

Challenges and Strategies in Cross-Cultural NLP

Daniel Hershcovich¹ Stella Frank² Heather Lent¹ Miryam de Lhoneux^{1,3,4}
Mostafa Abdou¹ Stephanie Brandl¹ Emanuele Bugliarello¹ Laura Cabello Piqueras¹
Ilias Chalkidis¹ Ruixiang Cui¹ Constanza Fierro¹ Katerina Margatina⁵
Phillip Rust¹ Anders Søgaard¹

¹University of Copenhagen ²University of Trento ³Uppsala University

⁴KU Leuven ⁵University of Sheffield

dh@di.ku.dk

Abstract

Various efforts in the Natural Language Processing (NLP) community have been made to accommodate linguistic diversity and serve speakers of many different languages. However, it is important to acknowledge that speakers and the content they produce and require, vary not just by language, but also by culture. Although language and culture are tightly linked, there are important differences. Analogous to cross-lingual and multilingual NLP, cross-cultural and multicultural NLP considers these differences in order to better serve users of NLP systems. We propose a principled framework to frame these efforts, and survey existing and potential strategies.

1 A Framework for Cultural Awareness

Language technology is rapidly advancing for a minority of the world’s languages. At the same time, the majority of languages are falling behind (Joshi et al., 2020). It is essential that language technology can serve the speakers of a wide variety of languages, who come from a wide variety of cultures. In this paper, we argue that doing so requires accommodating these speakers not only on a linguistic level, but also on a cultural level.

Culture, like language, is a term that is hard to pin down, but generally describes the way of life of a collective group of people, and distinguishes them from other groups with other cultures (Mora, 2013; Shweder et al., 2007). Culture encompasses both material as well as non-material aspects, such as beliefs and linguistic practices (Kendall et al., 2005). Moreover, since “[c]ulture is the acquired knowledge people use to interpret experience and generate behavior” (Spradley, 1972), it is also the lens through which people understand linguistic messages. Culturally maladapted messages can

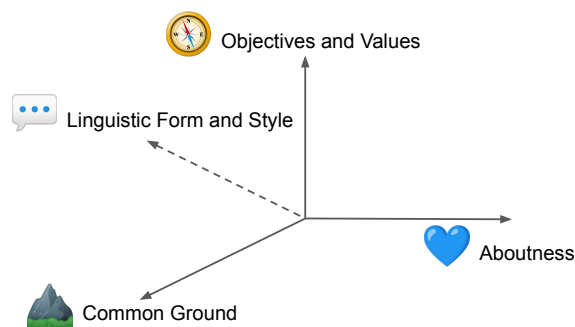


Figure 1: The role of culture in NLP, illustrated by four dimensions along which cultures vary, and for which NLP can be culturally biased: linguistic form and style, common ground, aboutness, and objectives (values).

and will be misinterpreted; NLP must be culturally sensitive in order to avoid doing harm.

Language and culture interact in a number of ways (see Figure 1). This paper aims to illuminate these connections, in order to motivate and inform future NLP work. Beyond **linguistic form and style**, *how* things are expressed in language, they include **common ground**, the shared knowledge based on which people reason and communicate; **aboutness**, *what* information people care to convey; and **objectives** or values, the goals people strive for (e.g., when developing language technology).

Since language and culture are intertwined, the different dimensions may be difficult to tease apart (Hovy and Yang, 2021): for example, the lexicon is shaped both by the need to convey information people care about, and by the conceptual categories grounded in the environment. Style is related to values and societal structure: formality levels in Japanese, for example, reflect the hierarchy of Japanese society (Gao, 2005). Nevertheless, language and culture are not interchangeable terms. Culture varies greatly *within* languages (Lin et al.,

2018). For example, the assumption that “English” in and of itself carries a single set of worldviews, interests and norms is unjustified (Paul and Girju, 2009; Wilson et al., 2016). On the other hand, a relatively homogeneous culture can span multiple languages, as in the Nordic countries (Sahlgren et al., 2021).

Contributions. We propose a framework for understanding the challenges that cultural diversity poses for NLP to serve all users, as well as the opportunities that NLP creates to understand these differences better. We consider four elements: linguistic form, common ground, aboutness and values. We then survey existing strategies in NLP to address these challenges. Highlighting limitations in current strategies, as well as successful examples, we propose directions for future development of cross-cultural NLP.

2 Linguistic Form and Style

Linguistic form refers to non-semantic questions of *how* to formulate an utterance. Most work on cross-lingual NLP has focused on how linguistic form varies between languages. However, the impact of social and cultural factors on linguistic form and stylistic variations is rarely discussed (Hovy and Yang, 2021).

Variation within language. How to circumscribe and define a particular language is a difficult problem (Eberhard et al., 2021). A language spoken in different countries often becomes standardized in slightly different ways (e.g., German in Austria and Germany). Geo-cultural variation within a language also gives rise to *dialects* (Zampieri et al., 2020; Wolfram and Friday, 1997; Brown et al., 2020), which in turn operate as an important sign of cultural identity (Falck et al., 2012). In addition, *sociolects* vary across social groups, which can include subcultures (McCormack et al., 2011).

Due to these variations, treating “a language” as a homogeneous mass limits cultural adaptation, and runs the risk of privileging certain cultures over others. Zhang et al. (2021) find that pre-trained language models (PLMs; see §6) reflect certain sociolects more than others. For example, there are considerable morphosyntactic variations between Spanish spoken in Spain and Argentina (Bentivoglio and Sedano, 2011), but they are not considered separately in a Spanish PLM (Cañete et al., 2020). A PLM specific to the Algerian di-

lect of Arabic performs better than a multilingual or general Arabic one (Antoun et al., 2020) in sentiment classification (Abdaoui et al., 2021).

Stylistic variation. Factors such as directness and formality are often associated with different communicative styles across cultures. For example, the level of politeness and thus also how offensive something is perceived depends very much on communicative norms (Gao, 2005; Larina, 2015). Misunderstandings that arise from different communicative styles can occur in any interaction between people of different cultural backgrounds. Thomas (1983) refers to this as *pragmatic failure*, namely the “inability to understand ‘what is meant by what is said’”, due to *how* it is being said. Comparative stylistics (Vinay and Darbelnet, 1995) aims to characterise these differences.

The same intention (e.g., to be polite) can lead to different forms in different cultures. For example, an offer of help can be made in an imperative form in Polish, while in English speaking Anglo-Saxon cultures, this form would be considered rude and so an interrogative form would be used instead (Wierzbicka, 2009). Similarly, German native speakers tend to use a high level of directness, which would be considered offensive in English (House and Kasper, 2011). Based on the observation that such differences exist, Ringel et al. (2019) applied distant supervision by inducing labels for English formality and sarcasm detection based on language—in this case German and Japanese, respectively.

The expression and perception of emotion also varies across cultures, both in text (Kirmayer et al., 2001; Ryder et al., 2008) and in face-to-face communication (Hareli et al., 2015). These differences are critical for cross-cultural sentiment analysis (Bautin et al., 2008) and for text-based recognition of medical conditions such as depression. For example, Loveys et al. (2018) find clear differences in linguistic form across English speakers of different ethnic identities with self-reported depression (e.g., the ratio of positive to negative emotion expression). Not understanding cross-cultural variation could lead to misclassification of movies, or worse, mis-diagnosing people.

3 Common Ground

A culture is in part defined by a shared *common ground*: the shared body of knowledge that can be talked about and that can be assumed as known by

others. This common ground varies from culture to culture, and thus cross-cultural language use has to take into account shifts in common ground.

These cross-cultural shifts, often correlating with cross-lingual shifts, are neglected in NLP. An underlying assumption in many approaches to multilingual NLP is that “different languages share a similar semantic structure” (Miceli Barone, 2016). However, the assumption that there is a cross-lingual, cross-culturally common semantics to preserve fails when the common grounding does not match between cultures. Two relevant aspects here are the set of relevant *concepts*, closely identified with problems of lexicalisation, and *common sense*, i.e., the relevant propositional knowledge used in reasoning and entailment.

