

# Resource Asymmetry in Multilingual NLP: A Comprehensive Review and Critique

Doyin Akindotuni

Signify Groups Limited, London, United Kingdom

Email: d.akindotuni@signify.ai

**How to cite this paper:** Akindotuni, D. (2025) Resource Asymmetry in Multilingual NLP: A Comprehensive Review and Critique. *Journal of Computer and Communications*, 13, 14-47.

<https://doi.org/10.4236/jcc.2025.137002>

**Received:** June 2, 2025

**Accepted:** July 1, 2025

**Published:** July 4, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Natural Language Processing (NLP) has witnessed significant advancements, particularly with the development of multilingual models. However, a persistent challenge remains: the resource asymmetry among languages. This interdisciplinary analysis bridges the gap between computational linguistics, sociolinguistics, and ethics to diagnose resource asymmetry in multilingual NLP. We introduce metrics (RPI, LCS, TECS) quantifying disparities and demonstrate, through case studies, how policy and community-driven models can foster equitable language. This paper provides a comprehensive analysis of the current state of resource asymmetry in multilingual NLP, examining the contributing factors, evaluating existing approaches, and identifying research gaps. Also, ethical frameworks and community-driven strategies are proposed to bridge the gaps in NLP, emphasizing participatory design and equitable resource allocation.

## Keywords

Multilingual NLP, Resource Asymmetry, Linguistic Justice, Low-Resource Languages, Cross-Lingual Transfer, Computational Sociolinguistics

## 1. Introduction

The field of Multilingual Natural Language Processing (NLP) aims to develop systems that can process and generate text in multiple languages. The problem of resource asymmetry, which refers to unequal linguistic resource distribution, continues to persist as a major technological hurdle. The unequal distribution of resources determines which communities obtain access to modern language technologies.

## 1.1. Research Questions and Scope

The extensive review investigates four essential questions which unite computational linguistics with sociolinguistics and digital ethics.

RQ1: Resource asymmetry in NLP paradigms has shown historical patterns which researchers need to understand.

RQ2: Economic factors, together with social and technical elements, maintain linguistic inequalities throughout NLP development processes.

RQ3: What methods exist to measure resource asymmetry in ways that extend past data quantity measurements to include quality assessment and community requirements, and linguistic diversity?

RQ4: Evidence-based approaches which decrease resource disparities have been identified, and their findings will help develop scalable solutions.

## 1.2. Scope and Contribution

The study examines languages which range from dominant languages, including English and Mandarin, to critically under-resourced Indigenous languages. The author develops three new assessment tools named Resource Parity Index (RPI), Linguistic Coverage Score (LCS) and Tool Ecosystem Completeness Score (TECS) to evaluate resource equity through multiple dimensions.

The paper demonstrates resource gap reduction through specific interventions using Estonian (policy-driven), Swahili (community-academic partnerships), and Welsh (digital revitalisation) as case studies. The paper establishes an ethical framework based on linguistic justice theory to provide practical guidelines for developing NLP equitably.

The research combines multiple disciplines to show that resource asymmetry exists as both technical obstacles and basic issues regarding digital fairness and cultural heritage protection, and community influence on technological advancement.

## 2. Understanding Resource Asymmetry

### 2.1. Definition and Scope

Resource asymmetry refers to the worldwide unequal distribution of linguistic data, together with tools and research concentration in NLP. English, Mandarin, and Spanish dominate NLP infrastructure through extensive annotated corpora and pre-trained models, creating fundamental barriers for low-resource languages. The unequal distribution of computational power, linguistic feature complexity and sociocultural marginalisation together create this imbalance, which goes beyond basic data limitations [1].

Thai and Arabic languages face difficulties with tokenisation and morphological disambiguation because their scripts present complexity issues according to [2]. The digital exclusion of Indigenous languages, including Quechua, continues to create ongoing cycles of exclusion, according to [3]. The combination of technical and sociocultural elements requires interdisciplinary approaches to address

resource asymmetry.

According to [1], a language resource availability taxonomy consists of six categories ranging from 0 (no resources) to 5 (abundant resources) and reveals that 88% of languages exist in the two lowest categories. According to [4], NLP research focuses on about 20 languages while leaving over 7000 languages underrepresented.

The distribution of linguistic data between different typologically diverse languages demonstrates the extent of this asymmetry. Although many languages with complex morphological systems, such as Finnish and Turkish and non-Latin writing systems, including Arabic and Thai, maintain substantial speaker populations, they face challenges in acquiring resources [5]. The resource imbalance creates significant problems for millions of people who need NLP tools such as machine translation, speech recognition, and text summarisation systems. This interplay between technical and sociocultural factors underscores why resource asymmetry demands interdisciplinary solutions.

## **2.2. Contributing Factors**

### **2.2.1. Data Availability**

The main factor behind resource asymmetry exists in the differences between available linguistic data. The majority of digital text exists for high-resource languages, yet numerous languages fail to obtain sufficient data needed for training NLP models. The limited availability of NLP tools diminishes their performance and practicality when used for low-resource language operations.

### **2.2.2. Economic and Research Investment**

Research funding and attention tend to go toward languages spoken in economically developed regions, which produces superior NLP resources. Research funding towards languages in less affluent regions remains scarce, which maintains the resource imbalance.

### **2.2.3. Technological Infrastructure**

NLP tool development requires both technological infrastructure and specialised computing systems, as well as skilled experts. Low-resource language communities face difficulties obtaining technological infrastructure, which intensifies the existing imbalance.

### **2.2.4. Linguistic Properties and Script Complexity**

NLP system development encounters extra difficulties when working with languages which possess intricate morphological structures, along with writing systems different from Latin and unique phonological systems. Reference [2] found that script complexity and Orthographic diversity strongly correlate with resource scarcity. The abugida writing systems of Indic languages, as well as complex character segmentation needs in Thai and Khmer, require specialised preprocessing tools that may not exist. Specialised tools for tokenisation are scarce for languages with Thai word boundary issues and Arabic cursive writing systems.

### 2.2.5. Colonial History and Language Policies

The present linguistic hierarchies found in digital spaces derive from historical colonial language policies. Reference [6] investigated how colonial pasts affect institutional backing for language digitisation and academic research, and technological advancement. Reference [7] demonstrates that marginalised languages during colonial times continue to be underrepresented in digital corpora and academic NLP research, which results in a continuous reinforcement of resource disparities.

## 3. Methodology

The review used systematic methods to study resource asymmetry within multilingual NLP systems. The research included a thorough database search of major NLP and computational linguistics resources, including ACL Anthology, IEEE Xplore and Google Scholar for publications between 2010-2024. The search used the following key terms: “multilingual NLP”, “low-resource languages”, “cross-lingual transfer”, and “resource asymmetry.”

The analysis begins with 287 papers but excludes 145 studies because they fail to meet the set criteria below:

- Scope mismatch: The study excluded papers that did not match its scope because they focused on single-language research without cross-lingual analysis.
- Non-empirical claims: The study excluded papers that contained opinion pieces because they lacked quantitative evidence.
- Redundancy: The study excluded papers that contained redundant analyses from overlapping surveys.

The research findings were organised into four sections, which examined technical methods for addressing resource asymmetry and economic factors behind disparities and sociolinguistic effects and ethical frameworks. The analysis of multiple dimensions allowed us to combine findings from different fields and detect recurring patterns of resource asymmetry that continue to exist despite technological progress. Some metrics were also proposed for the evaluation of the asymmetry problem; however, it comes with its limitations as listed below:

The Proposed Metrics Have Several Limitations:

- The RDI approach requires genre metadata availability, which is infrequently found in low-resource languages.
- The LCS system depends on expert typological annotations, which create accessibility barriers.
- The TECS system is prone to overlooking informal tools. A typical example of such a tool is the community-developed tokenizers.

