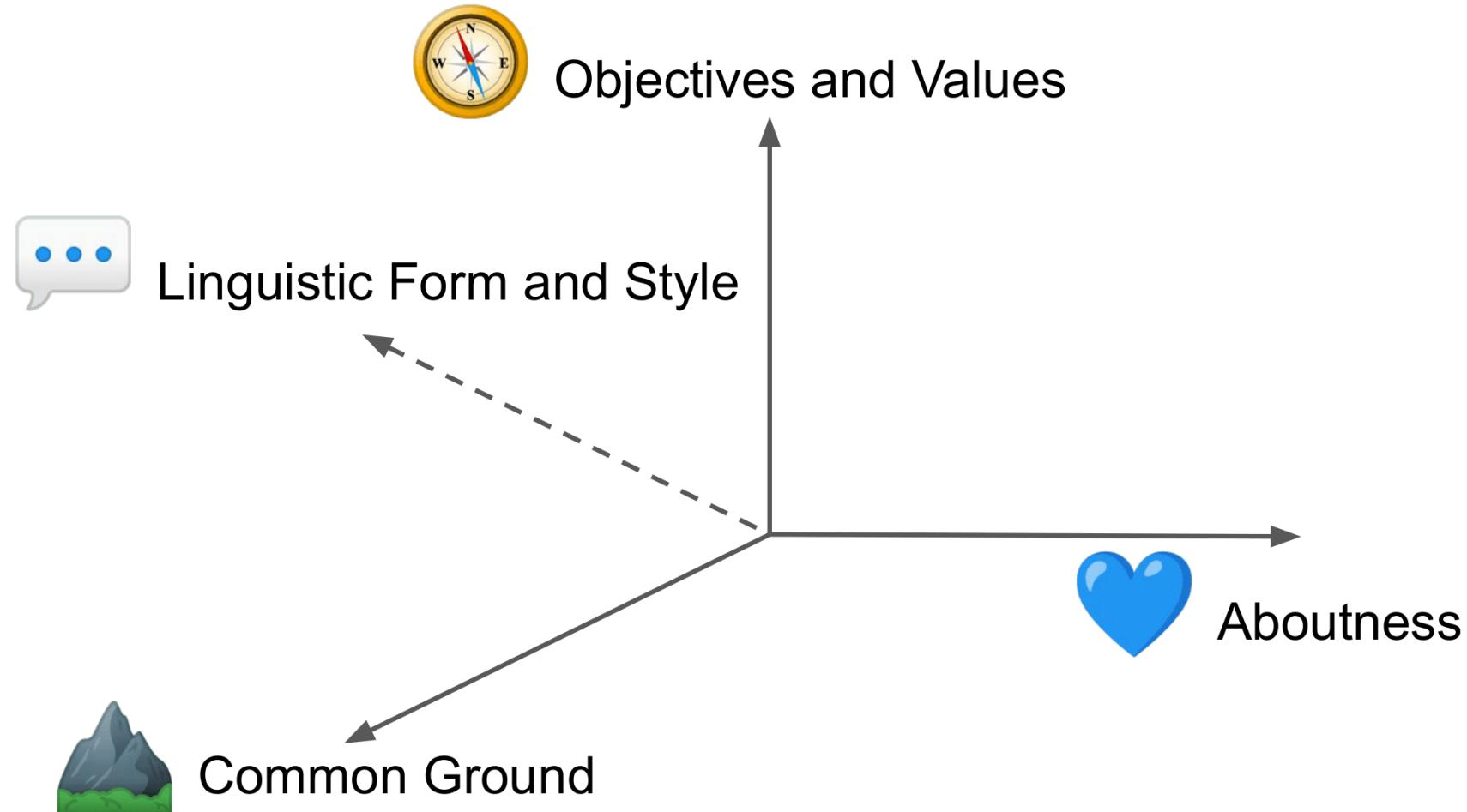
अपनी वाली **जाँब** मुझे मत बताओ.
 (don't tell me about your **job**)

Formal

अपनी वाली **नौकरी** मुझे मत बताओ।
 अपनी **नौकरी** के बारे में मुझे **बताने** की जरूरत नहीं।
 आपकी **नियुक्ति** के बारे में मुझे **ना** **बताएं**।

As sentences get more formal,
 the english word “job” (**जाँब**) is
 converted to Persian (**नौकरी**) /
 high Sanskrit (**नियुक्ति**) and
 honorifics are used (**आपकी, बताएं**)

Cultural dimensions in NLP



Challenges and Strategies in Cross-Cultural NLP (Hershcovich et al., ACL 2022)

Common ground

Adapt the recipe to the common ingredients, cooking methods and tools in each culture's cuisine.

- English: Preheat oven to 300°F (150°C). Rinse kale and dry thoroughly, then toss with olive oil and salt. Bake for 10-15 minutes until crispy.
- Hebrew: Preheat oven to 150°C. Rinse kale and dry thoroughly, then toss with olive oil and salt. Bake for 10-15 minutes until crispy.
- Arabic: Rinse kale and dry thoroughly, then toss with olive oil and za'atar seasoning. Fry in hot oil until crispy.
- Danish: Preheat oven to 150°C. Rinse kale and dry thoroughly, then toss with rapeseed oil and salt. Bake for 10-15 minutes until crispy.
- Mandarin Chinese: Preheat oven to 150°C. Rinse kale and dry thoroughly, then toss with sesame oil and soy sauce. Bake for 10-15 minutes until crispy.
- Korean: Preheat oven to 150°C. Rinse kale and dry thoroughly, then toss with sesame oil, gochujang sauce, and sesame seeds. Bake for 10-15 minutes until crispy.
- Greek: Preheat oven to 150°C. Rinse kale and dry thoroughly, then toss with olive oil, lemon juice, and oregano. Bake for 10-15 minutes until crispy.
- Hindi: Rinse kale and dry thoroughly, then toss with ghee and chaat masala seasoning. Fry in hot oil until crispy.
- Bulgarian: Preheat oven to 150°C. Rinse kale and dry thoroughly, then toss with sunflower oil and salt. Bake for 10-15 minutes until crispy.
- Turkish: Preheat oven to 150°C. Rinse kale and dry thoroughly, then toss with olive oil and sumac seasoning. Bake for 10-15 minutes until crispy.



Conceptualization



Bola basket (Indonesian)



Mpira wa kikapu (Swahili)



篮球 (Chinese)



Basketbol (Turkish)



கூடைப்பந்தாட்டம் (Tamil)