Conceptualisation. People carve up the world using conceptual categories (Margolis and Laurence, 2021). While some general or even universal patterns exist (Wierzbicka, 1996), these categories can and do differ between languages and cultures: the domain of colour is a famous example (Berlin and Kay, 1969), but cross-cultural differences are also reflected in kinship systems (the lexicalization of family structures Murdock, 1970), spatial relations (Bowerman and Choi, 2001) and basic objects (e.g., where to draw the distinction between a ‘cup’ and a ‘mug’, or the ‘hand’ and the ‘arm’; Majid et al., 2015). Translating between languages thus entails translating between conceptualisations, which can be impossible if the conceptual grounding is not available (is this Danish “kop” an English ‘cup’ or a ‘mug?’)—this is the motivation behind multimodal, i.e., visually grounded, translation (Specia et al., 2016). Data collected from a single culture may contain concepts that do not easily map across cultures, as for example seen in the recent MarVL dataset (Liu et al., 2021), in which about a quarter of concepts from a sample of non-Western cultures did not map to English concepts (as represented by WordNet).

Commonsense knowledge. “Common sense” is the knowledge that is held *in common* by a community and culture, the communal knowledge bank that can be presumed to be known by everyone. Common sense thus covers a diverse set of knowledge types, from physical and temporal reasoning to psychological, social and moral judgements (Sap et al., 2020). Some aspects of common sense are fairly universal, in particular those that arise from

inhabiting a grounded human body on earth (dropping a glass of juice onto a hard surface is likely to cause a mess; the baby might be crying because it’s hungry). On the other hand, axioms of social and moral common sense, for example the how and why of rituals such as marriages, vary between cultures (Acharya et al., 2021).

Moving between different banks of common sense can involve different strategies to either provide missing background/common knowledge, or to transfer the content to a target-culture appropriate setting (see §6.3).

4 Aboutness

Cultures promote different topics and issues (Seibert et al., 2002), sometimes by necessity, sometimes by accident.¹ Some of the domains commonly considered in (English-speaking, Anglo-western) NLP are irrelevant to some cultures, while others *mean* different things to different cultures, i.e., they involve completely different practices. An example of the former is beer reviews, a go-to domain in sentiment analysis (Zeng et al., 2019; Ji et al., 2020; Paranjape et al., 2020). Beer reviews are hardly meaningful in cultures with no beer consumption. An example of the latter, i.e., a domain that is used differently across cultures and social groups, is Twitter (Hine, 2020). As an example of multi-domain sentiment analysis, Liu et al. (2018) consider cameras, laptops, restaurants, and movies, within the context of data from U.S. American websites. These domains, however, apply differently in other cultures: restaurants are more important in some cultures (e.g., Copenhagen hipsters) than others (e.g., Sámi); similarly, laptops have not penetrated cultures equally and have in fact been found to *change* culture (Hansen et al., 2014).

Beyond conceptualisation differences (see §3), Liu et al. (2021) point out how the visual concepts at the core of many multimodal NLP tasks reflect a Northern American and Western European bias in the underlying data sources, from which images were scraped. As an alternative, in their new

¹What we, as human beings, are interested in is, in part, a product of our group memberships. Our group membership determines conditions, e.g., our dependence on weather, agriculture, livestock, etc., making some topics vital, and promotes others e.g., through fashion, trends, etc. Heidegger (1927, 1953) refers to the former as the *facticity* aspect of our *caring*; the latter as the *fallenness* aspect. Unbiased treatment of the facticity aspect of what cultures care about, has been argued to be particularly important from an ethical perspective (Nelson, 2008).

dataset (MarVL), they let the selection of both concepts and images be entirely driven by native speakers. This corrects for some existing asymmetries in what our evaluation data tends to be about. For example, Althoff et al. (2014) created a dataset entirely devoted to *pizzas*; in Liu et al. (2021), models’ ability to recognise *vadas*, an Indian food, is evaluated.

People describe the same events differently across cultures. News reports emphasise different goals, motivations, methods and content in different parts of the world (Bandura, 1986; Li, 1997; Loo, 2019). However, many news resources used in NLP research reflect only some cultures (Breed, 1955; Galtung and Ruge, 1965; Marcus et al., 1993). As NLP technologies have been widely adopted to build multilingual news generators (Zhang et al., 2016; Xu et al., 2020; Jin et al., 2020), it is important that these differences are taken into account in language generation.

5 Objectives and Values

Common objectives in NLP include progress, accuracy, fairness, robustness and interpretability. They are driven by the research community, and reflect the values of this community: novelty, for example, is highly esteemed, which is not a value held by more conservative, tradition-valuing cultures. Grounding these objectives explicitly in ethical values and norms, and acknowledging that these may differ across cultures, is essential for cross-cultural NLP. Jiang et al. (2021), for example, introduce COMMONSENSE NORM BANK and Delphi, a moral reasoning resource and model, respectively, reflecting “English-speaking cultures of the United States in the 21st century”. However, researchers, practitioners, regulators and users may belong to different cultures and have different objectives and values (Talat et al., 2021).

Another common goal in the NLP community is the eventual expansion of all NLP technologies to low-resource languages. However, as Bird (2020) points out, this goal must not overshadow, or be placed above, the specific desires and needs of a given language community, as every community will have their own objectives for what they do and do not want for their language with regards to technology. Without building relationships with the communities for which we aim to develop NLP, there is a serious risk for imposing our own objectives on the community.

Importantly, entire categories of applications and tasks may be driven by cultural assumptions. For example, the main objective of text summarisation is brevity (Jørgensen and Søgaard, 2021). Likewise, while fluency is a common objective of machine translation, it may in fact, be less important than comprehensibility for users (Castilho et al., 2018).

Values differ across cultures. Values are an important part of non-materialistic culture, as these values define what a specific culture deems to be good or valuable, as opposed to bad and undesirable (Kendall et al., 2005). Accordingly, these values help inform what is accepted as normal and ethical behaviour, and shape common cultural attitudes. Aspects that clash against a culture’s values, norms, and ethics, may often be taboo or even illegal. As cultures around the world hold different values, their norms and ethics are inevitably different, too (Schäfer et al., 2015; Hanel et al., 2018). For example, in some cultures, alcohol consumption is prohibited for religious reasons, and thus may not be normal or acceptable behaviour; meanwhile drinking alcohol can be seen as normal behaviour by some cultures without this shared value. For this reason, beer reviews in an application, for example, may be seen not only as irrelevant (see §4), but even as offensive. While these values can and do change greatly from culture to culture, *end-users of NLP systems deserve technologies that are suitable to the culture that they belong to*. To this end, the preservation of cultural values, norms, and ethics in models intended for users from different cultures, is another important aspect of cross-cultural NLP (Solaiman and Dennison, 2021).

Fairness and combating biases. Many recent studies in NLP have explored strategies to counteract unwanted biases (according to the community’s values, see above) in models, resulting from prejudices present in society. These biases are implicit in the data used to train our models, and become baked into the models themselves (Bolukbasi et al., 2016; Sap et al., 2019; Sun et al., 2019; Blodgett et al., 2020; Bender et al., 2021). Because NLP datasets are composed of utterances from members of a larger culture, the resulting biases constitute a partial reflection of that culture’s values, norms, and ethics. For example, work in NLP on gender bias demonstrate that there is a pervasive problem with sexism in models (a direct result of sexist attitudes present in society), and explore ways to

remedy this (Friedman et al., 2019; Zhao et al., 2019; Webster et al., 2021; Baker Gillis, 2021). Accordingly, just as members of any culture deserve language technology that truly serves them, *members of vulnerable groups and marginalised members in larger cultures deserve such as well*. Current research in addressing fairness and combating biases in models demonstrates how NLP can be used to tackle the challenge of inequitable cultural attitudes. In other words, the goal of countering bias in NLP can be seen as shifting existing cultures into hopefully more equitable ones.

Conflicting objectives. With regards to cultural values, cross-cultural NLP lies at the intersection of the two important, but potentially conflicting, crossroads of multicultural pluralism and societal equity: cultures will have different values, which should be respected and represented for end-users; meanwhile, cultures will also have some inequitable attitudes that hurt different end-users, and NLP should be part of the solution. Together, these conflicting aspects land cross-cultural NLP in the impossible position of needing to both *preserve* cultural values, while also *minimising* harmful cultural biases. We acknowledge that these issues become increasingly difficult to disentangle, especially for more difficult or taboo aspects of culture. On the one hand, it is dangerous for members of one culture to impose values onto the others (i.e. this would, in fact, contribute to NLP colonisation—see §7), and on the other hand, it is dangerous to leave marginalised groups vulnerable, when we can diminish bias against them. While none of the strategies we discuss in §6 offers a simple solution to this challenge, our goal is to bring this important conundrum in cross-cultural NLP to the community’s attention. To help navigate this convoluted space, and thus ensure end-users have access to relevant NLP systems, we encourage people working in this space to collaborate with members of the culture relevant to their work, as they are the best equipped to judge what cultural aspects can appropriately be challenged. Additionally, new works in AI ethics should also be considered (Nanayakkara et al., 2021; Hagendorff, 2020; Prunkl et al., 2021; Pitta et al., 1999; Mohamed et al., 2020). An international regulatory instrument is essential for the responsible development of AI, a task that UNESCO is in the process of undertaking.²

²<https://en.unesco.org/courier/2018-3/towards-global-code-ethics-artificial-intelligence-research>

6 Strategies

We have identified several challenges along four dimensions of culture. Here, we highlight general strategies for cross-cultural NLP. We identify three main areas where we could direct efforts towards mitigating cross-cultural disparities: data collection, model training, and translation. Data is the backbone of NLP and any efforts towards cross-cultural NLP needs to consider the strategies involved in collecting and annotating it. Transfer learning is central to cross-lingual NLP and can serve an important role in cross-cultural NLP. Finally, translation is used to communicate between languages and will often be necessary when communicating between cultures. We explain what can be done within each of these areas each in turn.