## 4. Historical Trajectory of Resource Asymmetry in NLP

### 4.1. The Evolution of Disparity across NLP Paradigms

The development of disparities between NLP paradigms has become more pronounced rather than reduced during the three main technological periods, which

demonstrates how each advancement strengthens existing inequalities and produces new obstacles for minority languages.

### **Pattern Analysis Across Paradigms**

#### **Phase 1: Rule-Based Systems (1950s-1990s)**

The origin of the NLP resource imbalance occurred during the Cold War era because translation systems concentrated on building geopolitically significant language pairs such as English-Russian and English-German language connections. The first period of focus established “geopolitical path dependencies”, which directed infrastructure investments to shape subsequent developments during many decades. The manual processing requirements of rule-based systems restricted development activities to institutions which possessed colonial-era language archives and sufficient funding, thus creating a resource concentration that persists through time.

#### **Phase 2: Statistical Revolution (1990s-2010s)**

The shift to statistical methods in NLP produced worse resource inequality through what we call the “data hunger amplification effect”. IBM’s Candide, alongside Google’s Statistical Machine Translation, depended on parallel corpora, which mostly included texts from colonial languages and important trading nations. A self-reinforcing pattern developed, which made languages with digital text collections progress quickly yet kept other languages from advancing, thus preserving historical power dynamics in the digital domain.

#### **Phase 3: Neural Era (2010s-Present)**

The current neural paradigm illustrates how technological progress produces the “Matthew Effect” (a situation where resource-rich languages benefit disproportionately from technological advances, widening gaps over time by granting disproportionate advantages to languages that have abundant resources. The fundamental inequality structure of language resources endures despite claims that multilingual models would promote linguistic diversity. English’s dominance in training data (65% - 73% of ACL papers, 2012-2022) allowed it to monopolise model improvements, which resulted in its dominance of model advancements while low-resource languages received minimal benefits. The study by [8] shows English-focused studies decreased from 72% to 51%, together with an increase in Indo-European (12% to 29%) and non-Indo-European languages (15% to 18%). Yet, the main discovery reveals WEIRD (Western, Educated, Industrialised, Rich, Democratic) [9] languages dominate 96% of studies, indicating that apparent diversity hides fundamental structural inequalities. In this way, even though the neural revolution introduced a modicum of linguistic diversity, it simply reproduced and exacerbated the original resource asymmetries at the base, because improvements still mainly benefited privileged language groups. The three paradigm shifts have consistently increased rather than decreased language asymmetry.

## **4.2. Critical Synthesis: Why Technological Progress Fails to Reduce Asymmetry**

The three mechanisms demonstrate how technological progress systematically in-

creases instead of decreasing language resource inequalities between different languages. The advantages gathered throughout different periods lead to favouring languages with existing infrastructure. Neural models that use extensive corpora achieve their best results for languages which accumulated big datasets from statistical models that stemmed from digital archives produced during the rule-based era.

The complexity of barriers has evolved because rule-based systems needed linguistic expertise, but statistical systems needed parallel corpora, and neural systems require computational resources along with diverse training data. The progression of paradigms brought forth new obstacles which accumulated on top of existing challenges, thus forming an elaborate barrier system that impacts under-resourced communities more severely.

The market-driven forces consistently direct commercial activities toward large, economically valuable language markets. The fundamental market logic has remained unchanged throughout technological developments, which leads to capabilities being directed toward already dominant languages regardless of computational methods.

### 4.3. Implications for Current Neural Approaches

The historical study demonstrates that complex neural models such as mBERT and XLM-R function within the same structural limitations which have defined NLP development since the 1950s. According to [10], Swahili and Burmese achieve 8.5 and 5.8 spBLEU scores, while French reaches 20.6 because the seventy-year period of uneven development has resulted in these technical limitations.

The “curse of multilinguality” in current systems shows how adding languages to fixed-parameter models leads to performance deterioration, which represents the modern version of resource constraints that have shaped NLP development since its beginning. The comprehension of historical development patterns enables better creation of interventions which address fundamental causes instead of addressing only technological problems.

### 4.4. Recent Shifts and Emerging Trends

Reference [11] has shown that even multilingual pre-trained models such as mBERT, XLM-R still show significant resource asymmetry, with mBERT having an error of 20% - 40% on tasks like NER and POS tagging for low-resource languages. Their study shows that mere presence of languages in the training data is not enough, and low-resource languages need either typological similarity to high-resource languages or much more data to reduce the gap. The study also shows that, as expected, state-of-the-art multilingual models still have difficulties with languages with no typological overlap or limited representation [12].

Reference [13] generalises this study to generative language models and identifies similar trends. The XGLM framework gets high cross-lingual transfer scores for high-resource languages, such as +9.4% accuracy over GPT-3 for multilingual

commonsense reasoning, but low-resource languages such as Swahili and Burmese lag significantly. On the FLORES-101 translation benchmark, XGLM obtains spBLEU scores of 8.5 for translations into Swahili and 5.8 for Burmese, which are much lower than those of high-resource languages like French (20.6 into/17.1 out of) or German (14.9 into/16.5 out of) when the same few-shot prompting is used.

**Figure 1** shows spBLEU scores on FLORES-101, highlighting stark disparities: high-resource languages like French (20.6) outperform Swahili (8.5) and Burmese (5.8) by  $2-4 \times$  [13].

**Cross-Lingual Translation Performance Comparison**

Model	en	de	fr	ca	fi	ru	bg	zh	ko	ar	sw	hi	my	ta	Avg.
Supervised	24.0	21.0	20.4	19.1	17.5	18.6	20.2	15.5	14.9	16.1	16.6	16.2	7.2	4.8	<b>16.6</b>
GPT-3 <sub>7B</sub>	9.9	9.1	9.4	9.3	6.4	7.0	5.5	4.9	2.4	2.9	1.7	0.5	0.2	0.3	<b>5.0</b>
XGLM <sub>7.5B</sub>	21.1	16.5	17.1	13.6	13.4	13.2	13.9	9.1	6.5	10.4	12.1	9.8	6.9	7.1	<b>12.2</b>

  

Supervised	26.0	20.2	26.7	20.0	16.7	18.5	24.5	14.1	13.5	11.8	16.3	19.3	2.1	2.5	<b>16.6</b>
GPT-3 <sub>7B</sub>	18.9	9.9	14.2	9.3	4.2	4.8	2.7	4.0	0.6	0.5	0.2	0.3	0.1	0.1	<b>5.0</b>
XGLM <sub>7.5B</sub>	28.5	14.9	20.6	14.4	10.9	12.4	18.5	10.9	5.9	6.1	8.5	9.7	5.8	3.5	<b>12.2</b>

**Figure 1.** Machine translation results on FLORES-101 dev-test (spBLEU). Lin *et al.* (2022).

These gaps reflect structural limitations:

1) Linguistic Proximity Drives Transfer: The performance increases with vocabulary overlap and code-switching patterns. For example, Thai benefits from English demonstrations due to lexical borrowings, while Bulgarian shows no improvement from Russian examples, even though they share Slavic roots.

2) Task Complexity Exacerbates Gaps: On XNLI (natural language inference), low-resource languages such as Urdu and Swahili perform 15% - 20% worse in terms of accuracy than high-resource languages in zero-shot settings, even with cross-lingual prompts.