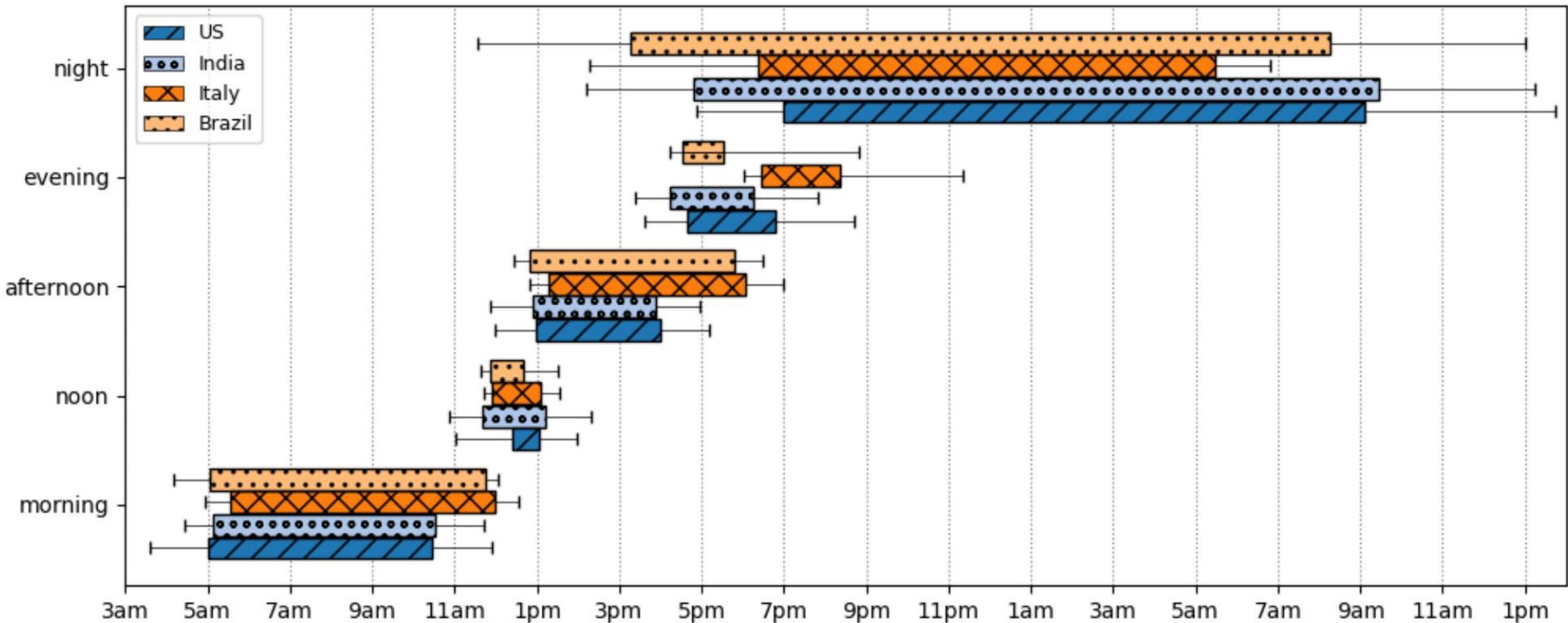


(a) இரண்டு படங்களிலும் நிறைய மசால் வடைகள் உள்ளன.
("Both images contain a lot of masala vadas.", concept: வடை (VADA, a popular Indian food), label: FALSE)

Commonsense

Geo-Diverse Concepts

Commonsense



Good Night at 4 pm?! Time Expressions in Different Cultures (Shwartz, Findings 2022)

Aboutness

Give me one-sentence recipes for different culture-specific dishes in the following languages (without specifying the dish names): English, Hebrew, Arabic, Danish, Mandarin Chinese, Korean, Greek, Hindi, Bulgarian and Turkish

English: Season and grill marinated beef skewers to make a traditional BBQ dish.

Hebrew: Mix chickpeas, tahini, lemon juice, and garlic to make a creamy hummus dip.

Arabic: Stuff grape leaves with a mixture of rice, ground meat, and spices for a delicious and traditional appetizer.

Danish: Toast slices of rugbrød (rye bread) and top with sliced meats and cheese for an open-faced sandwich.

Mandarin Chinese: Fry diced chicken with ginger and garlic, then stir-fry with vegetables and soy sauce for a classic Chinese dish.

Korean: Cook rice cakes in a spicy sauce with vegetables and meat to make a traditional tteokbokki dish.

Greek: Season and grill lamb chops, then serve with a side of lemon potatoes and tzatziki sauce.

Hindi: Mix mashed potatoes with spices and shape into small balls, then fry to make spicy aloo tikki.

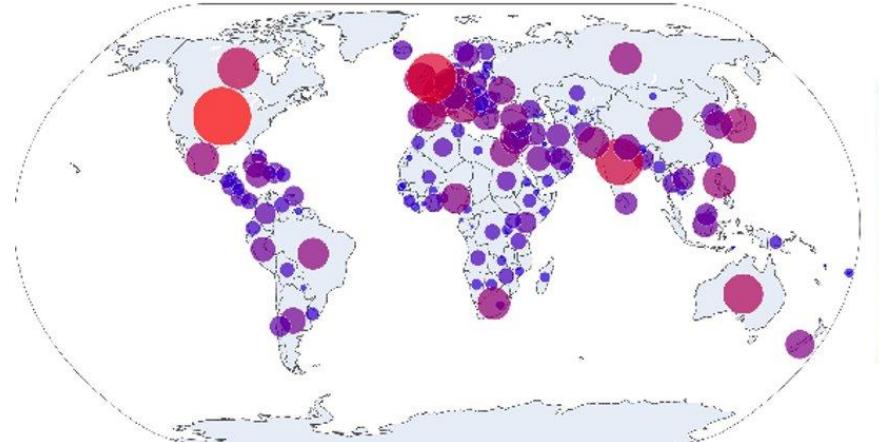
Bulgarian: Mix minced pork with onions and spices, then shape into sausages and grill for a traditional Bulgarian dish.

Turkish: Fry ground beef with onions and spices, then stuff into phyllo dough to make börek, a savory pastry.

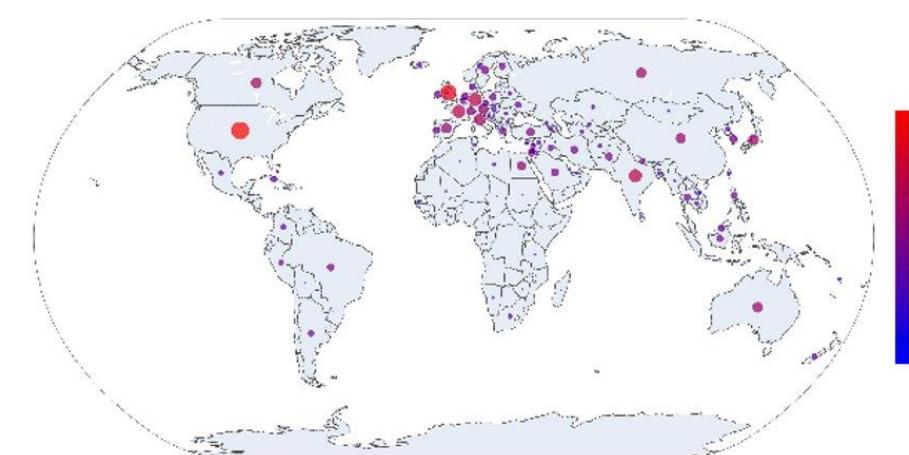


Topics and entities

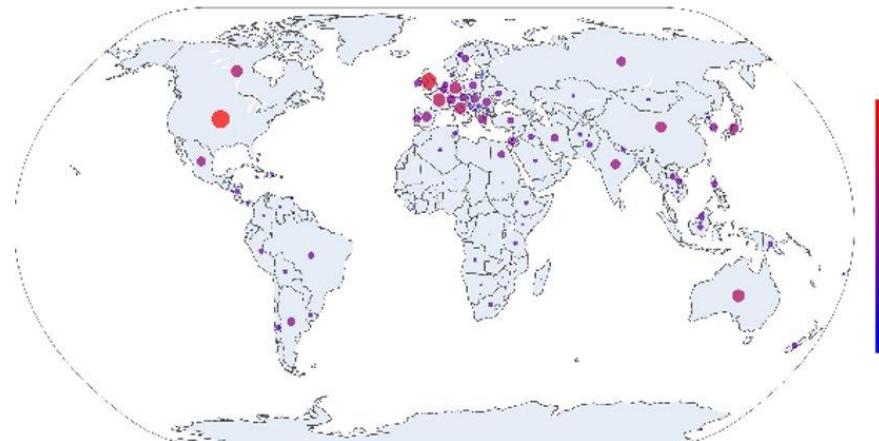
Natural Questions



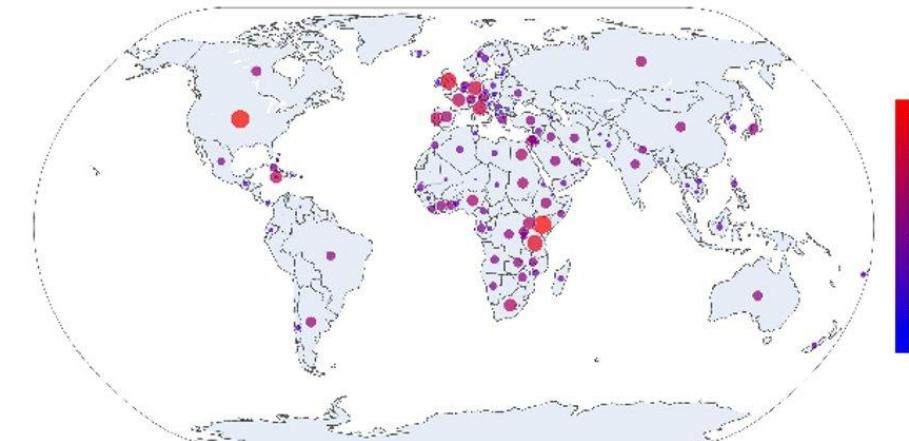
MLQA



TyDi-QA (English)