6.1 Data Collection

The most fundamental issue is the representation disparity in our data, i.e., not all cultures are (equally) represented. While the volume of multilingual datasets increases and multilingual NLU benchmarks are becoming available (Hu et al., 2020; Liang et al., 2020; Ruder et al., 2021), this does not guarantee cross-cultural representation. There are two main factors: source of data (e.g., media outlet) and origin of annotations (e.g., automatic vs human-generated). To have truly diverse datasets, we should both ensure that they not only represent diverse sources, but also multiple perspectives in terms of annotations, when applicable.

Data selection and curation. Relevant to the impact of the source of data, Dodge et al. (2021) discuss how the Colossal Clean Crawled Corpus (C4; Raffel et al., 2020), an English web-based corpus, is skewed in favour of US governmental institutions and main-stream US media, while data filtering, which tries to remove slurs, or obscene words, disproportionately removes text from and about minorities (e.g., African American and LGBTQ+). Similarly, data from Wikipedia is heavily used in multilingual NLP even though it has been shown to be culturally biased (Callahan and Herring, 2011). To mitigate the risks associated with using culturally biased data, data selection or curation methods should strive to use data sources that are appropriate for the target culture of downstream NLP applications. Large, general-purpose datasets should be curated so as to be as unbiased as possible, and carefully documented (Jo and Gebru, 2020; Bender et al., 2021).

As examples of culturally diverse data collection, Liu et al. (2021) and Yin et al. (2021) recruit geographically diverse annotators from Africa, Asia and the Americas who are tasked with also providing the data (the images) along with the annotations (image descriptions). However, these datasets are small and thus limited to evaluation only; the cost and effort associated with collecting sufficient data to train modern NLP models means that finding culture-specific data to cover the enormous diversity of cultures represented on earth remains a formidable challenge. A diverse and open community, however, facilitates scalability in this regard. Examples for such communities include Universal Dependencies, dedicated to manual annotation of morphosyntactic datasets for over one hundred languages, covering the highest typological diversity to date among such datasets (Nivre et al., 2020). As another example, the Masakhane community aims to strengthen NLP research for African languages. It has created MasakhaNER (Adelani et al., 2021), a Named Entity Recognition dataset for 10 African languages, collected by native speakers of these languages.

Data annotation. With respect to annotation practices, a diverse pool of annotators reduces the risk of cultural bias. In the case of subjective tasks such as detecting affect, aggression, and hate speech, annotators may systematically disagree with one another due to cultural differences that are often reflected by their biases and values (Davani et al., 2021). Annotator disagreements may capture important nuances in such tasks that are often ignored while aggregating annotations to a single (possibly hegemonic) ground truth. Release of all annotations with each dataset, even disagreeing ones, allows training models that generalise better (Plank et al., 2014; Pavlick and Kwiatkowski, 2019; Prabhakaran et al., 2021; Gordon et al., 2022). Careful documentation of the annotation process also plays an important role. In a recent survey of machine learning papers (including NLP ones), Geiger et al. (2020) reported a wide divergence in the level of documentation about methodological practices in human annotation. They advocate for the importance of human annotation within the research process, arguing that it ought to be given as much attention, care, and concern as is currently placed on performance based metrics.

When arguing for an increasingly diverse ‘participatory design’ (Bodker et al., 2009), however, it is important to consider the values, ideologies,

codes, narratives, and power relations which govern the interaction between the assemblies of actors involved in data collection and annotation. This is the perspective of social-cultural studies and other critical viewpoints from the humanities (Mainsah and Morrison, 2014; Gray and Suri, 2019).

Annotation projection. Of course, data collection for many languages and cultures can be costly. Using parallel data and a word alignment tool, annotation (for example a syntactic tree) in a source language can be transferred to a target language without extra annotation (Yarowsky et al., 2001; Hwa et al., 2005). A related method to create multilingual datasets is translating a dataset (typically in English) into other languages (often using machine translation). While not usually called annotation projection, it can be considered a variation of this method, since the source annotation is ported to the translated data. For example, in XNLI (Conneau et al., 2018), the premise and hypothesis pair is translated from English to other languages and the label is reused for the translated pair. These methods help leverage data from high-resource languages to create more data for low-resource languages (Agić, 2017).

However, these methods risk ignoring target culture complexity or forcing the source culture concepts on the target culture. For example, English common sense datasets (Singh et al., 2021) include culture-specific concepts such as food ingredients,³ rituals and celebrations,⁴ and societal expectations.⁵ Translating this dataset into other languages will require making decisions about how, and whether, to modify these items to make them more intelligible in the target culture. Lin et al. (2021a) use machine translation to translate two common-sense reasoning datasets from English into 14 other languages. They attempt to deal with difficult cases by automatically flagging and removing examples which contain ‘social keywords’ from the dataset, or that are (again, automatically) labeled as containing non-neutral sentiment. However, these methods are unlikely to capture all examples of social behaviour and cannot identify examples of cultural over-specificity (sports teams,

³A hot sauce is going to be hotter if it uses jalapeño peppers rather than habanero.

⁴Colt doesn’t have any kids. Finley has four kids. Therefore, Finley is more likely to go Trick or Treating.

⁵Many people disapproved of the widow waiting one week after his [sic] wife’s death to start dating again, rather than one year.

jalapeños). Automatically translated *training* data can lead to worse performance than native target language data (Liu et al., 2021). However, if *evaluation* data is automatically translated too, we have no trivial way of exposing cultural biases introduced by the projection process. Culturally-aware evaluation thus necessitates data annotated directly in the target language, or at least culturally-sensitive human translations.

Ponti et al. (2020) point out that literal translation of datasets is sometimes impossible or undesirable due to culture-specific concepts in the source that may be missing or unnatural in the target. In their multilingual extension of the English Choice of Plausible Alternatives (COPA; Roemmele et al., 2011) dataset, they therefore asked “carefully chosen” human translators to perform culturally-sensitive translation, and either paraphrase, substitute the original concepts with similar ones that exist in the target language, or leverage phonetically transcribed loan words.

Human translation, or original data from the target culture, is clearly the expensive option, but will often be the only way to avoid cultural bias. Only translating/generating high-quality *evaluation* data is becoming an attractive middle ground option (Liu et al., 2021; Ponti et al., 2020; Yin et al., 2021) that at least allows us to judge the success of cross-lingual transfer in a culturally appropriate way.

6.2 Model Training

Models can be culturally biased even when trained on culturally diverse data. On the other end, diverse cultural representation can in some cases be achieved with specific training strategies, not only data selection and annotation.

Transfer learning and pre-training. A common strategy in machine learning is to *transfer* the knowledge acquired for a task, domain or language to another. It is used extensively in the context of cross-lingual learning, motivated by similarities between languages and cultures: if only the form differs between languages, then models can learn to abstract away from it and transfer from resource-rich languages, obviating task-specific training data in many languages (Agić, 2017; Wu and Dredze, 2019; Blloshmi et al., 2020).

In this context, approaches for creating cross-lingual word embeddings (Klementiev et al., 2012; Ammar et al., 2016) are based on the assumption that the semantic spaces of different languages are

approximately *isomorphic*. However, this assumption is violated in practice (Søgaard et al., 2018, 2019). Among linguistic reasons, cultural factors (e.g., conceptualisation) can play a role in the mismatch between the spaces. Though hardly ever considered when selecting the source language for model transfer, considering cultural factors improves performance on target languages in pragmatically motivated tasks (Sun et al., 2021).