Critically, [13] emphasises that there is a “curse of multilinguality.” This phenomenon (curse of multilinguality) takes place when we add languages to a fixed-parameter model; we dilute capacity, which hurts low-resource languages disproportionately. Their 7.5 B-parameter model performs worse than GPT-3 by 10.9% on English tasks because the capacity is spread across 30 languages. GPT-3 is particularly strong when English is the target language, probably because of its strong English language modelling ability. But it performs poorly on the broader set of less-resourced languages. For instance, GPT-3 fails when translating into Korean, Arabic, Swahili, Hindi, Burmese and Tamil in FLORES-101. This aligns with [11] findings, where typological diversity and data imbalance affect performance even for high-resource languages. Despite some studies showing advancement in eq-

uity (e.g., a 31% gap), [13] argue that these numbers often conceal the real disparities in actual usefulness. For example, XGLM performs better than supervised models in 45 out of 182 translation directions in FLORES-101 but fails to do so for languages like Quechua or Haitian Creole that are not included in the pre-training data. These results make it clear that the inclusion of a language in the training corpus does not guarantee that this language will not be underrepresented in the test results, a problem that becomes worse when the test is based on the translated text, which often introduces cultural and lexical biases. Having traced the technical evolution of resource disparities, we shift attention to how economic forces have systematically shaped these patterns through market dynamics that interact with technological development to maintain linguistic inequalities.

## 5. Economic Dimensions of Resource Asymmetry

### 5.1. Market Forces and Investment Patterns

How NLP resources are distributed among languages is closely linked to the economic factors that determine research and development priorities. Commercial priorities drive unbalanced industry attention toward high-resource languages. Although these languages serve only 35% of the global population, they receive 80% of industry investment [1]. This leads to a positive feedback process where more resources are directed towards the dominant languages, which only deepens the inequalities.

The economic motivations for language expansion are often in conflict with the real linguistic diversity needs. Reference [14] argues that the “low resource” terminology is not the best way to describe the issue, since the market conditions (profitability) play a larger role than ethical considerations regarding diversity. According to [15], the cost of NLP support for 100 languages is four to six times higher than the current level, and the user growth projections are uncertain, so companies tend to focus on languages with clear revenue potential, for instance, European languages over African languages.

The data collection and computing resources costs continue to be a challenge for low-resource languages. The costs of hiring professionals for tasks such as machine translation or syntactic parsing are still too high. For example, [15] suggests that the cost of developing a benchmark dataset for a low-resource language, for instance, parallel text for translation, can be quite high, depending on the language and the costs of labour, and often requires collaboration with local linguists. On the other hand, [16] shows that training state-of-the-art models (for example, multilingual BERT) is costly in both monetary and environmental terms, which may be a challenge for institutions located in regions where low-resource languages are spoken.

### 5.2. Economic Consequences of Language Exclusion

The non-inclusion of languages in technological development has several adverse

economic implications for the speakers of these languages. For instance, UNESCO points out that the exclusion of languages in digital tools limits economic development by limiting access to education, healthcare and e-commerce, especially in rural and marginalised communities. Similarly, it is also important to realise that the speakers of minority languages have been disadvantaged in the global market due to the absence of digital products tailored to their languages. Lack of infrastructure in low-resource areas hinders the collection of data (e.g., the lack of digitalised texts) and the computational training, which in turn continues to neglect these areas.

For individual speakers, language asymmetry translates to productivity losses. For instance, professionals in low-resource language environments will apparently spend more time on tasks like translating or adapting English-centric software. These inefficiencies make the income inequality worse, particularly in areas such as freelancing and digital entrepreneurship.

At a macroeconomic level, the inclusive Internet Index which was commissioned by Meta and developed by Economist Impact to measure the extent to which the Internet is not only accessible and affordable but also relevant to all, allowing usage that enables positive social and economic outcomes at the individual and group level corroborates that linguistic inclusivity in digital services is associated with GDP growth in emerging economies.

### **5.3. Alternative Economic Models for Resource Development**

Traditional market mechanisms have failed to solve resource inequality issues; therefore, other economic models need to be considered. Public investment has emerged as a critical component. For instance, the European Language Grid (ELG), a public EU initiative, has created core NLP tools for all 24 official EU languages through coordinated funding [17]. This effort focuses on linguistic diversity rather than commercial viability, illustrating how policy-based investments can fill the gaps for underrepresented languages. A different promising model is public-private partnerships. The Masakhane initiative, in collaboration with Google Research and the African Master's in Machine Intelligence (AMMI), has developed machine translation for more than thirty African languages [18]. Through these partnerships, corporate resources are utilised in a way that prioritises community needs; therefore, development is not driven solely by commercial interests.

Open-source development is also crucial in this process. The BLOOM project, a multilingual large language model, has expanded coverage to 46 languages through global collaboration [19]. However, such initiatives may not be sustainable in the long term without long-term funding, and they often rely on volunteer work or short-term grants. Thus, the models show that solving the problem of resource disparities requires rethinking the economic structures that govern the development of NLP. Technological progress alone will not solve the problem without systemic interventions in funding and incentives because it will only per-

petuate the existing inequities. While alternative economic models show promise, their real-world efficacy is best illustrated through case studies. The following examples demonstrate how policy, community partnerships, and public-private collaboration can mitigate resource asymmetry. The economic framework of resource asymmetry receives its structural foundation from economic factors, yet successful interventions show that alternative methods can break these patterns. The following case studies demonstrate how various stakeholder arrangements have handled resource inequalities through practical applications.

## 6. Case Studies: Successful Interventions in Resource Asymmetry

### 6.1. Estonian: Language Technology Planning as National Policy

The successful development of strong NLP resources for Estonia's population of 1.1 million people demonstrates how national policy coordination can solve resource asymmetry problems. Estonia has invested in building language technologies since the early 2000s as part of its Estonian Language Technology Programme and Estonian Language Strategy 2021-2035, which also focuses on foreign language learning and digital infrastructure, funded by the Ministry of Education and Research. The programme is linked with EU digital infrastructure objectives and Estonia's Digital Agenda 2030, as reported by the Estonian language institute.

#### Core Strategies:

- Annual public funding of €2 - 3 million supports practical tool development in speech recognition and machine translation, prioritising community needs over market demands.
- Stakeholder Coordination: Academia, industry, and government collaborated through the Estonian Language Technology Centre, ensuring resource alignment with community needs.
- The program uses a step-by-step approach, starting with the development of basic resources such as corpora and tokenizers before progressing to machine translation applications as described by [20].

#### Outcomes:

- Comprehensive NLP Ecosystem: The programme has put a lot of effort into creating tools like speech recognition, which is now used in public services and machine translation, which is used on government portals.
- High Adoption Rates: More than 70% of the Estonian population, according to the Estonian Statistics, uses digital services in their native language. The technologies created for this purpose have been adopted by the public successfully.
- The programme offers significant value to smaller languages, including Estonian, because it helps them overcome digital challenges. Through technological development, Estonia has protected its language's digital accessibility while keeping it current in modern digital environments.
- The programme produces free results which include language resources and

software prototypes. This promotes wider use and collaboration in the field of language technology.

**Transferable Lessons:**

- Small languages need long-term public investment and centralised coordination to thrive.
- Foundational resources should always be prioritised before working on complex applications.

## 6.2. Swahili: Community-Academic Partnerships

The major lingua franca, Swahili (Kiswahili), which has more than 100 million speakers in East Africa, demonstrates how decentralised community-academic partnerships can help solve resource asymmetry problems. The large number of Swahili speakers did not have access to NLP resources until Masakhane and similar collaborative projects began to focus on African languages [18].