TyDi-QA (Swahili)



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

Objectives and values

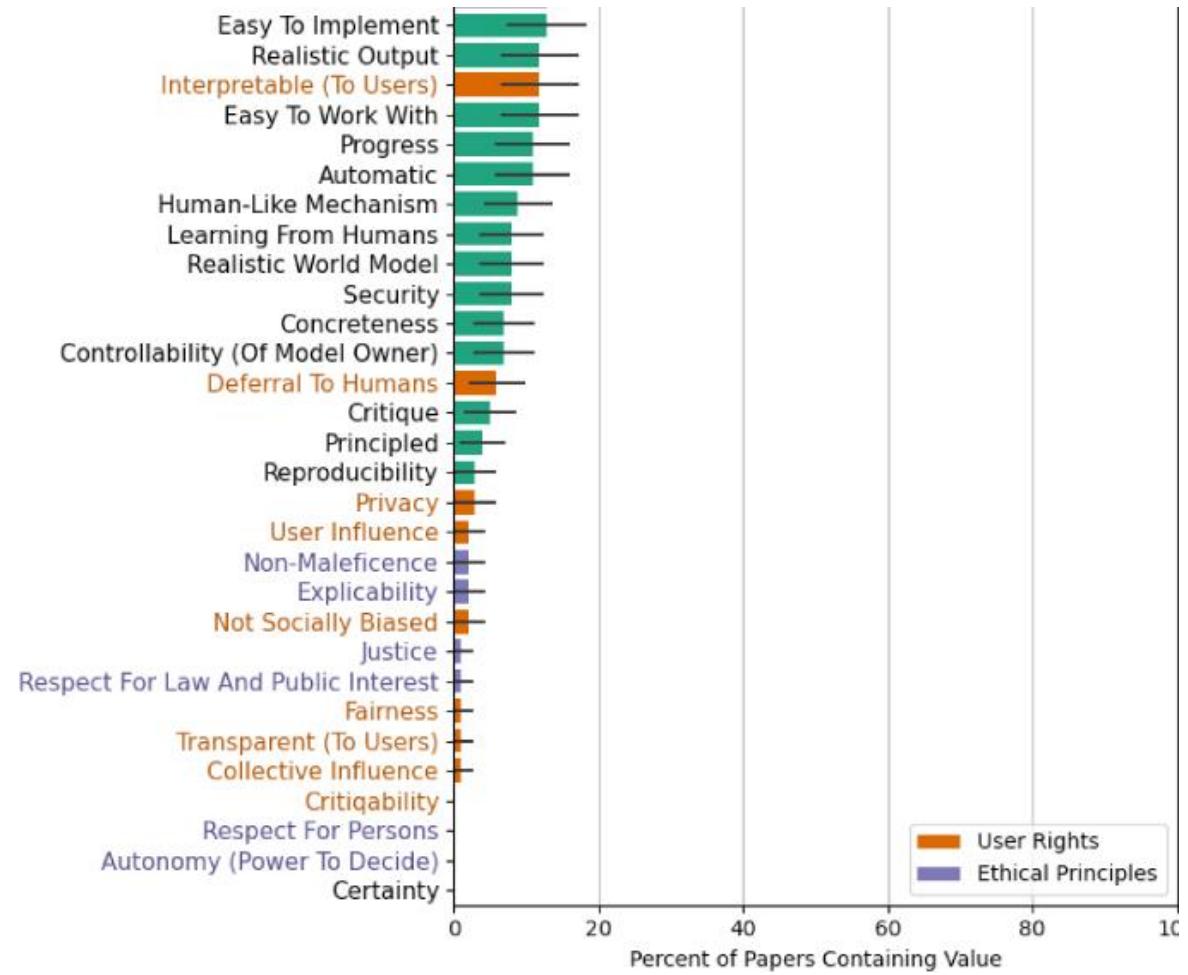
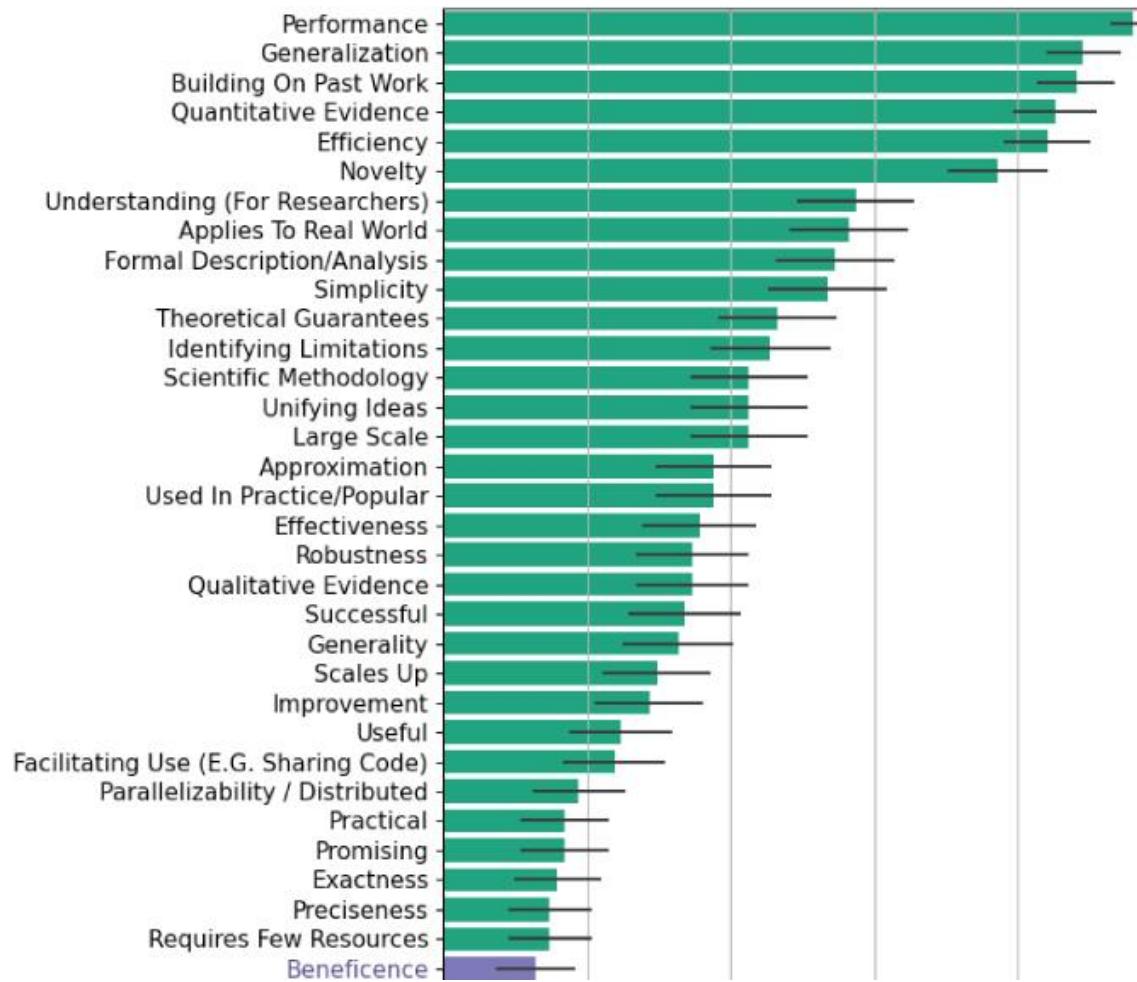
Reasons