In recent years, *model transfer* has been the mainstream paradigm with the advent of pre-trained language models (PLMs), and specifically, multilingual PLMs such as mBERT (Devlin et al., 2019) and XLM-R (Conneau et al., 2020). Language models require massive amounts of data to *pre-train*. Since corpora are largely skewed in favour of few languages, this leads to cross-lingual disparities. To mitigate this issue, Conneau and Lample (2019) introduced an exponential smoothing of the language *sampling* rate leading to a less skewed data selection. Conneau et al. (2020) and Xue et al. (2021) further studied the effect of the language sampling rate, and found that more uniform sampling improves performance in low-resource languages, but hurts high-resource languages.

The reliance on pre-training means that what PLMs encode has far-reaching consequences. Several works find differences in culture-specific commonsense knowledge in multilingual PLMs, depending on the language used to probe them. These include differences in factual knowledge (Kassner et al., 2021), grounding of time-of-day expressions (Shwartz, 2022), and social value (Lin et al., 2021b). Understanding these differences better will facilitate cultural adaptation and debiasing of NLP systems.

Training. The common methodology for training machine learning models (e.g., empirical loss minimisation) relies on maximising average performance across training examples (instead of groups, e.g., languages), which often leads to low minority performance, a phenomenon named *representation disparity* (Hashimoto et al., 2018). Model performance for minorities is often disregarded in favour of majority groups, as shown for race (Blodgett and O’Connor, 2017), gender (Jørgensen and Søgaard, 2021), and age (Zhang et al., 2021). Deriving fair models from biased data is a promising countermeasure (Mehrabi et al., 2021). In a cross-cultural setting, methodologies that account for model up-

dates for different groups (Hashimoto et al., 2018; Sagawa et al., 2020; Levy et al., 2020) could potentially reduce cultural biases. Most methods to balance for bias require access to protected demographic attributes (Sagawa et al., 2020), and these only partially reflect culture.

In this direction, Zhou et al. (2021) and Ponti et al. (2021) propose using *Group Distributionally Robust Optimisation* (group DRO; Oren et al., 2019) to optimise worst-case performance across languages. Similarly, de Lhoneux et al. (2022) used Worst-Case-Aware Curriculum Learning (Zhang et al., 2020) to improve group (language) performance parity in cross-lingual dependency parsing. However, Lent et al. (2021) find that group DRO has no benefit in a low-resource settings, namely for Creole languages. Such techniques can potentially be applied to pre-training multilingual language models with better cross-lingual parity, possibly in addition to improved data sampling. Furthermore, a fairer model with respect to other attributes, besides the language dimension, can lead to less culturally biased models. While these measures are widely discussed in a multilingual framework, they are also applicable in a monolingual setting to favour a more equal representation of different cultures in a single language (e.g., a fairer representation of English-speaking communities for English NLP).

Limitations. Nonetheless, cross-lingual countermeasures are *culture-agnostic*. In other words, even if we sample languages equally, cross-cultural disparity persists, as data for a given language is also not balanced in terms of sources, and hence cultures. However, one can directly target a more diverse cross-cultural representation by applying the same principles in terms of cultures instead of languages.

Ideally, methods such as data sampling or group DRO should facilitate generalisation of universally-common knowledge from highly-represented cultures, while granting equal representation to minority cultures. Overemphasising the former is problematic as it contributes to *cultural homogenisation*, while the latter enables mitigation of cross-cultural biases present in the underlying training data and models. There is no substitute for a larger representation of minority cultures in the data used for training and evaluation.

Finally, practical concerns include the level of granularity at which cultural groups are defined and

annotated in the data (e.g., as metadata attributes or criteria for splitting text corpora). There is a continuum from representing each individual separately (which may raise privacy concerns) to considering large, culturally diverse groups (such as all Spanish speakers) as a homogeneous mass. Again, NLP dataset creators as well as model developers must be aware of the trade-off between generalisation and adaptation here.

6.3 Translation

Beyond NLP for various cultures, cross-cultural NLP can be used for bridging *between* cultures, investigating cross-cultural communication. The classic example is machine translation (MT) between languages. In MT, semantic divergences are usually treated as noise, or in any case, as imperfect translations (Briakou and Carpuat, 2021). A commonly accepted criterion for translation adequacy is that the semantics of the source are preserved in the output (Carl, 1998; Sulem et al., 2015). However, when applied for translation across cultures, translation may have different objectives. The same meaning may be inappropriate in the target culture, regardless of fluency, and require adjustments: this is referred to as *adaptation* in the translation field (Vinay and Darbelnet, 1995). Peskov et al. (2021) observed that translated sentences are often opaque without cultural context (e.g., “I saw Merkel eating a Berliner from Dietsch on the ICE”), and propose adapting entities to the target culture by substituting them with their approximate counterparts. As an alternative strategy, adding explanations, for example in the form of appositives (Kementchedjhieva et al., 2020), can elucidate entities that are known in the source culture but not in the target culture. These are examples of *cross-cultural translation* (Sperber et al., 1994). This kind of adaptation can also be helpful in culturally-situated dialogue, where some users may be less familiar with the common ground of a particular culture; in such cases, Abou-Khalil et al. (2018) propose to explain by analogy to concepts from the user’s home culture.

Cross-cultural translation is not necessarily always between languages, but could take the form of style transfer within a language (Shen et al., 2017; Prabhumoye et al., 2018). For example, Jhamtani et al. (2017) transform text from modern English to Shakespearean English. Roy et al. (2015) approach personalised marketing by adapting the style of marketing messages for specific audience segments,

defined by geographic location and occupation.

Evaluation of cross-cultural translation is challenging, as the task is not always well-defined. In particular, for style transfer, human evaluation is more reliable than automatic evaluation, but still suffers from non-standard evaluation protocols (Briakou et al., 2021a,b). Reference-based automatic evaluation methods are particularly unreliable in this case, as the assumption that there is just one (or a few) *correct* translations is ostensibly violated (Song et al., 2013; Reiter, 2018). This stresses the need for culture-sensitive human evaluation.

7 Conclusion

In this paper, we have touched on a number of ways in which cultural knowledge, preferences and values can affect NLP practices. It is important to acknowledge the breadth of culture in all of these aspects, if we should aspire to a more cross-cultural NLP. However, we do not pretend to cover all relevant aspects of culture in our taxonomy: many works in sociology posture frameworks to better explain and catalogue the specific elements that compose a culture (Dant, 1999; Hofstede, 2001; Kendall et al., 2005; Woodward, 2007). We encourage further investigation of their impact on NLP.

Finally, present-day computational science have inherited colonising practices, which in NLP are realised as “homogenisation of perspectives” and “algorithmic monoculture” (Kleinberg and Raghavan, 2021; Bommasani et al., 2021). While there is general agreement in the NLP community that we need to represent cultures from outside Western, Educated, Industrialised, Rich, and Democratic (WEIRD) societies (Henrich et al., 2010), representation alone is not enough if we do not also allow them to prioritise their goals and values, rather than the goals and values held by the NLP community. Decolonisation (Bird, 2020; Mohamed et al., 2020; Birhane and Guest, 2021) aims to dismantle harmful power asymmetries and concepts of knowledge, turning us instead towards a “pluriversal epistemology of the future” (Mignolo, 2012) that unlike universalisms, acknowledges and supports a wider radius of socio-political, ecological, cultural, and economic needs (Mohamed et al., 2020).

Acknowledgements

The authors thank Desmond Elliott and Vinit Ravishankar for participating in preliminary dis-

cussions about the ideas presented in this paper. We are grateful to Omri Abend, members of the CoAStAL NLP group and the anonymous reviewers for their constructive feedback. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 801199 (*Heather Lent* and *Emanuele Bugliarello*). This work is also partly funded by the Innovation Fund Denmark (IFD)⁶ under File No. 0175-00011A (*Ilias Chalkidis* and *Anders Søgaard*). *Miryam de Lhoneux* was funded by the Swedish Research Council (grant 2020-00437).

8 Ethical Aspects and Broader Impact

As discussed in §5, norms and ethics are culture-dependent, and in some cases there is a conflict between maintaining researchers’ and practitioners’ ethical values (such as social equity) and multicultural acceptance. It is not our place to settle this conflict or to provide answers, but we stress that asking the question, of which values should prevail in each case, is essential for cross-cultural NLP. Furthermore, language technology for local communities must involve the members of those communities in an active, participatory manner, in order to decolonise language technology and respect the sovereignty of local people over their data (Bird, 2020; Mukhija et al., 2021).