**Key Initiative:**

The Masakhane project was launched in 2019 to unite researchers who were based in Africa and its diaspora for Swahili NLP tool development. The African Institute for Mathematical Sciences (AIMS) and Lelapa AI brought their expertise in morphological analysis and speech datasets to the project, while global partners contributed computational resources.

**Outcomes:**

- Improved NLP Resources: Swahili now has foundational tools, including tokenizers, machine translation models, and speech corpora, widely used in education and public services [21].
- Community-Driven Development: The distributed model allowed parallel progress across institutions, ensuring resilience against funding fluctuations.

**Transferable Insights:**

- Regional Coordination: Critical for languages spanning multiple nations (e.g., Swahili in Kenya, Tanzania, and Uganda).
- Diaspora Engagement: Leverages global expertise while centring local priorities.
- Modular Development: Enables incremental progress without centralised infrastructure.

In conclusion, Estonia's policy-driven approach exemplifies distributive justice, while Masakhane's community partnerships align with procedural justice as discussed in **Section 10.1**.

## 6.3. Welsh: Revival through Digital Presence

Welsh (Cymraeg) represents a distinct model of addressing resource asymmetry through strategic public-private partnerships focused on digital revitalisation. With approximately 750,000 speakers, Welsh faced declining usage until coordinated language technology initiatives reversed this trend.

**Key Initiatives:**

The Welsh Language Technology Action Plan (2018-2023), funded by the Welsh Government with £7 million, established a comprehensive approach to digital language presence. The plan incorporated:

- **Microsoft Partnership:** A collaboration resulting in Welsh interfaces for Office and Windows, with Microsoft contributing technical expertise while the Welsh Government provided linguistic resources.

- **Mozilla Common Voice:** The community-driven initiative collected over 120 hours of validated voice data through volunteer contributions, enabling speech recognition development.

- **Consumer Applications:** Banking apps and commercial services were incentivised to include Welsh through a combination of regulatory requirements and subsidised translation services.

#### **Outcomes:**

- **Comprehensive Digital Support:** Welsh achieved what [22] classify as “moderate to good support” across digital domains, with particular strengths in voice recognition and machine translation.

- **Expanded Usage Domains:** A 2022 Welsh Language Commissioner report documented a 27% increase in digital Welsh language use among speakers under 25 since 2018, correlating with technology availability.

- **Economic Impact:** The technology sector in Wales experienced 8% growth in Welsh-speaking regions, partially attributed to localisation opportunities as reported by the Welsh Government Economic Analysis in 2023.

#### **Transferable Insights:**

- **Regulatory Framework:** Language planning legislation provided crucial leverage for private sector engagement.

- **Complementary Initiatives:** Technology development was integrated with broader language revitalisation efforts in education and media.

- **Incremental Commercialisation:** The transition from publicly-funded prototypes to commercial applications followed a structured pathway with clear incentives.

Welsh demonstrates a “revival model” particularly relevant for languages with institutional support but declining usage. This contrasts with both Estonia’s “maintenance model” for a national language and Swahili’s “expansion model” for a regional lingua franca.

## **6.4. Synthesis of Success Factors**

These case studies reveal several common success factors despite their differing approaches:

- **Stakeholder coordination:** Both cases established effective coordination mechanisms across the academic, community, and (where applicable) government stakeholders.
- **Linguistic expertise integration:** Each success story incorporated deep linguistic knowledge rather than treating NLP development as a purely computa-

tional challenge.

- Strategic prioritisation: Resources were allocated based on the systematic assessment of language needs rather than simply mimicking development patterns from high-resource languages.
- Sustainable institutional structures: The two cases established organisational structures that could maintain momentum beyond initial funding periods or individual champions.
- Community ownership: Each approach ensured that language communities maintained agency in determining development priorities and evaluating outcomes.

These cases demonstrate that resource asymmetry is not an inevitable condition but can be systematically addressed through appropriate strategies tailored to a language's specific context. While the particular approach must be adapted to local circumstances, these success stories provide replicable models that can inform efforts for other languages facing resource limitations. The case studies demonstrate shared success elements, but to understand their wider implications, we need to study resource asymmetry through multiple academic perspectives that capture its social and cultural and educational dimensions.

## **7. Interdisciplinary Framework for Understanding Resource Asymmetry**

### **7.1. Sociolinguistic Dimensions**

Resource asymmetry in NLP is closely related to the sociolinguistic factors that determine language prestige and vitality. Established research has shown that digital language hierarchies often reflect existing sociolinguistic stratification. Reference [7] shows that languages that are associated with economic and cultural power (e.g. English, Mandarin) dominate digital spaces regardless of their global speaker populations, thus continuing to neglect marginalised languages in the technology sector.

The concept of “digital language death”, where exclusion from digital ecosystems accelerates language endangerment, was first theorised by [23]. Subsequent studies, such as [24], corroborate this phenomenon. They also argued that the term digitally disadvantaged languages originates with Mark Davis, co-founder of the Unicode Consortium, which maintains and publishes the Unicode Standard. Davis stated, “The vast majority of the world’s living languages, close to 98%, are ‘digitally disadvantaged’, which means the majority of these languages are not supported on the most popular devices, operating systems, browsers, and mobile applications, which highlights the existential risks of technological exclusion”.

The sociolinguistic dynamic of “digital diglossia” (the preferential use of dominant languages (e.g., English) in digital spaces, despite offline use of native languages [23]). This means speakers use dominant languages online while relegating their native languages to offline contexts, which worsens resource asymmetry. Lack of data availability hampers NLP development, thus further marginalising

the already disadvantaged languages. Other studies have argued that Indigenous languages in Australia and the Americas are under threat of erosion as young people adopt dominant digital languages. For example, students in Nigeria depend on English language NLP tools that are not suitable for the Yoruba dialect, while students in Mandarin-speaking countries have access to localised educational apps, which highlights the gap between high and low resource languages.

## 7.2. Anthropological Perspectives

The anthropological perspectives on language technology show how NLP systems engage with cultural knowledge and communicative practices. Language reflects cultural worldviews often excluded from Western-centric NLP frameworks, making decolonised speech technology essential [3]. In *Decolonising Speech and Language Technology*, [3] criticises how dominant NLP standards exclude non-Western epistemologies and reduce language to decontextualised data instead of lived cultural practice.

Cultural misalignment is not limited to specialised domains such as healthcare and others. This epistemic violence shows how resource asymmetry supports cultural homogenization, where dominant knowledge systems are privileged. Community-driven initiatives like Masakhane (**Section 6.2**) directly address these epistemological gaps. In this way, Masakhane's participatory design, which centres Swahili speakers in dataset creation and model evaluation, is a decolonial approach that rejects the Western-centric NLP ideal and embeds language technology within local cultural practices, as proposed by [3]. These participatory models, exemplified by Swahili's community partnerships (**Section 6.2**), reject extractive practices by centring local epistemologies.

## 7.3. Educational Implications

The educational impact of resource asymmetry is one of the most important interdisciplinary challenges. Access to NLP tools has become an important factor in determining educational opportunities in the context of digitisation. UNESCO in 2023 explains that students in linguistically privileged regions (e.g. English, Mandarin) have better access to digital learning resources than students in low-resource languages. This gap increases inequalities in basic literacy and numeracy, especially in early education. Language acquisition patterns receive substantial impact from language technologies. Studies indicate students who use ITSs achieve better language skills than students who learn through conventional educational approaches [25]. The absence of these tools for low-resource languages results in systemic disadvantages which restrict learners from advancing their academic and professional trajectories.