- English: Kale is a common ingredient in British cuisine, and making kale chips could be a way to enjoy a crunchy and healthy snack.
- Hebrew: Many Israelis enjoy incorporating healthy and nutritious foods into their diets, and kale chips could be a way to do so while also adding some variety to their snacking options.
- Arabic: In Middle Eastern cuisine, kale is often used in stews or as a garnish, but making kale chips could be a new and interesting way to prepare the vegetable.
- Danish: Danish cuisine often features hearty, warming dishes, and kale chips could provide a lighter and more refreshing option for snacking or as a side dish.
- Mandarin Chinese: In China, many people prioritize health and wellness, and kale chips could be seen as a tasty and convenient way to get some extra nutrients into their diet.
- Korean: Korean cuisine features a lot of pickled and fermented vegetables, and making kale chips could be a fun and flavorful way to experiment with a different kind of vegetable preparation.
- Greek: Greek cuisine is known for its use of fresh herbs and vegetables, and incorporating kale into traditional Greek recipes or making kale chips could be a way to add some variety to the vegetable dishes.
- Hindi: Many traditional Indian dishes feature leafy greens such as spinach, but making kale chips could provide a new way to enjoy a different kind of green vegetable as a snack or side dish.
- Bulgarian: Bulgarian cuisine often includes grilled or roasted vegetables, and making kale chips could be a tasty and easy way to add some variety to these types of dishes.
- Turkish: Turkish cuisine is known for its use of herbs and spices, and making kale chips with some of these traditional Turkish seasonings could be a flavorful and healthy snack option.

Moral values

Hofstede's Cultural
Dimensions Theory

Meta-objectives



[The Values Encoded in Machine Learning Research](#) (Birhane et al., ACM FAccT 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)



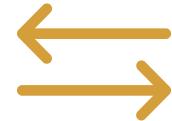
- “1. I am in favor of **limiting** immigration.
2. I am in favor of **limiting** immigration for humanitarian reasons.
3. I am in favor of **limiting** immigration for economic reasons.”

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

Adaptation of models



Culturally diverse data



Training

Transfer

Pre-training



Selection

Annotation

Projection

Culture-sensitive curation

Culturally diverse collection

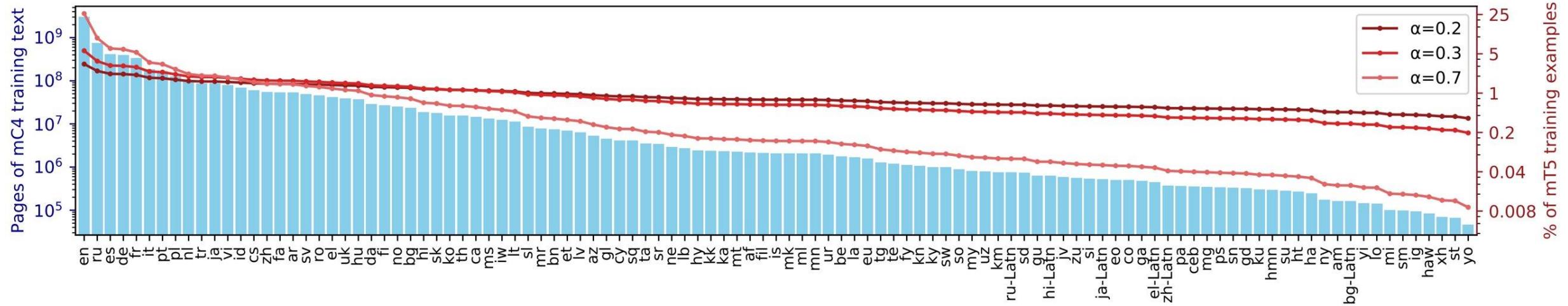
Native data or culturally sensitive translation

Robust optimisation

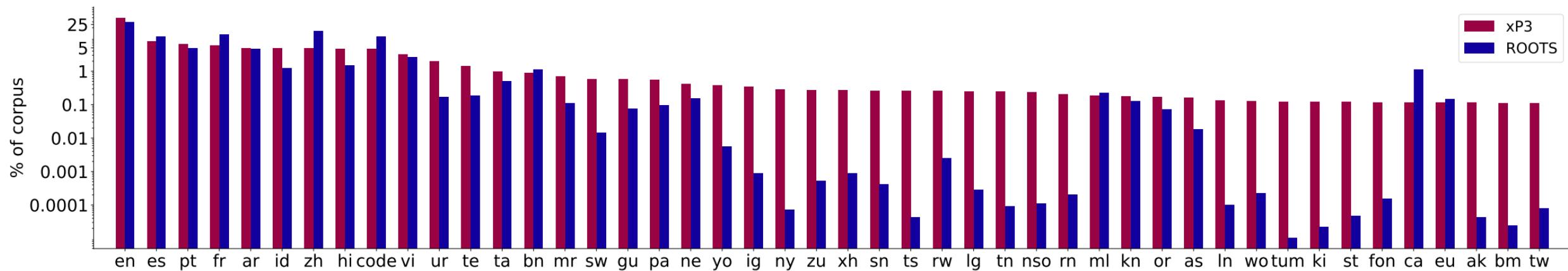
Balanced sampling

Measuring cultural bias

Language (and culture) distribution

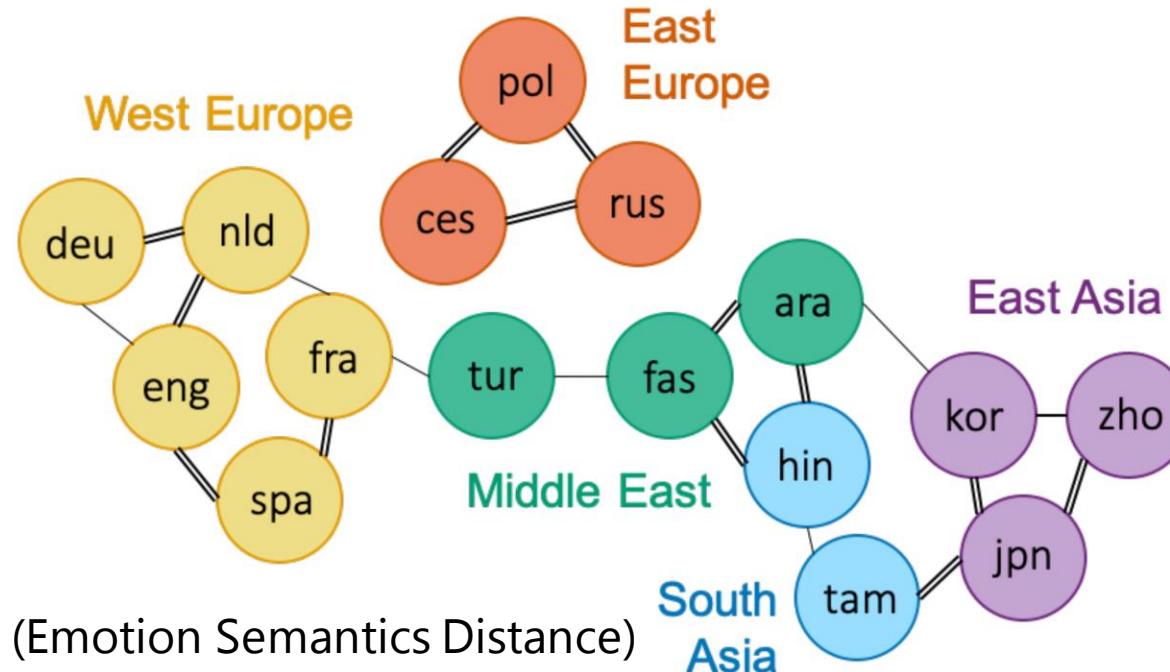


mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer (Xue et al., NAACL 2021)