References

- Amine Abdaoui, Mohamed Berrimi, Mourad Oussalah, and Abdelouahab Moussaoui. 2021. Dziribert: a pre-trained language model for the Algerian dialect. *arXiv preprint arXiv:2109.12346*.
- Victoria Abou-Khalil, Toru Ishida, Masayuki Otani, Brendan Flanagan, Hiroaki Ogata, and Donghui Lin. 2018. Learning culturally situated dialogue strategies to support language learners. *Research and Practice in Technology Enhanced Learning*, 13(1):1–20.
- Anurag Acharya, Kartik Talamadupula, and Mark A. Finlayson. 2021. [Toward an atlas of cultural commonsense for machine reasoning](#). In *Proceedings of the Workshop on Common Sense Knowledge Graphs (CSKGs)*, Online. Held in conjunction with the 35th AAAI Conference on Artificial Intelligence (AAAI 2021).
- David Ifeoluwa Adelani, Jade Z. Abbott, Graham Neubig, Daniel D’souza, Julia Kreutzer, Constantine Lignos, Chester Palen-Michel, Happy Buzaaba, Shruti

⁶<https://innovationsfonden.dk/en>

- Rijhwani, Sebastian Ruder, Stephen Mayhew, Israel Abebe Azime, Shamsuddeen Hassan Muhammad, Chris Chinenye Emezue, Joyce Nakatumba-Nabende, Perez Ogayo, Aremu Anuoluwapo, Catherine Gitau, Derguene Mbaye, Jesujoba O. Alabi, Seid Muhie Yimam, Tajuddeen Gwadabe, Ignatius Ezeani, Rubungo Andre Niyongabo, Jonathan Muki-ibi, Verrah Otiende, Iroro Orife, Davis David, Samba Ngom, Tosin P. Adewumi, Paul Rayson, Mofetoluwa Adeyemi, Gerald Muriuki, Emmanuel Anebi, Chiamaka Chukwunke, Nkiruka Odu, Eric Peter Wairagala, Samuel Oyerinde, Clemencia Siro, Tobius Saul Bateesa, Temilola Oloyede, Yvonne Wambui, Victor Akinode, Deborah Nabagereka, Maurice Katusiime, Ayodele Awokoya, Mouhamadane Mboup, Dibora Gebreyohannes, Henok Tilaye, Kelechi Nwaike, Degaga Wolde, Abdoulaye Faye, Blessing Sibanda, Orevaoghene Ahia, Bonaventure F. P. Dossou, Kelechi Ogueji, Thierno Ibrahima Diop, Abdoulaye Diallo, Adewale Akinfaderin, Tendai Marengereke, and Salomey Osei. 2021. [MasakhaNER: Named entity recognition for African languages](#). In *AfricaNLP workshop at EACL 2021*.
- Željko Agić. 2017. [Cross-lingual parser selection for low-resource languages](#). In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, pages 1–10, Gothenburg, Sweden. Association for Computational Linguistics.
- Tim Althoff, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2014. How to ask for a favor: A case study on the success of altruistic requests. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*.
- Waleed Ammar, George Mulcaire, Yulia Tsvetkov, Guillaume Lample, Chris Dyer, and Noah A. Smith. 2016. Massively multilingual word embeddings. *ArXiv*, abs/1602.01925.
- Wissam Antoun, Fady Baly, and Hazem Hajj. 2020. [AraBERT: Transformer-based model for Arabic language understanding](#). In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, pages 9–15, Marseille, France. European Language Resource Association.
- Noa Baker Gillis. 2021. [Sexism in the judiciary: The importance of bias definition in NLP and in our courts](#). In *Proceedings of the 3rd Workshop on Gender Bias in Natural Language Processing*, pages 45–54, Online. Association for Computational Linguistics.
- Albert Bandura. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice-Hall series in social learning theory. Prentice-Hall.
- Mikhail Bautin, Lohit Vijayarenu, and Steven Skiena. 2008. International sentiment analysis for news and blogs. In *ICWSM*.
- Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. [On the dangers of stochastic parrots: Can language models be too big?](#) In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21*, page 610–623, New York, NY, USA. Association for Computing Machinery.
- Paola Bentivoglio and Mercedes Sedano. 2011. Morphosyntactic variation in Spanish-speaking Latin America. *The handbook of Hispanic sociolinguistics*, pages 168–186.
- Brent Berlin and Paul Kay. 1969. *Basic color terms: Their universality and evolution*. University of California Press.
- Steven Bird. 2020. [Decolonising speech and language technology](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3504–3519, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Abeba Birhane and Olivia Guest. 2021. [Towards decolonising computational sciences](#). *Kvinder, Køn & Forskning*, 29(2):60–73.
- Rexhina Blloshmi, Rocco Tripodi, and Roberto Navigli. 2020. [XL-AMR: Enabling cross-lingual AMR parsing with transfer learning techniques](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2487–2500, Online. Association for Computational Linguistics.
- Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. [Language \(technology\) is power: A critical survey of “bias” in NLP](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5476, Online. Association for Computational Linguistics.
- Su Lin Blodgett and Brendan T. O’Connor. 2017. Racial disparity in natural language processing: A case study of social media African-American English. In *Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML 2017)*.
- Keld Bodker, Finn Kensing, and Jesper Simonsen. 2009. *Participatory IT design: Designing for business and workplace realities*. MIT press.
- Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. [Man is to computer programmer as woman is to home-maker? Debiasing word embeddings](#). In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.
- Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

- Melissa Bowerman and Soonja Choi. 2001. *Language acquisition and conceptual development*, chapter Shaping meanings for language: Universal and language-specific in the acquisition of semantic categories. Cambridge University Press.
- Warren Breed. 1955. Social control in the newsroom: A functional analysis. *Social Forces*, 33:326–335.
- Eleftheria Briakou, Sweta Agrawal, Joel Tetreault, and Marine Carpuat. 2021a. [Evaluating the evaluation metrics for style transfer: A case study in multilingual formality transfer](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1321–1336, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Eleftheria Briakou, Sweta Agrawal, Ke Zhang, Joel Tetreault, and Marine Carpuat. 2021b. [A review of human evaluation for style transfer](#). In *Proceedings of the 1st Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2021)*, pages 58–67, Online. Association for Computational Linguistics.
- Eleftheria Briakou and Marine Carpuat. 2021. [Beyond noise: Mitigating the impact of fine-grained semantic divergences on neural machine translation](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7236–7249, Online. Association for Computational Linguistics.
- Nina Brown, Thomas McIlwraith, and Laura Tubelle de González. 2020. *Perspectives: An Open Introduction to Cultural Anthropology*, 2 edition, chapter 4. Language. The American Anthropological Association.
- Ewa Callahan and Susan C. Herring. 2011. Cultural bias in Wikipedia content on famous persons. *J. Assoc. Inf. Sci. Technol.*, 62:1899–1915.
- Michael Carl. 1998. Meaning preservation in machine translation. *10th European summer school in logic, language and information (ESSLI'98)*.
- Sheila Castilho, Stephen Doherty, Federico Gaspari, and Joss Moorkens. 2018. *Approaches to Human and Machine Translation Quality Assessment*, pages 9–38. Springer International Publishing, Cham.
- José Cañete, Gabriel Chaperon, Rodrigo Fuentes, Jou-Hui Ho, Hojin Kang, and Jorge Pérez. 2020. Spanish pre-trained BERT model and evaluation data. In *PMLADC at ICLR 2020*.
- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. [Unsupervised cross-lingual representation learning at scale](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.
- Alexis Conneau and Guillaume Lample. 2019. [Cross-lingual language model pretraining](#). In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.
- Alexis Conneau, Ruty Rinott, Guillaume Lample, Adina Williams, Samuel Bowman, Holger Schwenk, and Veselin Stoyanov. 2018. [XNLI: Evaluating cross-lingual sentence representations](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2475–2485, Brussels, Belgium. Association for Computational Linguistics.
- Tim Dant. 1999. *Material culture in the social world*. McGraw-Hill Education (UK).
- Aida Mostafazadeh Davani, Mark Díaz, and Vinodkumar Prabhakaran. 2021. [Dealing with disagreements: Looking beyond the majority vote in subjective annotations](#).
- Miryam de Lhoneux, Sheng Zhang, and Anders Søgaard. 2022. [Zero-shot dependency parsing with worst-case aware automated curriculum learning](#). In *Proceedings of ACL*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Jesse Dodge, Maarten Sap, Ana Marasovic, William Agnew, Gabriel Ilharco, Dirk Groeneveld, and Matt Gardner. 2021. [Documenting the English colossal clean crawled corpus](#). *CoRR*, abs/2104.08758.
- David M. Eberhard, Gary F. Simons, and Charles D. Fennig. 2021. [Ethnologue: Languages of the world](#).
- Oliver Falck, Stephan Hebllich, Alfred Lameli, and Jens Südekum. 2012. [Dialects, cultural identity, and economic exchange](#). *Journal of Urban Economics*, 72(2):225–239.
- Scott Friedman, Sonja Schmer-Galunder, Anthony Chen, and Jeffrey Rye. 2019. [Relating word embedding gender biases to gender gaps: A cross-cultural analysis](#). In *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, pages 18–24, Florence, Italy. Association for Computational Linguistics.