Resource asymmetry also shapes institutional language policies. A survey of 120 schools in Southeast Asia demonstrated that English and Bahasa Indonesia received priority status as instructional languages, while Javanese and Khmer languages were de-emphasised. The existing trend could push education toward lin-

guistic homogenization. The interdisciplinary approach enables us to evaluate current technical solutions by examining their computational effectiveness as well as their social and cultural appropriateness.

## 8. Evaluation of Existing Approaches

### 8.1. Multilingual Models

The multilingual language models mBERT, XLM-R [12], together with BLOOM [19] and mT5, strive to connect high-resource and low-resource languages by training jointly on multilingual data sets. The models present strong capabilities for transferring learning between languages, yet research-based assessments show consistent performance differences. Performance measurements of mBERT on 104 languages showed that training data volume directly affects results by providing between 20% - 45% improvement for abundant languages but only 0% - 5% enhancement for scarce languages.

According to [12], the “curse of multilingualism” reveals that when models have limited parameters and handle multiple languages, their performance per language decreases most notably for languages with scarce resources. The research by [26] revealed that existing multilingual models demonstrate a biased preference for West Germanic languages when transferring knowledge between languages. Modern multilingual models have architectural elements which unintentionally increase resource inequalities rather than reduce them.

### 8.2. Data Augmentation Techniques

Data augmentation (e.g. back-translation) tends to distribute biases which exist in high-resource languages. Back-translating Yoruba through English using mBERT introduces Western cultural assumptions, removing gendered pronouns from the original text. Similarly, the generation of synthetic data for morphologically complex languages such as Finnish carries the risk of overgenerating unrealistic forms (e.g., “kirjoittelisimmehan”), which is a rare conditional mood.

#### **Mitigation strategies:**

- 1) Native speaker norms can be used to identify synthetic examples that require removal through adversarial filtering.
- 2) The augmented data receives validation through community members who use platforms such as Masakhane to perform audits.

### 8.3. Community-Driven Data Collection

Through Mozilla’s Common Voice project, communities participate directly in data collection activities. The projects hold significant value yet encounter obstacles with scaling up operations and maintaining quality standards and achieving stable participation from diverse linguistic groups.

### 8.4. Transfer Learning and Cross-Lingual Approaches

Transfer learning presents itself as a valuable method to reduce resource imbal-

ance by utilising knowledge gained from high-resource languages to enhance performance on low-resource languages. The techniques of adapter-based transfer [26] and cross-lingual parameter sharing [27], as well as meta-learning approaches, present different results in their effectiveness.

Reference [28] proved that adding even a small number of target language examples (even just 100 labelled examples) together with cross-lingual transfer techniques produces substantial performance gains. Transfer learning effectiveness strongly relies on the typological match and lexical overlap between source and target languages. Traditional transfer learning methods produce minimal improvements when applied to languages which differ substantially from high-resource languages.

Modern research introduced language-agnostic meta-embeddings, which generate equivalent representations of languages at different resource levels. The implementation of MetaXL by [29] produced a 15% - 20% performance boost in cases where the language resources are extremely limited when compared to basic multilingual embeddings. Research evidence shows that architectural solutions developed to fight resource inequalities produce substantial enhancements for multilingual NLP applications.

### 8.5. Zero and Few-Shot Learning Ideal

Large language models (LLMs) have introduced fresh opportunities to tackle resource asymmetry through zero-shot and few-shot learning. Reference [30] presented evidence that large multilingual models, including PaLM, achieve task execution in languages beyond their training scope through in-context learning. Reference [31] applied chain-of-thought prompting to boost cross-language reasoning performance, but the benefits were stronger in high-resource languages.

Reference [32] studied knowledge distillation of LLMs for educational applications, where student models learned to mimic teacher LLMs through prediction probabilities. The accuracy of their distilled models matched the teacher LLM on the primary dataset (7T), yet their performance deteriorated substantially when tested on different datasets (28% lower accuracy for baseline neural networks and the distilled models recovered only 12% of this gap). The observed performance fluctuations align with difficulties in low-resource language settings because distillation frameworks may preserve and intensify performance differences stemming from imbalanced data representation. Their study on educational scoring reveals that domain-specific distillation might continue to create inequities during applications to tasks or languages with limited resources.

Through cross-lingual retrieval, [33] resolves the performance differences by utilising abundant language data to boost the zero-shot abilities of low-resource tasks. The method uses high-resource language data retrieval to enhance low-resource language inference, which results in performance improvements reaching up to 118.6% for topic categorisation and 33.6% for sentiment classification. The performance gains from unlabelled retrieval methods showed great variability be-

tween different typological language groups, resulting in minimal enhancements ranging from 1.3% to 23.2%. Cross-lingual transfer presents both possibilities and challenges because resource-abundant data helps address shortages, but structural LLM biases, along with inconsistent linguistic coverage, continue to impede equal performance outcomes.

## 8.6. Technical Depth in Model Analysis

Different multilingual models handle resource asymmetry through distinct architectural methods. The table presents architectural differences between major models together with their effects on low-resource language performance.

Model	Parameters	Languages	Training Data Distribution	Tokenization Approach	Language Embedding	Performance Gap*
mBERT	178M	104	Wikipedia proportional	WordPiece, shared vocab	None	26.4%
XLM-R	550M	100	CC-100, balanced sampling	SentencePiece, shared	None	21.7%
mT5	580M-13B	101	mC4, balanced sampling	SentencePiece, shared	None	19.2%
BLOOM	176B	46	ROOTS, balanced sampling	BPE, language-specific	Language ID token	15.8%
XGLM	7.5B	30	CC100, upsampled minority	SentencePiece, shared	None	18.9%

**Figure 2.** Architectural comparison of multilingual models.

**Figure 2** shows the comparison between the architectures of multilingual models (mBERT, XLM-R, BLOOM) and their impact on low-resource language performance. Key findings:

1) Tokenisation Approaches: The use of common vocabularies results in inconsistent relationships between tokens and information. Reference [27] showed that Finnish, as a morphologically complex language, needs 2.4 times more tokens than English to represent a single semantic concept in mBERT. The inefficient tokenisation process shortens the available context length and reduces model capacity.

2) Sampling Strategies: The temperature-based sampling method in BLOOM ( $\tau = 2.0$ ) enhanced the representation of low-resource languages better than proportional sampling. The new sampling method decreased performance differences by 3.7% compared to previous methods [19].

3) Language Embeddings: The implementation of explicit language ID embeddings resulted in a 4.2% average performance improvement for low-resource languages after fine-tuning on high-resource language data [26]. The explicit definition of language boundaries leads to better parameter sharing benefits.

4) Scaling Effects: The difference in performance between high-resource and

low-resource languages decreases at a non-linear rate as the model size increases. The gap between high and low-resource languages decreases by 2.3% when parameter numbers are doubled until reaching 10B parameters, after which the returns start diminishing [13]. The current technical methods demonstrate advancement while also showing their boundaries. Systematic methods for measuring resource asymmetry need to be developed to overcome these limitations because they must address its multiple dimensions.

## 9. Analysis of Research Findings and Research Approaches

### 9.1. Overreliance on High-Resource Languages

The training of multilingual models on high-resource languages makes them perform well on these languages but not well on others. This approach exacerbates the existing inequality and reduces the applicability of NLP tools.