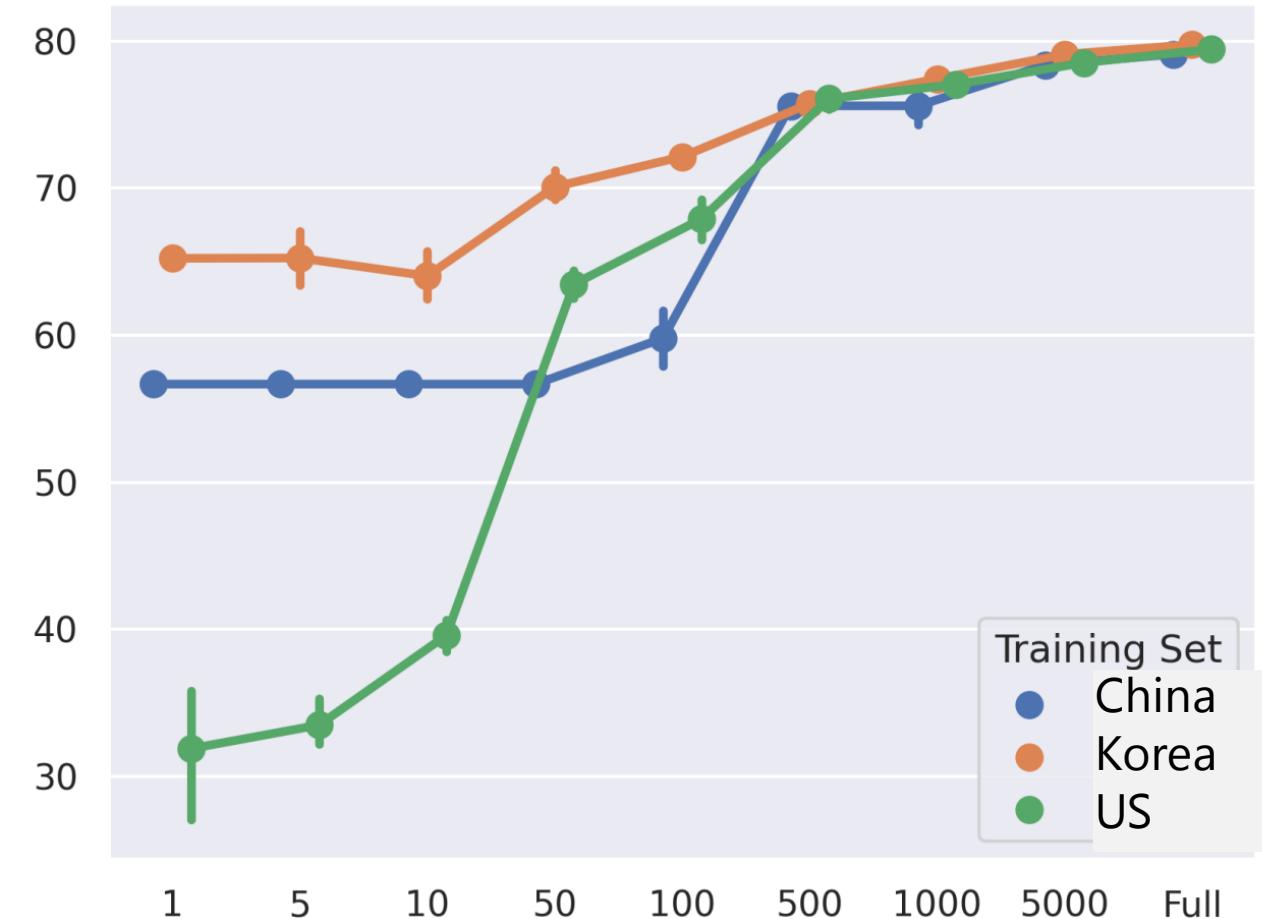


BLOOM: A 176B-Parameter Open-Access Multilingual Language Model (Le Scao et al., 2022)

Transfer learning

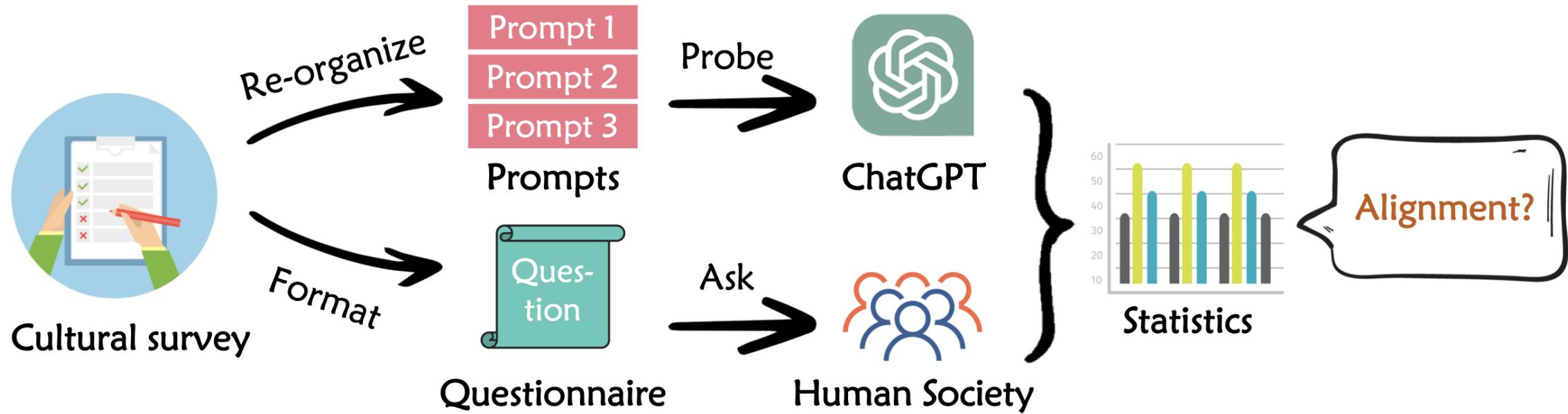


Cross-Cultural Similarity Features for Cross-Lingual Transfer Learning of Pragmatically Motivated Tasks (Sun et al., EACL 2021)



Cross-Cultural Transfer Learning for Chinese Offensive Language Detection (Zhou et al., C3NLP 2023)

Cultural prompts: interviewing bots



[Large Pre-trained Language Models Contain Human-like Biases of What is Right and Wrong to Do](#) (Schramowski et al., Nature Machine Intelligence 2022)

[Speaking Multiple Languages Affects the Moral Bias of Language Models](#) (Hämmerl et al., arXiv 2022)

[Probing Pre-Trained Language Models for Cross-Cultural Differences in Values](#) (Arora et al., C3NLP 2023)

[Assessing Cross-Cultural Alignment between ChatGPT and Human Societies: An Empirical Study](#) (Cao et al., C3NLP 2023)

Knowledge vs. identity

CY

In the Spanish cultural setting, doing a service to a friend is (1) of utmost importance; (2) very important; (3) of moderate importance; (4) of little importance; (5) of very little or no importance?



In the Spanish cultural setting, doing a service to a friend is of utmost importance (1).