- Johan Galtung and Mari Holmboe Ruge. 1965. The structure of foreign news: The presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers. *Journal of Peace Research*, 2(1):64–90.
- Fengping Gao. 2005. Japanese: A heavily culture-laden language. *Journal of Intercultural Communication*, 10:1404–1634.
- R Stuart Geiger, Kevin Yu, Yanlai Yang, Mindy Dai, Jie Qiu, Rebekah Tang, and Jenny Huang. 2020. Garbage in, garbage out? Do machine learning application papers in social computing report where human-labeled training data comes from? In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 325–336.
- Mitchell L Gordon, Michelle S Lam, Joon Sung Park, Kayur Patel, Jeffrey T Hancock, Tatsunori Hashimoto, and Michael S Bernstein. 2022. Jury learning: Integrating dissenting voices into machine learning models. *arXiv preprint arXiv:2202.02950*.
- Mary L Gray and Siddharth Suri. 2019. *Ghost work: How to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books.
- Thilo Hagendorff. 2020. [The ethics of AI ethics: An evaluation of guidelines](#). *Minds and Machines*, 30(1):99–120.
- Paul H. P. Hanel, Gregory R. Maio, Ana K. S. Soares, Katia C. Vione, Gabriel L. de Holanda Coelho, Valdiney V. Gouveia, Appasaheb C. Patil, Shanmukh V. Kamble, and Antony S. R. Manstead. 2018. [Cross-cultural differences and similarities in human value instantiation](#). *Frontiers in Psychology*, 9:849.
- Nina Hansen, Tom Postmes, Annika Tovote, and Anemarie Bos. 2014. How modernization instigates social change: Laptop usage as a driver of cultural value change and gender equality in a developing country. *Journal of Cross-Cultural Psychology*, 45(8):1229–1248.
- Shlomo Hareli, Konstantinos Kafetsios, and Ursula Hess. 2015. A cross-cultural study on emotion expression and the learning of social norms. *Frontiers in psychology*, 6:1501.
- Tatsunori Hashimoto, Megha Srivastava, Hongseok Namkoong, and Percy Liang. 2018. [Fairness without demographics in repeated loss minimization](#). In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1929–1938, Stockholm, Sweden. PMLR.
- Martin Heidegger. 1927, 1953. *Sein und Zeit*, 16. aufl. edition. Niemeyer, Tübingen.
- Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and brain sciences*, 33(2-3):61–83.
- Christine Hine. 2020. [The evolution and diversification of Twitter as a cultural artefact in the British press 2007–2014](#). *Journalism Studies*, 21(5):678–696.
- Geert Hofstede. 2001. *Culture’s consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage publications.
- Juliane House and Gabriele Kasper. 2011. Politeness markers in English and German. In *Conversational routine*, pages 157–186. De Gruyter Mouton.
- Dirk Hovy and Diyi Yang. 2021. [The importance of modeling social factors of language: Theory and practice](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 588–602, Online. Association for Computational Linguistics.
- Junjie Hu, Sebastian Ruder, Aditya Siddhant, Graham Neubig, Orhan Firat, and Melvin Johnson. 2020. [XTREME: A massively multilingual multi-task benchmark for evaluating cross-lingual generalization](#). *CoRR*, abs/2003.11080.
- Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara I. Cabezas, and Okan Kolak. 2005. Bootstrapping parsers via syntactic projection across parallel texts. *Natural Language Engineering*, 11:311 – 325.
- Harsh Jhamtani, Varun Gangal, Eduard Hovy, and Eric Nyberg. 2017. [Shakespeareizing modern language using copy-enriched sequence to sequence models](#). In *Proceedings of the Workshop on Stylistic Variation*, pages 10–19, Copenhagen, Denmark. Association for Computational Linguistics.
- Yunjie Ji, Hao Liu, Bolei He, Xinyan Xiao, Hua Wu, and Yanhua Yu. 2020. [Diversified multiple instance learning for document-level multi-aspect sentiment classification](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7012–7023, Online. Association for Computational Linguistics.
- Liwei Jiang, Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Maxwell Forbes, Jon Borchardt, Jenny Liang, Oren Etzioni, Maarten Sap, and Yejin Choi. 2021. [Delphi: Towards machine ethics and norms](#).
- Di Jin, Zhijing Jin, Joey Tianyi Zhou, Lisa Orri, and Peter Szolovits. 2020. [Hooks in the headline: Learning to generate headlines with controlled styles](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5082–5093, Online. Association for Computational Linguistics.
- Eun Seo Jo and Timnit Gebru. 2020. [Lessons from archives: Strategies for collecting sociocultural data in machine learning](#). In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT* ’20*, page 306–316, New York, NY, USA. Association for Computing Machinery.

- Anna Jørgensen and Anders Søgaard. 2021. [Evaluation of summarization systems across gender, age, and race](#). In *Proceedings of the Third Workshop on New Frontiers in Summarization*, pages 51–56, Online and in Dominican Republic. Association for Computational Linguistics.
- Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. [The state and fate of linguistic diversity and inclusion in the NLP world](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293, Online. Association for Computational Linguistics.
- Nora Kassner, Philipp Dufter, and Hinrich Schütze. 2021. [Multilingual LAMA: Investigating knowledge in multilingual pretrained language models](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 3250–3258, Online. Association for Computational Linguistics.
- Yova Kementchedjheva, Di Lu, and Joel Tetreault. 2020. [The ApposCorpus: a new multilingual, multi-domain dataset for factual appositive generation](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1989–2003, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Diana Elizabeth Kendall, Rick Linden, and Jane Lothian Murray. 2005. *Sociology in our times*. Thomson Wadsworth Belmont, CA.
- Laurence J Kirmayer et al. 2001. Cultural variations in the clinical presentation of depression and anxiety: implications for diagnosis and treatment. *Journal of clinical psychiatry*, 62:22–30.
- Jon Kleinberg and Manish Raghavan. 2021. Algorithmic monoculture and social welfare. *Proceedings of the National Academy of Sciences*, 118(22).
- Alexandre Klementiev, Ivan Titov, and Binod Bhattarai. 2012. [Inducing crosslingual distributed representations of words](#). In *Proceedings of COLING 2012*, pages 1459–1474, Mumbai, India. The COLING 2012 Organizing Committee.
- Tatiana Larina. 2015. [Culture-specific communicative styles as a framework for interpreting linguistic and cultural idiosyncrasies](#). *International Review of Pragmatics*, 7(2):195 – 215.
- Heather Lent, Emanuele Bugliarello, Miryam de Lhoneux, Chen Qiu, and Anders Søgaard. 2021. [On language models for creoles](#). In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 58–71, Online. Association for Computational Linguistics.
- Daniel Levy, Yair Carmon, John C. Duchi, and Aaron Sidford. 2020. [Large-scale methods for distributionally robust optimization](#). In *NeurIPS*.
- D. Li. 1997. New values in contrast: Western versus Chinese news reconstruction. *Asia-Pacific Media Educator*, 1:66–77.
- Yaobo Liang, Nan Duan, Yeyun Gong, Ning Wu, Fenfei Guo, Weizhen Qi, Ming Gong, Linjun Shou, Daxin Jiang, Guihong Cao, Xiaodong Fan, Ruofei Zhang, Rahul Agrawal, Edward Cui, Sining Wei, Taroon Bharti, Ying Qiao, Jiun-Hung Chen, Winnie Wu, Shuguang Liu, Fan Yang, Daniel Campos, Rangan Majumder, and Ming Zhou. 2020. [XGLUE: A new benchmark dataset for cross-lingual pre-training, understanding and generation](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6008–6018, Online. Association for Computational Linguistics.
- Bill Yuchen Lin, Seyeon Lee, Xiaoyang Qiao, and Xiang Ren. 2021a. [Common sense beyond English: Evaluating and improving multilingual language models for commonsense reasoning](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1274–1287, Online. Association for Computational Linguistics.
- Bill Yuchen Lin, Frank F. Xu, Kenny Zhu, and Seungwon Hwang. 2018. [Mining cross-cultural differences and similarities in social media](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 709–719, Melbourne, Australia. Association for Computational Linguistics.
- Xi Victoria Lin, Todor Mihaylov, Mikel Artetxe, Tianlu Wang, Shuohui Chen, Daniel Simig, Myle Ott, Nanman Goyal, Shruti Bhosale, Jingfei Du, Ramakanth Pasunuru, Sam Shleifer, Punit Singh Koura, Vishrav Chaudhary, Brian O’Horo, Jeff Wang, Luke Zettlemoyer, Zornitsa Kozareva, Mona Diab, Veselin Stoyanov, and Xian Li. 2021b. [Few-shot learning with multilingual language models](#).
- Fangyu Liu, Emanuele Bugliarello, Edoardo Maria Ponti, Siva Reddy, Nigel Collier, and Desmond Elliott. 2021. [Visually grounded reasoning across languages and cultures](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10467–10485, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Qi Liu, Yue Zhang, and Jiangming Liu. 2018. [Learning domain representation for multi-domain sentiment classification](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 541–550, New Orleans, Louisiana. Association for Computational Linguistics.
- Eric Loo. 2019. Reading “Asian values” into journalism practices in Asia. *Oxford Research Encyclopedia of Communication*.