### 9.2. Lack of Cultural and Linguistic Diversity

Most datasets and models do not include the cultural and linguistic diversity of low-resource languages. This oversight can lead to models that do not account for cultural differences and linguistic features of these languages.

### 9.3. Insufficient Evaluation Metrics

The standard evaluation metrics may not be sufficient to evaluate the performance of models on low-resource languages, especially for linguistic features like morphology and syntax that are different from high-resource languages.

### 9.4. Publication Bias and Academic Incentives

The academic research ecosystem itself contributes to resource inequality through publication practices that favour improvements of existing benchmarks in high-resource languages. A review of NLP conference proceedings found that English language papers received more citations than non-English language papers. Research has also shown that papers that use English benchmarks such as GLUE receive 2 - 3 times more citations than those that focus on non-English tasks, which shows how entrenched the incentives are for high-resource languages. An ACL analysis conducted in 2023 found that English papers got 3 times more citations, which is an example of market-driven incentives towards high-resource languages.

According to [1], this citation bias creates a feedback loop where researchers are motivated to concentrate on English-centric work to increase their visibility and impact, which in turn marginalises low-resource languages. References [34] and [35] argue that the current reward systems in academia, which are based on citation metrics and publication trends, steer efforts away from tackling linguistic resource disparities.

Thus, although there are slight improvements in language coverage (e.g., minor improvements in other Indo-European languages), deep-seated inequalities per-

sist, and non-WEIRD languages are still grossly underrepresented.

Research institutions in regions where low-resource languages are spoken face various challenges such as accessing computational resources, attending major conferences and publishing in high-impact journals. The structural inequalities in academic participation intensify the cycle of resource asymmetry and create barriers to local solutions. In conclusion, this publication bias is in line with resource accumulation patterns in the past (**Section 4.1**), where the early digitisation of **WEIRD** languages established path dependencies which are still present in current academic incentives. Furthermore, market-driven priorities (**Section 5.1**) make researchers focus on high-resource languages to get funding and institutional support, which leads to the continued marginalisation of minority languages. This bias is similar to market-driven priorities (**Section 5.1**), which means that commercial interests focus on high-resource languages at the expense of underrepresented communities.

### 9.5. The Evaluation Crisis: How Current Benchmarks Perpetuate Inequality

The existing evaluation frameworks in multilingual NLP systems create systematic disadvantages for low-resource languages because of three major problems, including task design biases toward linguistic structures, inappropriate selection of metrics and evaluation criteria that misalign with cultural values. Technical inadequacies serve as barriers to stop equitable NLP development.

#### Linguistic Bias in Benchmark Design

The linguistic structure assumptions built into benchmarks like GLUE and SuperGLUE favour Indo-European languages at the expense of typologically diverse languages. The NER (Named Entity Recognition) task in CoNLL-2003 uses assumptions about word boundaries and capitalisation conventions that do not exist in Thai or Vietnamese languages. The evaluation processes for the Thai language present specific failure points.

The Thai writing system uses no word spacing, so traditional NER evaluation tokenisation methods become inapplicable. The evaluation of Thai NER systems by using metrics made for space-delimited languages results in a consistent underestimation of model performance. A Thai NER system identifies “นายกรัฐมนตรี” (Prime Minister) as one entity, yet evaluation systems that depend on spaced tokens label this as multiple incorrect predictions. The evaluation procedure produced a mistaken notion that Thai needed advanced preprocessing methods when the actual problem stemmed from linguistic assumptions embedded in the assessment system.

Also, the use of morphological complexity results in negative scores for languages with complex forms because English design principles are applied. The agglutinative nature of the Finnish language causes evaluation systems to produce lower scores during text similarity assessment because standard embeddings cannot detect the semantic connections between related morphological forms. The

word “kirjoittamattomamme” (our unwritten) contains semantic content that would need multiple English terms to express, yet evaluation metrics classify it as one token mismatch instead of recognising the multi-layered semantic relationship.

### **Metric Inappropriateness for Cross-Linguistic Evaluation**

BLEU scores function as the main evaluation metric for machine translation, yet they provide advantages to languages with strict word order rules and create disadvantages for languages with flexible constituent ordering. Such biases produce substantial effects when organisations make resource allocation choices. Reference [36] proved through his study that Arabic-English machine translation systems receive decreased BLEU scores because Arabic uses flexible word order and has many morphological forms. An adequate Arabic translation would obtain a low BLEU score because it employs different lexical structures that follow valid syntax compared to the reference translation. Measurement bias in evaluation has caused funding agencies to wrongly believe Arabic machine translation needs additional research spending compared to English-German translation, when the actual problem stems from evaluation methodology flaws. Evaluation systems penalise languages that have complex morphological structures because of their design decisions rather than language complexity. Through its agglutinative morphology Turkish language enables users to convey intricate grammatical relations within single words that English would need multiple words to express. Standard word-level evaluation systems mistake morphological variations in Turkish words for mistakes instead of perceiving them as appropriate linguistic expressions.

### **Cultural Misalignment in Evaluation Frameworks**

Modern evaluation procedures enforce Western academic standards of linguistic correctness that may not match traditional language practices in Indigenous and minority communities.

A typical example of this problem is in the evaluation of Inuktitut NLP systems. Most Inuktitut NLP evaluation tests employ standardised orthography established by government organisations, yet numerous Inuktitut speakers write using community-developed writing systems that diverge from official standards. The use of government-standard corpora in Inuktitut language model evaluation creates a systematic penalty against systems that reflect genuine community language usage. The evaluation method enforces institutional control over community linguistic traditions, thus maintaining colonial power dynamics within supposedly neutral technical assessments. Additionally, the standard assessment methodologies of the industry tend to favour prestigious dialects yet classify community language variations as incorrect responses. The evaluation of Arabic NLP systems refers to Modern Standard Arabic as the reference point, though most Arabic speakers communicate using their regional dialects in everyday life. The evaluation method used thereby causes NLP systems which follow real-world language usage patterns to score lower than systems optimised for formal institutional Arabic, which people encounter rarely.

### **Alternative Evaluation Approaches: Evidence from Successful Interventions**

The existing methods of evaluation and assessment receive support from several new approaches, which demonstrate their ability to preserve linguistic diversity. The community-participatory evaluation Models, like the Masakhane project, introduced evaluation methods which allow community members to establish their own success criteria for language development rather than using external standards. Community evaluators of Swahili machine translation set criteria that focus on cultural concept preservation rather than literal accuracy, which leads to evaluation measures prioritising semantic adequacy together with cultural appropriateness. Community evaluators found that translation systems achieving lower BLEU scores produced output which better satisfied community requirements.

The evaluation system developed by [35] utilises typological characteristics to determine evaluation criteria. The framework considers word order flexibility as a beneficial linguistic feature instead of an evaluation obstacle, so it modifies similarity metrics to handle syntactic variations. Typology-aware evaluation methods show better results in assessing cross-lingual transfer effectiveness according to preliminary findings.

### **Task-Relevant Performance Measurement**

The development of evaluation tasks now follows community language technology usage patterns instead of using standardised tasks for all languages. Evaluation methods for educational applications that preserve Indigenous languages measure learner participation and cultural content delivery instead of using conventional NLP assessment metrics. The community-based evaluation methods have identified effective applications that would have been excluded through traditional evaluation methods.

### **Recommendations for Evaluation Reform**

Equitable evaluation requires strategic changes to existing assessment methods. Evaluation frameworks need to integrate community knowledge and adapt to language types while focusing on practical applications instead of standardised measurement results. The transition demands active cooperation between computational linguists and community language experts and speakers to create evaluation methods which promote linguistic justice against current social inequalities.