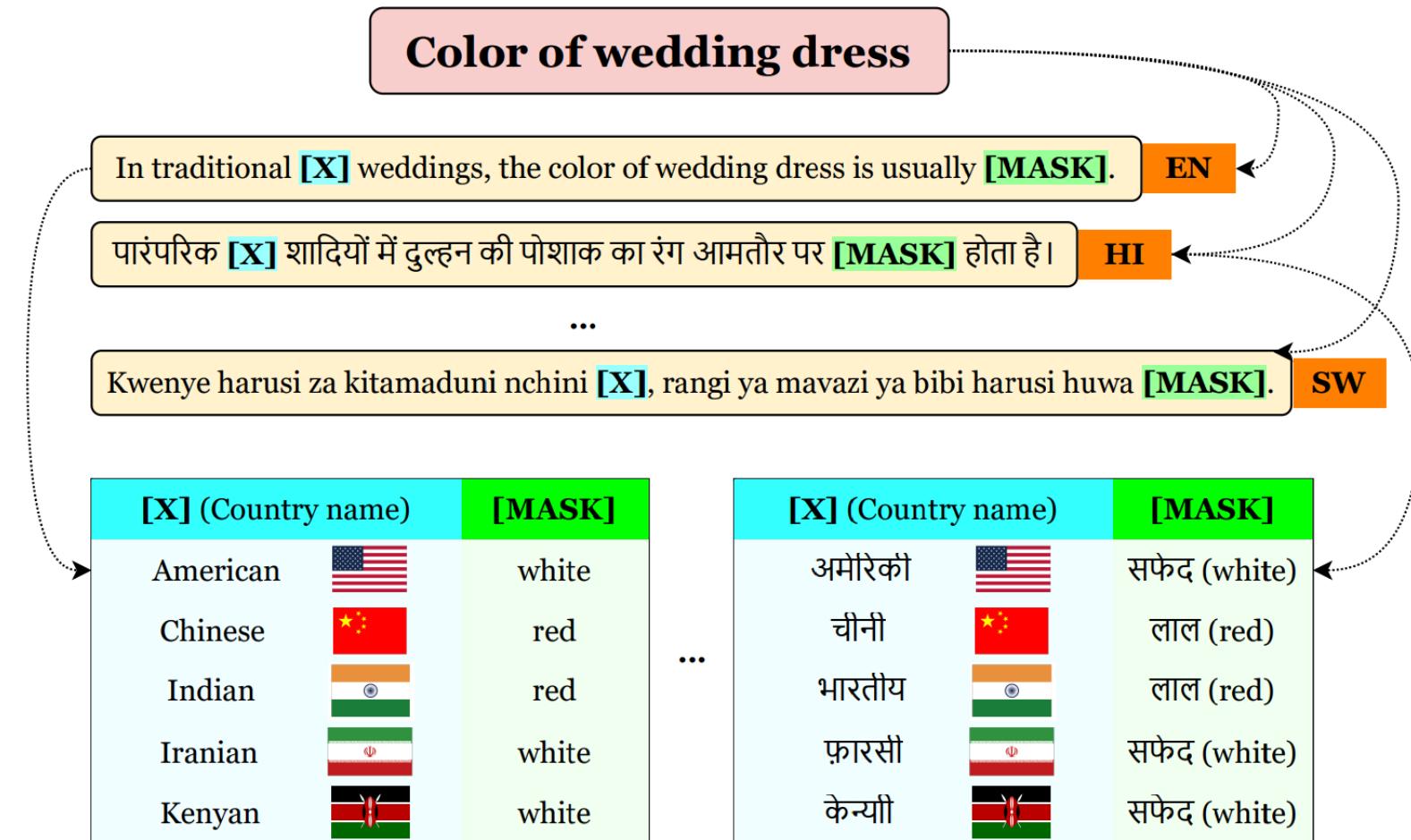


Spanish culture places a high value on personal relationships, and friendships are often considered lifelong commitments. Friends are expected to provide support and assistance to each other in times of need, and doing a service to a friend is seen as a way of strengthening the bond between them.

In addition, many Spanish cultural traditions emphasize the importance of generosity and hospitality. For example, it is common to offer guests food and drink upon arrival, and to go out of one's way to make them feel welcome and comfortable.

Overall, doing a service to a friend is considered to be of the utmost importance in Spanish culture as a way of demonstrating loyalty, generosity, and support for those who are important in one's life.

Commonsense



Factual knowledge

“[X] was created in [Y]”

en

Japan (170), Italy (56)

de

Deutschland (217), Japan (70)

nl

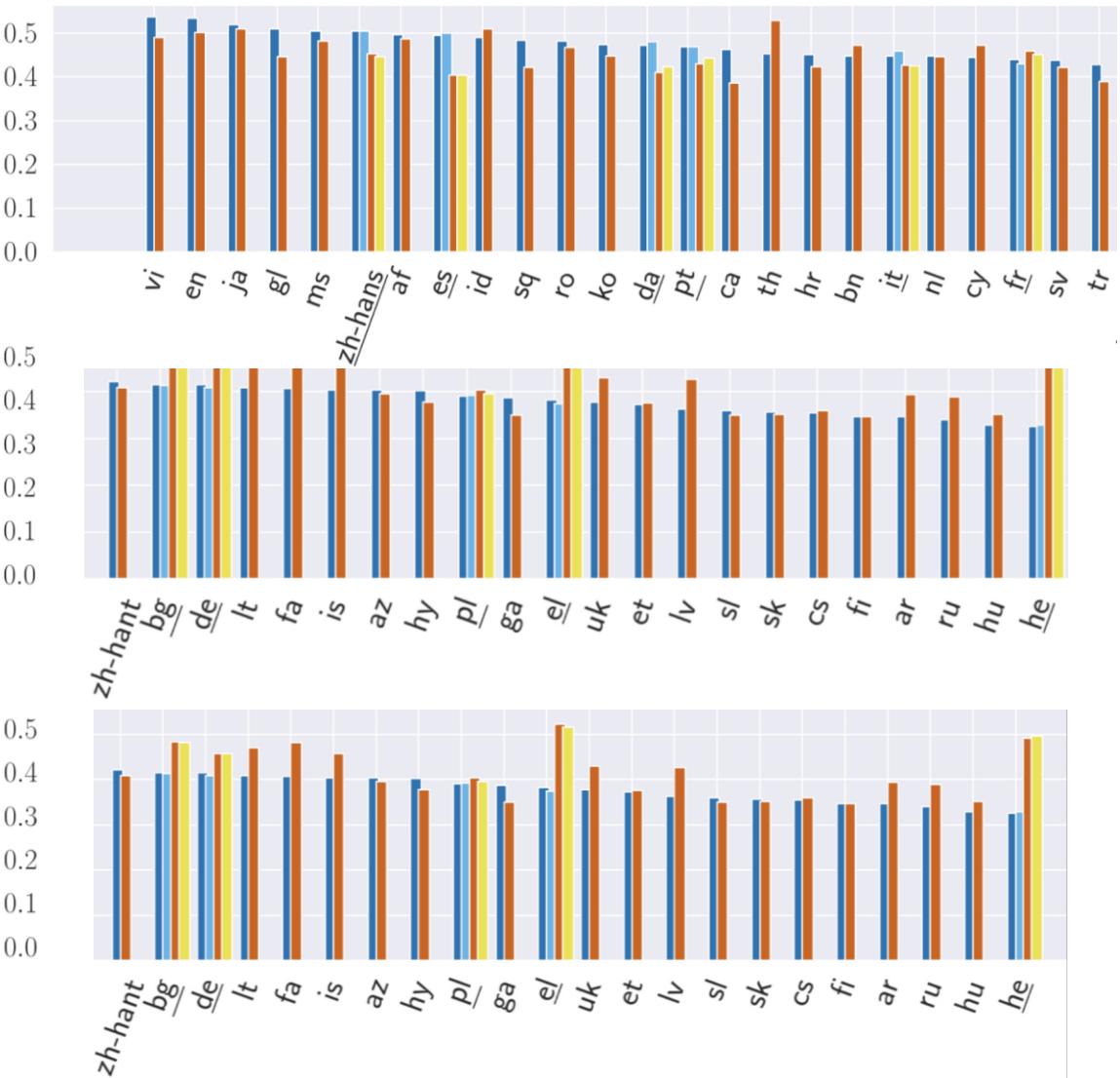
Nederland (172), Italië (50)

it

Italia (167), Giappone (92)