- Kate Loveys, Jonathan Torrez, Alex Fine, Glen Moriarty, and Glen Coppersmith. 2018. Cross-cultural differences in language markers of depression online. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 78–87.
- Henry Mainsah and Andrew Morrison. 2014. Participatory design through a cultural lens: Insights from postcolonial theory. In *Proceedings of the 13th Participatory Design Conference: Short Papers, Industry Cases, Workshop Descriptions, Doctoral Consortium papers, and Keynote abstracts-Volume 2*, pages 83–86.
- Asifa Majid, Fiona Jordan, and Michael Dunn. 2015. Semantic systems in closely related languages. *Language Sciences*, 49:1–18.
- Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.
- Eric Margolis and Stephen Laurence. 2021. Concepts. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, Spring 2021 edition. Metaphysics Research Lab, Stanford University.
- Jo McCormack, Murray Pratt, and Alistair Rolls Alistair Rolls. 2011. *Hexagonal variations: Diversity, plurality and reinvention in contemporary France*, volume 359. Rodopi.
- Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A survey on bias and fairness in machine learning. *ACM Comput. Surv.*, 54(6).
- Antonio Valerio Miceli Barone. 2016. Towards cross-lingual distributed representations without parallel text trained with adversarial autoencoders. In *Proceedings of the 1st Workshop on Representation Learning for NLP*, pages 121–126, Berlin, Germany. Association for Computational Linguistics.
- Walter D. Mignolo. 2012. *Local Histories/Global Designs: Coloniality, Subaltern Knowledges, and Border Thinking*. Princeton University Press.
- Shakir Mohamed, Marie-Therese Png, and William Isaac. 2020. Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33(4):659–684.
- Cristina Mora. 2013. Cultures and organizations: Software of the mind intercultural cooperation and its importance for survival. *Journal of Media Research*, 6(1):65.
- Namrata Mukhija, Monojit Choudhury, and Kalika Bali. 2021. Designing language technologies for social good: The road not taken.
- G. P. Murdock. 1970. Kin term patterns and their distribution. *Ethnology*, 9(165).
- Priyanka Nanayakkara, Jessica Hullman, and Nicholas Diakopoulos. 2021. Unpacking the expressed consequences of AI research in broader impact statements. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, page 795–806, New York, NY, USA. Association for Computing Machinery.
- Eric Nelson. 2008. *Heidegger and the Ethics of Facticity*. SUNY Press.
- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. 2020. Universal Dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4034–4043, Marseille, France. European Language Resources Association.
- Yonatan Oren, Shiori Sagawa, Tatsunori B. Hashimoto, and Percy Liang. 2019. Distributionally robust language modeling. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4227–4237, Hong Kong, China. Association for Computational Linguistics.
- Bhargavi Paranjape, Mandar Joshi, John Thickstun, Hannaneh Hajishirzi, and Luke Zettlemoyer. 2020. An information bottleneck approach for controlling conciseness in rationale extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1938–1952, Online. Association for Computational Linguistics.
- Michael Paul and Roxana Girju. 2009. Cross-cultural analysis of blogs and forums with mixed-collection topic models. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1408–1417, Singapore. Association for Computational Linguistics.
- Ellie Pavlick and Tom Kwiatkowski. 2019. Inherent disagreements in human textual inferences. *Transactions of the Association for Computational Linguistics*, 7:677–694.
- Denis Peskov, Viktor Hangya, Jordan Boyd-Graber, and Alexander Fraser. 2021. Adapting entities across languages and cultures. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3725–3750, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Dennis A. Pitta, Fung Hung-Gay, and Steven Isberg. 1999. Ethical issues across cultures: Managing the differing perspectives of China and the USA. *The Journal of Consumer Marketing*, 16(3):240–246. Copyright - Copyright MCB UP Limited (MCB) 1999; Last updated - 2021-09-09; SubjectsTermNotLitGenreText - China; United States–US.

- Barbara Plank, Dirk Hovy, and Anders Søgaard. 2014. [Learning part-of-speech taggers with inter-annotator agreement loss](#). In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 742–751, Gothenburg, Sweden. Association for Computational Linguistics.
- Edoardo Maria Ponti, Rahul Aralikkatte, Disha Shrivastava, Siva Reddy, and Anders Søgaard. 2021. [Minimax and neyman–Pearson meta-learning for outlier languages](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1245–1260, Online. Association for Computational Linguistics.
- Edoardo Maria Ponti, Goran Glavaš, Olga Majewska, Qianchu Liu, Ivan Vulić, and Anna Korhonen. 2020. [XCOPA: A multilingual dataset for causal commonsense reasoning](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2362–2376, Online. Association for Computational Linguistics.
- Vinodkumar Prabhakaran, Aida Mostafazadeh Davani, and Mark Diaz. 2021. [On releasing annotator-level labels and information in datasets](#). In *Proceedings of The Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop*, pages 133–138, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Shrimai Prabhumoye, Yulia Tsvetkov, Ruslan Salakhutdinov, and Alan W Black. 2018. [Style transfer through back-translation](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 866–876, Melbourne, Australia. Association for Computational Linguistics.
- Carina E. A. Prunkl, Carolyn Ashurst, Markus Anderljung, Helena Webb, Jan Leike, and Allan Dafoe. 2021. [Institutionalizing ethics in AI through broader impact requirements](#). *Nature Machine Intelligence*, 3(2):104–110.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. [Exploring the limits of transfer learning with a unified text-to-text transformer](#). *Journal of Machine Learning Research*, 21(140):1–67.
- Ehud Reiter. 2018. [A structured review of the validity of BLEU](#). *Computational Linguistics*, 44(3):393–401.
- Dor Ringel, Gal Lavee, Ido Guy, and Kira Radinsky. 2019. [Cross-cultural transfer learning for text classification](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3873–3883, Hong Kong, China. Association for Computational Linguistics.
- Melissa Roemmele, Cosmin Adrian Bejan, and Andrew S. Gordon. 2011. [Choice of plausible alternatives: An evaluation of commonsense causal reasoning](#). In *Logical Formalizations of Commonsense Reasoning, Papers from the 2011 AAAI Spring Symposium, Technical Report SS-11-06, Stanford, California, USA, March 21-23, 2011*. AAAI.
- Rishiraj Saha Roy, Aishwarya Padmakumar, Guna Prasaad Jeganathan, and Ponnurangam Kumaraguru. 2015. Automated linguistic personalization of targeted marketing messages mining user-generated text on social media. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 203–224. Springer.
- Sebastian Ruder, Noah Constant, Jan Botha, Aditya Siddhant, Orhan Firat, Jinlan Fu, Pengfei Liu, Junjie Hu, Graham Neubig, and Melvin Johnson. 2021. [XTREME-R: Towards more challenging and nuanced multilingual evaluation](#). *CoRR*, abs/2104.07412.
- Andrew G Ryder, Jian Yang, Xiongzhao Zhu, Shuqiao Yao, Jinyao Yi, Steven J Heine, and R Michael Bagby. 2008. The cultural shaping of depression: Somatic symptoms in China, psychological symptoms in North America? *Journal of abnormal psychology*, 117(2):300.
- Shiori Sagawa, Pang Wei Koh, Tatsunori B. Hashimoto, and Percy Liang. 2020. [Distributionally Robust Neural Networks](#). In *International Conference on Learning Representations*.
- Magnus Sahlgren, Fredrik Carlsson, Fredrik Olsson, and Love Börjeson. 2021. [It’s basically the same language anyway: the case for a nordic language model](#). In *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 367–372, Reykjavik, Iceland (Online). Linköping University Electronic Press, Sweden.
- Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith. 2019. [The risk of racial bias in hate speech detection](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1668–1678, Florence, Italy. Association for Computational Linguistics.
- Maarten Sap, Vered Shwartz, Antoine Bosselut, Yejin Choi, and Dan Roth. 2020. [Commonsense reasoning for natural language processing](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 27–33, Online. Association for Computational Linguistics.
- Marie Schäfer, Daniel B. M. Haun, and Michael Tomasello. 2015. [Fair is not fair everywhere](#). *Psychological Science*, 26(8):1252–1260. PMID: 26115962.
- P S Seibert, P Stridh-Igo, and C G Zimmerman. 2002. [A checklist to facilitate cultural awareness and sensitivity](#). *Journal of Medical Ethics*, 28(3):143–146.