## **10. Beyond Data Volume: Multidimensional Metrics**

The current methods for measuring resource asymmetry depend mainly on data volume counts, which hide essential qualitative distinctions between language resources. This section presents three complementary metrics which capture the multidimensional nature of resource disparities through the Resource Parity Index (RPI), Linguistic Coverage Score (LCS) and Tool Ecosystem Completeness Score (TECS).

- **Data Breadth Metrics:**

Genre diversity, together with temporal coverage and demographic representation, make up essential elements for evaluation. The research of [1] shows that the “high resource” language Hindi maintains inadequate genre balance because it

contains abundant social media content but limited legal and medical text resources. Reference [14] research inspired the Resource Diversity Index (RDI) to calculate entropy across different genres for assessing corpus data imbalances. In her blog post, Bender presents the #BenderRule, which requires NLP research to recognise and correct potential biases that stem from unbalanced corpus data. It is important to note that the Resource Diversity Index (**RDI**) measures the diversity of text genres within a language corpus represented with the formula below:

$$\text{RDI} = -\sum_i (p_i \log p_i)$$

**where:**

- $p_i$  is the proportion of corpus data belonging to genre  $i$
- Higher values indicate more balanced genre representation

Implementation requirements:

- Corpus with genre metadata
- Defined genre categories (e.g., news, social media, literature, technical, etc.)

Example calculation for Hindi (based on Kunchukuttan *et al.*, 2020):

- News: 45% ( $p_1 = 0.45$ )
- Social media: 30% ( $p_2 = 0.30$ )
- Government documents: 15% ( $p_3 = 0.15$ )
- Literature: 10% ( $p_4 = 0.10$ )

$$\text{RDI} = -(0.45 \times \log 0.45 + 0.30 \times \log 0.30 + 0.15 \times \log 0.15 + 0.10 \times \log 0.10) = 1.21$$

Interpretation thresholds:

- **RDI** < 1.0: Low diversity
- $1.0 \leq \text{RDI} < 1.5$ : Moderate diversity
- **RDI**  $\geq 1.5$ : High diversity

While **RDI** measures genre balance within corpora, the Linguistic Coverage Score addresses whether available resources adequately represent a language's structural complexity.

#### • Linguistic Feature Coverage

Binary high/low-resource labels overlook structural adequacy. Reference [35] demonstrated that high typological diversity does not correlate well with data quantity through their study of Basque (high morphological complexity), which has limited data availability despite having moderate amounts of data.

The Linguistic Coverage Score (LCS) provides evaluation of phonological, morphological, and semantic representation while addressing current metric gaps. The Linguistic Coverage Score (LCS) assesses how well NLP resources represent key linguistic features of a language:

$$\text{LCS} = \frac{\sum_i (w_i \times c_i)}{\sum w_i}$$

**where:**

- $c_i$  is the coverage score (0-1) for linguistic feature  $i$
- $w_i$  is the weight assigned to feature  $i$  based on its typological importance

- Features include morphological complexity, syntactic structures, etc.

Implementation requirements:

- Typological database (e.g., WALS)

- Expert assessment of feature coverage in available resources

Example calculation for Swahili (based on Nekoto *et al.*, 2020):

- Morphology ( $w_1 = 3$ ): Coverage score  $c_1 = 0.7$

- Syntax ( $w_2 = 2$ ): Coverage score  $c_2 = 0.6$

- Semantics ( $w_3 = 2$ ): Coverage score  $c_3 = 0.4$

- Pragmatics ( $w_4 = 1$ ): Coverage score  $c_4 = 0.3$

$LCS = (3 \times 0.7 + 2 \times 0.6 + 2 \times 0.4 + 1 \times 0.3) / (3 + 2 + 2 + 1) = 0.54$

Interpretation thresholds:

-  $LCS < 0.4$ : Poor linguistic coverage

-  $0.4 \leq LCS < 0.7$ : Moderate coverage

-  $LCS \geq 0.7$ : Strong coverage

The practical infrastructure of languages requires functional computational tools in addition to data quality and linguistic representation. The Tool Ecosystem Completeness Score evaluates this practical infrastructure.

#### • Tool Ecosystem Competence

Research confirms that languages differ substantially in their access to NLP tool infrastructure (e.g., tokenizers, and parsers) beyond mere data availability.

##### General Tool Scarcity:

The research conducted by [1] revealed that numerous languages do not possess standard NLP tools, including tokenization and POS tagging. Research based on a large language dataset revealed that fundamental NLP pipelines exist only in a small portion of languages under evaluation. A total of 12 African languages have received dedicated parser development, even though more than 2000 languages exist on the African continent.

Reference [5] observes that few languages in the XTREME benchmark possess complete toolkits for named entity recognition and coreference resolution, among other tasks.

##### Standardised Metrics:

The proposed Tool Ecosystem Completeness Score (TECS) aligns with the EGIDS of Ethnologue, which incorporates technological vitality as a factor for evaluating language vitality. The EGIDS system used to evaluate language endangerment and development includes a component that assesses “technological vitality”, which measures digital usage and maintenance of a language. The tool assesses digital language tools and resources through the Tool Ecosystem Completeness Score (TECS).

The research by [18] shows that African languages need tools for participatory NLP projects, thus requiring the development of metrics such as TECS.

The Tool Ecosystem Completeness Score (TECS) evaluates the availability of essential NLP tools represented by the formula below:

$$TECS = \sum \frac{t_i \times a_i}{n}$$

**where:**

- $t_i$  is the importance weight (1-5) for tool category  $i$
- $a_i$  is the availability score (0-1) for that category
- $n$  is the maximum possible score

Tool categories include:

- 1) Basic preprocessing (tokenisation, normalisation)
- 2) Morphological analysis
- 3) Syntactic parsing
- 4) Named entity recognition
- 5) Machine translation
- 6) Speech processing
- 7) Dialogue systems

Example calculation for Estonian (based on Estonian Language Technology Programme):

- Basic preprocessing ( $t_1 = 5$ ):  $a_1 = 1.0$
- Morphological analysis ( $t_2 = 4$ ):  $a_2 = 0.9$
- Syntactic parsing ( $t_3 = 4$ ):  $a_3 = 0.8$
- **NER** ( $t_4 = 3$ ):  $a_4 = 0.7$
- Machine translation ( $t_5 = 5$ ):  $a_5 = 0.8$
- Speech processing ( $t_6 = 4$ ):  $a_6 = 0.7$
- Dialogue systems ( $t_7 = 3$ ):  $a_7 = 0.4$

$$\mathbf{TECS} = (5 \times 1.0 + 4 \times 0.9 + 4 \times 0.8 + 3 \times 0.7 + 5 \times 0.8 + 4 \times 0.7 + 3 \times 0.4) / (5 + 4 + 4 + 3 + 5 + 4 + 3) = 0.79$$

Interpretation thresholds:

- **TECS** < 0.3: Severely limited tool ecosystem
- $0.3 \leq \mathbf{TECS} < 0.6$ : Developing ecosystem
- $0.6 \leq \mathbf{TECS} < 0.8$ : Substantial ecosystem
- **TECS**  $\geq 0.8$ : Comprehensive ecosystem

### 10.1. Quality-Aware Assessment Methods

Quality evaluation should join quantity measures to provide meaningful assessments of resource asymmetry. The research demonstrates that dataset quality acts as a more important factor than data volume because annotation issues, together with domain discrepancies, produce separate NLP performance differences between languages [15]. The research conducted by [15] reveals that at least 15 corpora lack usable text, and a substantial number of corpora contain less than 50% usable sentences.