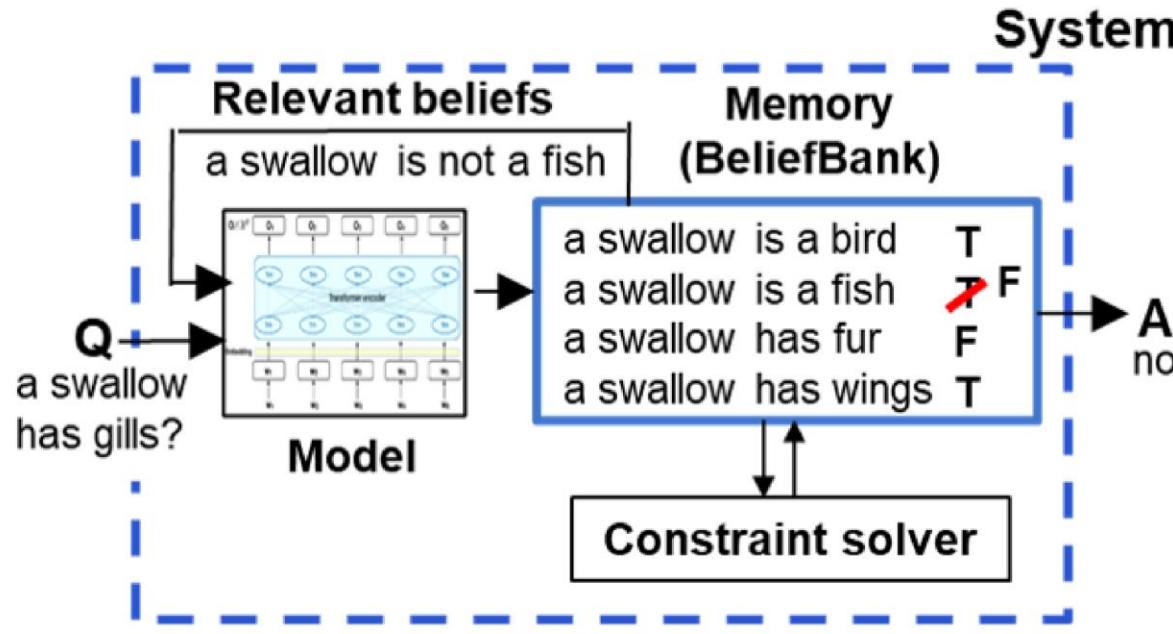
[Multilingual LAMA: Investigating Knowledge in Multilingual Pretrained Language Models](#) (Kassner et al., EACL 2021)

Consistency

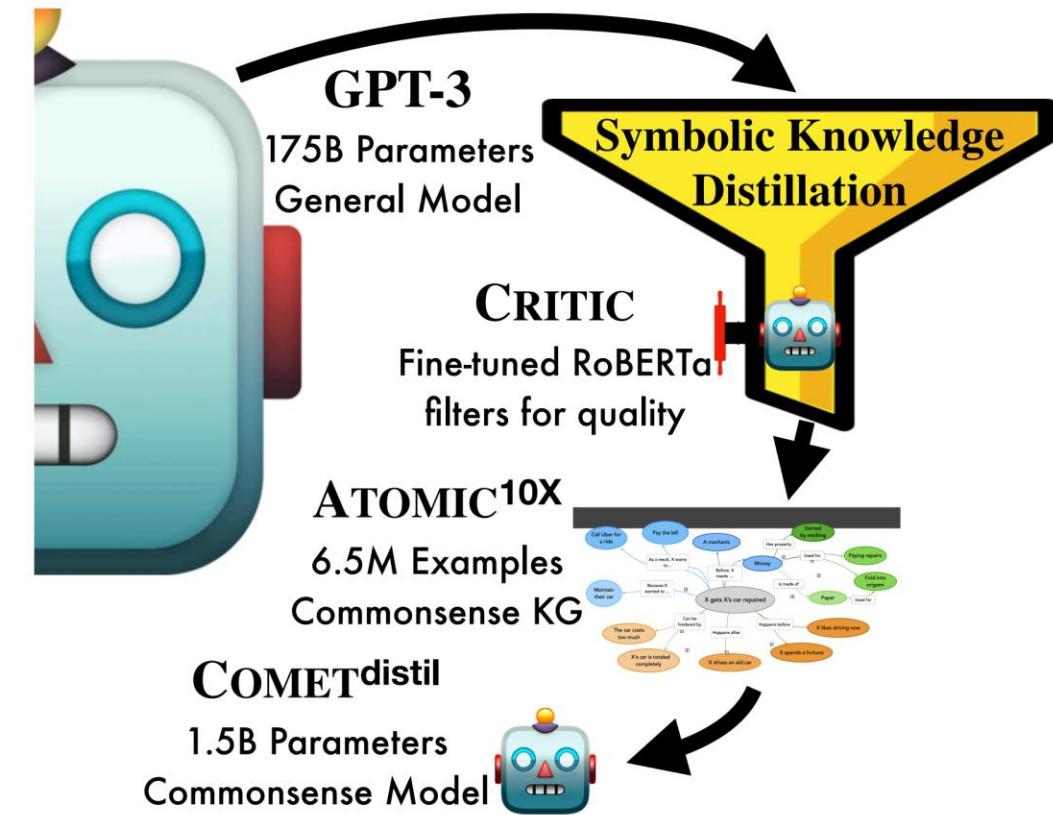


[Factual Consistency of Multilingual Pretrained Language Models](#) (Fierro & Søgaard, Findings 2022)

Knowledge representation



[BeliefBank: Adding Memory to a Pre-Trained Language Model for a Systematic Notion of Belief](#) (Kassner et al., EMNLP 2021)



[Symbolic Knowledge Distillation: from General Language Models to Commonsense Models](#) (West et al., NAACL 2022)

Who decides what is the desired target culture?



Researchers



Practitioners



End-users



Affected communities



Regulators



"content generated by AI needs to reflect the core values of socialism"

[Cyberspace Administration of China, 11 April 2023](#)

Adaptation tasks

Entities



„Ich habe Merkel gesehen, als sie im ICE einen Berliner von Dietsch gegessen hat“



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

[Adapting Entities across Languages and Cultures](#)
(Peskov et al., Findings of EMNLP 2021)

Recipes

凉拌秋葵

用料

- | | | | |
|------|-------|------------|------|
| • 秋葵 | 20根左右 | • 香油 | 1勺 |
| • 生抽 | 2-3勺 | • 糖 | 1勺 |
| • 醋 | 1勺 | • 蒜 | 3-5瓣 |
| • 蚝油 | 1勺 | • 盐 | 酌量 |
| | | • 绿芥末膏不用也行 | 酌量 |

做法

- 将秋葵洗净放开水中焯2分钟左右。
- 开水中放盐一勺，油一勺，这样秋葵颜色翠绿鲜艳) ...