- Tianxiao Shen, Tao Lei, Regina Barzilay, and Tommi Jaakkola. 2017. [Style transfer from non-parallel text by cross-alignment](#). In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- Vered Shwartz. 2022. Good night at 4 pm?! Time expressions in different cultures. In *Findings of ACL*.
- Richard A Shweder, Jacqueline J Goodnow, Giyoo Hatano, Robert A LeVine, Hazel R Markus, and Peggy J Miller. 2007. The cultural psychology of development: One mind, many mentalities. *Handbook of child psychology*, 1.
- Shikhar Singh, Nuan Wen, Yu Hou, Pegah Alipoormolabashi, Te-lin Wu, Xuezhe Ma, and Nanyun Peng. 2021. [COM2SENSE: A commonsense reasoning benchmark with complementary sentences](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 883–898, Online. Association for Computational Linguistics.
- Anders Søgaard, Sebastian Ruder, and Ivan Vulić. 2018. [On the limitations of unsupervised bilingual dictionary induction](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 778–788, Melbourne, Australia. Association for Computational Linguistics.
- Anders Søgaard, Ivan Vulić, Sebastian Ruder, and Manaal Faruqi. 2019. Cross-lingual word embeddings. *Synthesis Lectures on Human Language Technologies*, 12(2):1–132.
- Irene Solaiman and Christy Dennison. 2021. Process for adapting language models to society (PALMS) with values-targeted datasets. *Advances in Neural Information Processing Systems*, 34.
- Xingyi Song, Trevor Cohn, and Lucia Specia. 2013. BLEU deconstructed: Designing a better MT evaluation metric. *Int. J. Comput. Linguistics Appl.*, 4:29–44.
- Lucia Specia, Stella Frank, Khalil Sima’an, and Desmond Elliott. 2016. [A shared task on multimodal machine translation and crosslingual image description](#). In *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers*, pages 543–553, Berlin, Germany. Association for Computational Linguistics.
- Ami D Sperber, Robert F Devellis, and Brian Boehlecke. 1994. Cross-cultural translation: Methodology and validation. *Journal of cross-cultural psychology*, 25(4):501–524.
- James P. Spradley. 1972. *Culture and Cognition: Rules, Maps, and Plans*. Chandler Publishing Company.
- Elior Sulem, Omri Abend, and Ari Rappoport. 2015. [Conceptual annotations preserve structure across translations: A French-English case study](#). In *Proceedings of the 1st Workshop on Semantics-Driven Statistical Machine Translation (S2MT 2015)*, pages 11–22, Beijing, China. Association for Computational Linguistics.
- Jimin Sun, Hwijee Ahn, Chan Young Park, Yulia Tsvetkov, and David R. Mortensen. 2021. [Cross-cultural similarity features for cross-lingual transfer learning of pragmatically motivated tasks](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2403–2414, Online. Association for Computational Linguistics.
- Tony Sun, Andrew Gaut, Shirlyn Tang, Yuxin Huang, Mai ElSherief, Jieyu Zhao, Diba Mirza, Elizabeth Belding, Kai-Wei Chang, and William Yang Wang. 2019. Mitigating gender bias in natural language processing: Literature review. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1630–1640, Florence, Italy. Association for Computational Linguistics.
- Zeerak Talat, Hagen Blix, Josef Valvoda, Maya Indira Ganesh, Ryan Cotterell, and Adina Williams. 2021. A word on machine ethics: A response to Jiang et al. (2021).
- Jenny Thomas. 1983. [Cross-Cultural Pragmatic Failure](#). *Applied Linguistics*, 4(2):91–112.
- Jean-Paul Vinay and Jean Darbelnet. 1995. *Comparative stylistics of French and English: A methodology for translation*, volume 11. John Benjamins Publishing.
- Kellie Webster, Xuezhi Wang, Ian Tenney, Alex Beutel, Emily Pitler, Ellie Pavlick, Jilin Chen, Ed Chi, and Slav Petrov. 2021. [Measuring and reducing gendered correlations in pre-trained models](#). *arXiv preprint*.
- Anna Wierzbicka. 1996. *Semantics: Primes and Universals*. Oxford University Press.
- Anna Wierzbicka. 2009. *Cross-cultural pragmatics*. De Gruyter Mouton.
- Steven Wilson, Rada Mihalcea, Ryan Boyd, and James Pennebaker. 2016. [Disentangling topic models: A cross-cultural analysis of personal values through words](#). In *Proceedings of the First Workshop on NLP and Computational Social Science*, pages 143–152, Austin, Texas. Association for Computational Linguistics.
- Walt Wolfram and William C. Friday. 1997. [The role of dialect differences in cross-cultural communication: Proactive dialect awareness](#). *Bulletin VALS-ASLA (Association suisse de linguistique appliquée)*, 65:143–154.
- Ian Woodward. 2007. *Understanding Material Culture*. SAGE Publications.
- Shijie Wu and Mark Dredze. 2019. [Beto, bentz, becas: The surprising cross-lingual effectiveness of BERT](#). In *Proceedings of the 2019 Conference on Empirical*

- Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 833–844, Hong Kong, China. Association for Computational Linguistics.
- Runxin Xu, Jun Cao, Mingxuan Wang, Jiase Chen, Hao Zhou, Ying Zeng, Yuping Wang, Li Chen, Xiang Yin, Xijin Zhang, Songcheng Jiang, Yuxuan Wang, and Lei Li. 2020. **Xiaomingbot: A Multilingual Robot News Reporter**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 1–8, Online. Association for Computational Linguistics.
- Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. **mT5: A massively multilingual pre-trained text-to-text transformer**. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.
- David Yarowsky, Grace Ngai, and Richard Wicentowski. 2001. **Inducing multilingual text analysis tools via robust projection across aligned corpora**. In *Proceedings of the First International Conference on Human Language Technology Research*.
- Da Yin, Liunian Harold Li, Ziniu Hu, Nanyun Peng, and Kai-Wei Chang. 2021. **Broaden the vision: Geodiverse visual commonsense reasoning**. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2115–2129, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Marcos Zampieri, Preslav Nakov, and Yves Scherrer. 2020. **Natural language processing for similar languages, varieties, and dialects: A survey**. *Natural Language Engineering*, 26(6):595–612.
- Ziqian Zeng, Wenxuan Zhou, Xin Liu, and Yangqiu Song. 2019. **A variational approach to weakly supervised document-level multi-aspect sentiment classification**. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 386–396, Minneapolis, Minnesota. Association for Computational Linguistics.
- Jianmin Zhang, Jin-ge Yao, and Xiaojun Wan. 2016. **Towards constructing sports news from live text commentary**. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1361–1371, Berlin, Germany. Association for Computational Linguistics.
- Sheng Zhang, Xin Zhang, Weiming Zhang, and Anders Søgaard. 2020. **Worst-case-aware curriculum learning for zero and few shot transfer**. *CoRR*, abs/2009.11138.
- Sheng Zhang, Xin Zhang, Weiming Zhang, and Anders Søgaard. 2021. **Sociolectal analysis of pretrained language models**. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4581–4588, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Jieyu Zhao, Tianlu Wang, Mark Yatskar, Ryan Cotterell, Vicente Ordonez, and Kai-Wei Chang. 2019. **Gender bias in contextualized word embeddings**. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 629–634, Minneapolis, Minnesota. Association for Computational Linguistics.
- Chunting Zhou, Daniel Levy, Xian Li, Marjan Ghazvininejad, and Graham Neubig. 2021. **Distributionally robust multilingual machine translation**.