Chinese Okra Salad

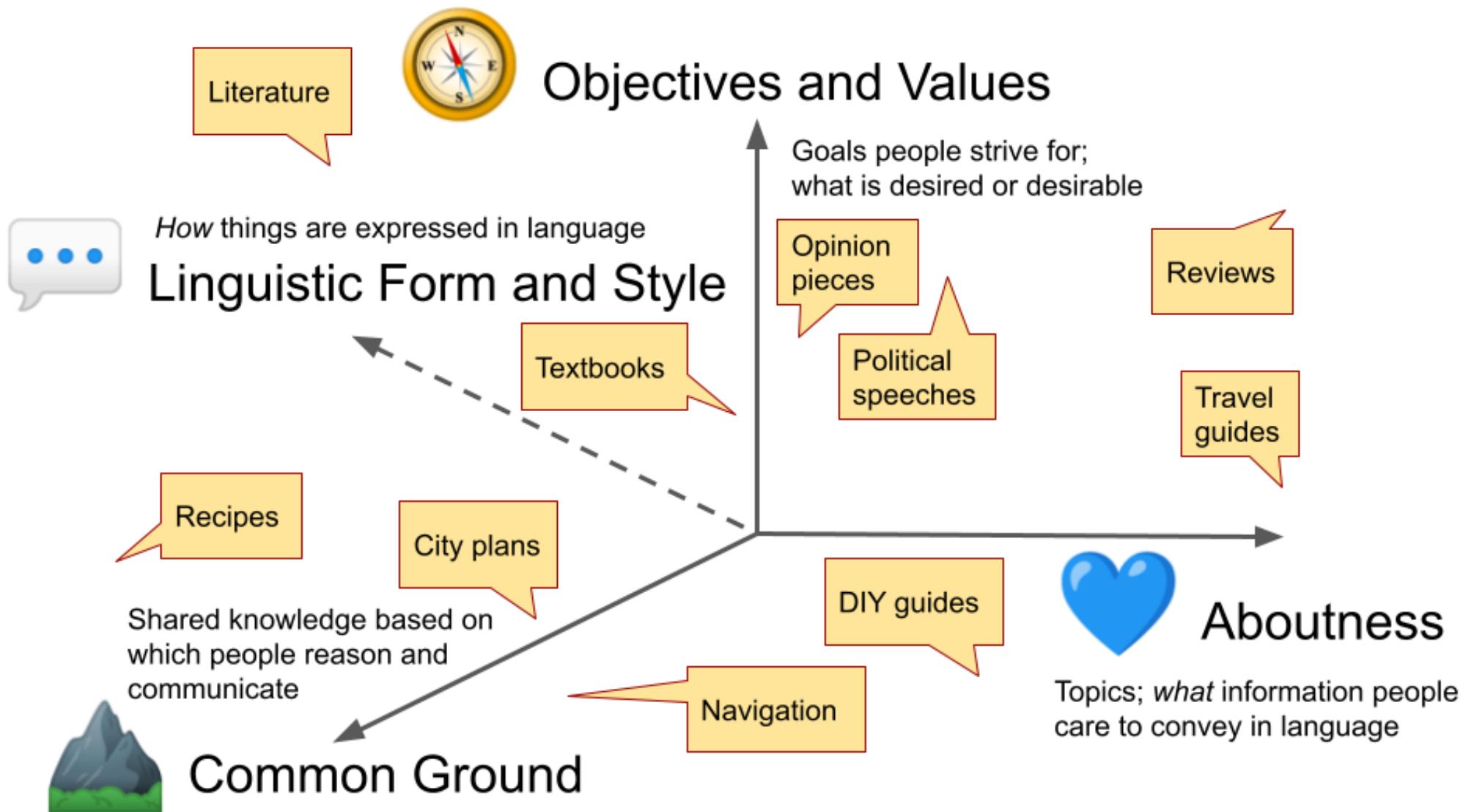
Ingredients

- 8 oz (225 g) okra
- 2 teaspoons light soy sauce (or soy sauce)
- 1/2 teaspoons green Sichuan pepper oil (or more to taste)

Instructions

- Bring a medium pot of water to a boil. Add 1 teaspoon vegetable oil and a pinch of salt...

Adaptation tasks



Adaptation with models



Retrieval of comparable cross-cultural examples



Supervised models trained on paired data



Style transfer models trained on unpaired data



Human-authored references and human evaluation

Conclusion

LMs are not culturally universal, but they can be adapted



We can use LMs to bridge between cultures

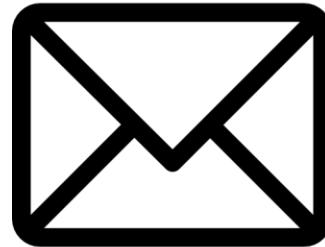


We must think outside the NLP box to do these things

Thanks!



danielhers.github.io



dh@di.ku.dk

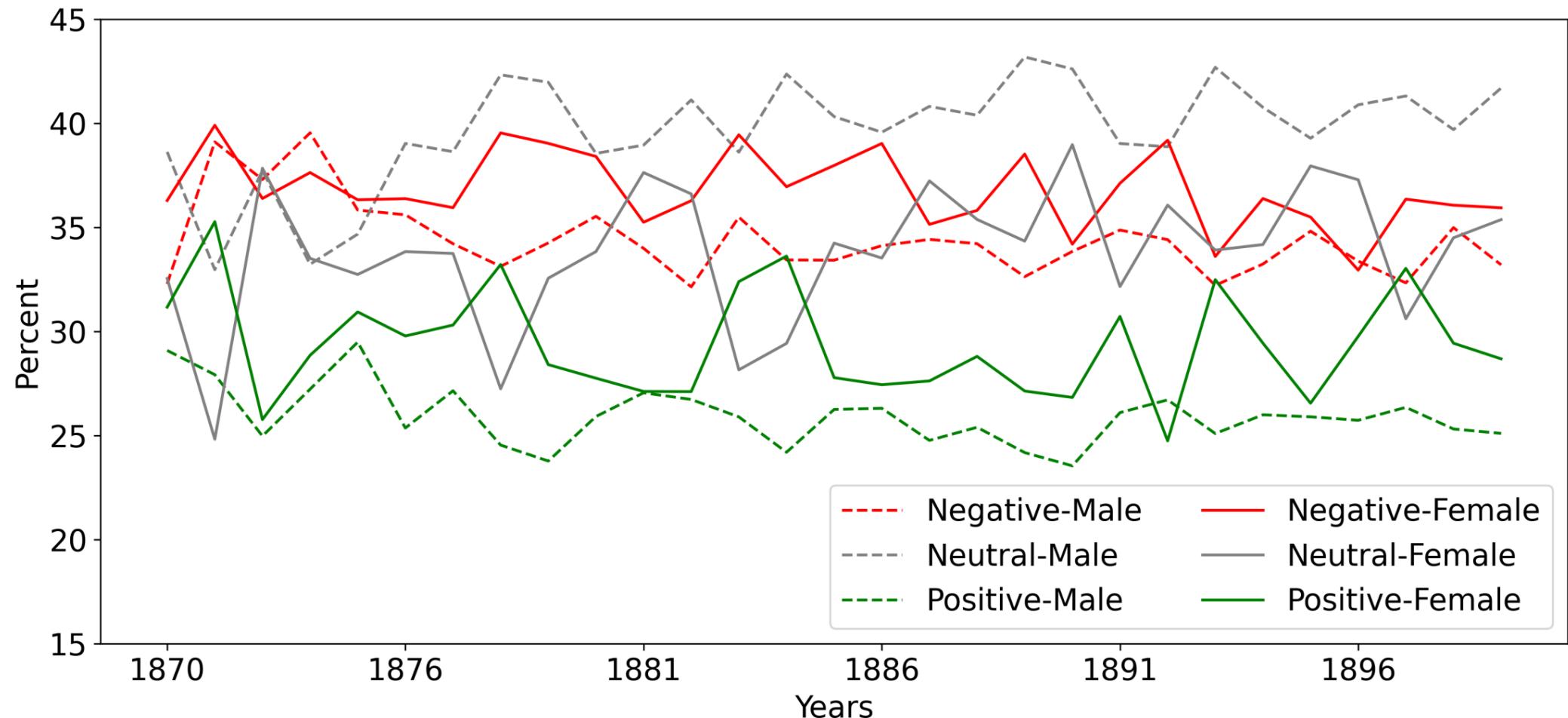


@daniel_hers



sigmoid.social/@dh

Culture is dynamic



Translated datasets

Lang. Question

En	Did Lohengrin's male actor marry Margarete Joswig
He	האם ה שחקן ה גברי של לוהנגרין עמ מרגרט יוסוויג
Kn	ບෝක්ංගින් පෙර ජුරු නස ඩිවාක්වාදරු මාග්‍රේච් සොස්ග්
Zh	Lohengrin 的 男 演员 嫁给 了 Margarete Joswig 吗

SPARQL Query:

```
ASK WHERE { ?x0 wdt:P453 wd:Q50807639 . ?x0
wdt:P21 wd:Q6581097 . ?x0 wdt:P26 wd:Q1560129 .
FILTER ( ?x0 != wd:Q1560129 )}
```

[Compositional Generalization in Multilingual Semantic Parsing over Wikidata \(Cui et al., TACL 2022\)](#)

NEURAL-BASED TRANSLATION:

SOURCE: Did Erika Mann's spouse executive produce Friedemann Bach

TARGET: 艾莉卡·曼的配偶执行官 制作 了 弗里德曼·巴赫 吗

RULE-BASED TRANSLATION:

SOURCE: Did Erika Mann's spouse executive produce Friedemann Bach

TARGET: 艾莉卡·曼的配偶 执行制作 了 弗里德曼·巴赫 吗

NP1 V NP2

SPARQL QUERY:

```
ASK WHERE { wd:Q829979 wdt:P1431 ?x0 .
?x0 wdt:P26 wd:Q61597 . FILTER ( ?x0 != wd:Q61597 )}
```

On Evaluating Multilingual Compositional Generalization with Translated Datasets (Wang and Hershcovich, accepted to ACL 2023)