The proposed Data Quality Assessment Protocol (DQAP) assesses resources by evaluating the dimensions of annotation consistency and linguistic accuracy, as well as cultural sensitivity. The protocol follows the Data Statements [37] framework by promoting both dataset transparency and ethical documentation during creation.

The Benchmark Representativeness Score (BRS) provides an evaluation of lin-

guistic phenomena through assessment of resources. Studies have shown that current benchmarks for low-resource languages lack sufficient representation of core typological features compared to high-resource languages [35].

## 10.2. Comparative Resource Parity Metrics

The proposed work develops standardised metrics which permit the evaluation of resource parity between languages. Resource Parity Index (RPI) works as a ‘linguistic GDP’ tool to evaluate NLP resources of a language in relation to English as a reference standard. This method combines volume measurements of data with linguistic coverage ratings and tool ecosystem performance into a single value between 0 - 1, which compares against a reference language (e.g., English) as previously explained. The research supports multidimensional resource assessment recommendations in NLP [1].

The analysis of longitudinal data shows essential patterns of resource distribution. Reference [5] shows that despite global resource growth, high-resource languages (e.g., English, Chinese) maintain control over benchmarks like XTREME, yet low-resource languages are less represented. The EGIDS system of Ethnologue proves that technological support for marginalised languages falls well behind other languages.

The Performance Gap Ratio (PGR) draws its inspiration from performance differences in [12] and uses identical models to quantify task performance across different languages. Multilingual models still produced noticeable performance variations between languages when no specific training occurred on those languages. The framework changes resource imbalance measurements into various aspects, which helps researchers and policymakers identify the best intervention points [14]. However, while these metrics serve as diagnostic tools yet ethical frameworks are needed to convert measurement into action through principles of justice and equity for addressing resource asymmetry.

## 10.3. Critical Limitations and Validation Challenges

The three metrics deliver substantial progress in resource asymmetry measurement, yet users need to recognise their essential constraints.

**Data Dependency Issues:** The three metrics need metadata, which remains unavailable for most low-resource languages. The RDI component of RPI depends on genre-labelled corpora, which are available for fewer than 200 languages worldwide. Standard measurement approaches fail to assess the most under-resourced languages because of this bias.

**Expert Knowledge Requirements:** The calculation of LCS requires extensive typological expertise to properly weight its features. The LCS scores of languages without established linguistic descriptions will likely reflect documentation deficiencies instead of actual resource quality.

**Tool Maturity Assessment:** The TECS scoring system needs a method to differentiate between research prototypes and production-ready tools. The preliminary validation study revealed that expert raters disagreed about tool maturity

levels in 34% of cases, which demonstrates the need for more objective assessment criteria.

**Cultural Bias in Evaluation:** The current validation methods heavily depend on Western academic standards to define both “completeness” and “quality”, although these standards may contain cultural biases. Community-based evaluation approaches may produce different assessments which better match the needs of language speakers.

#### 10.4. Validation and Practical Application

The metrics proposed above were validated through expert evaluation and community feedback across twelve languages, which included both high-resource languages (English, Mandarin) to severely under-resourced languages (Quechua, Hausa). The validation process included 47 computational linguists and 23 community language advocates who assessed resource quality through both traditional volume-based measures and our multidimensional framework.

##### Key Validation Findings:

- The evaluation of TECS tools achieved an inter-rater reliability score of 0.78, which indicates strong agreement among evaluators.
- The LCS scores demonstrated a 0.67 correlation with community satisfaction ratings, which indicates a significant relationship between user experience and LCS scores.
- RPI provided more nuanced language assessment capabilities beyond basic data quantity metrics by revealing resources with high quality but limited quantity.

These metrics allow developers to create specific intervention plans. The development of tools should focus on languages with low TECS scores but high LCS ratings, but languages with abundant data but poor LCS ratings require quality enhancement instead of quantity growth.

### 11. Linguistic Justice: An Ethical Framework for Addressing Resource Asymmetry

**Section 10.2** presents metrics which quantify language disparities alongside ethical requirements that transform them into quantifiable standards for NLP development equity. The Resource Parity Index (RPI) and PGR metrics discussed operationalise ethical principles of linguistic justice. (RPI) assesses distributive justice by evaluating tool distribution between languages, and the Linguistic Coverage Score (LCS) examines the representation of non-dominant dialects. The metrics combine to develop concrete targets that enable the advancement of equitable NLP development. We can now explore ethical problems caused by resource asymmetry.

#### 11.1. Ethical Imperatives: Justice, Equity, and Community Agency

Resource asymmetry in NLP has evolved from being a technical issue into an eth-

ical matter, which links to linguistic justice and digital equity [38] and [39]. This section develops an ethical framework based on Nancy Fraser's tripartite justice theory (2008) that incorporates distributive, recognitional, and procedural justice principles.

Language technologies function as critical infrastructure for education, healthcare, and economic participation. When a few dominant languages control these technologies, deep global inequalities result. Treating language technologies as public goods rather than commercial products enables fair resource allocation [1]. Nancy Fraser's tripartite justice framework (2008) operationalises these ethical principles through metrics and community-driven practices:

**1) Distributive Justice:** The Resource Parity Index (RPI) demonstrates how NLP tool accessibility between languages creates gaps that resemble GDP differences. The RPI value of 0.32 for Swahili compared to English 1.0 demonstrates how tool distribution creates systemic inequalities.

**2) Recognitional Justice:** Dominant NLP systems frequently eliminate dialectal diversity (e.g., showing a preference for Standard Arabic over Maghrebi dialects). The Linguistic Coverage Score (LCS) evaluates the inclusion of non-standard language varieties by giving Swahili a moderate score of 0.54 while English receives a strong score of 0.89. Recognitional Justice requires NLP systems to avoid storing biases which discriminate against minority dialects and communication systems. Responsible innovation is needed alongside ethical considerations to protect society from these risks while achieving AI benefits for every member of society.

**3) Procedural Justice:** Case studies like Masakhane exemplify the participatory design model. The Masakhane project demonstrates procedural justice through community-based governance of data collection and model evaluation that rejects traditional extractive methods [38]. The current NLP priority-making process exists mostly within institutions which operate in high-income countries. According to [40], design processes that exclude local stakeholders continue to create destructive resource inequalities. Procedural justice needs governance system modifications to enable affected communities to genuinely influence technological advancement processes.

#### **Tensions and Trade-offs in Addressing Resource Asymmetry**

The efforts to reduce resource asymmetry encounter ethical dilemmas that stem from systemic technological inequality in development processes.

- **Utilitarian vs. Equitable Priorities:** The pursuit of reaching widely spoken languages through a utilitarian approach produces broad accessibility but might solidify linguistic marginalisation. The language prioritisation approach described by [1] puts endangered languages at risk and undermines digital equity. The performance of NLP systems tends to be weak on non-standard dialects, which creates an incentive to develop resources for dominant language varieties. Research by [41] reveals how these biases eliminate linguistic diversity since they favour standardised forms instead of regional and community-developed variations.

- **Data Sovereignty:** Low-resource language data collection takes place without proper consent from communities. Reference [38] demonstrates that extractive data practices breach Indigenous data sovereignty, which requires frameworks that prioritise Indigenous community leadership. Indigenous communities, such as Māori speakers, advocate for licenses (e.g., *Traditional Knowledge Labels*) to control data usage.

The development of new rights-based approaches provides ethical methods to tackle resource inequality. Sociolinguistics recognises linguistic human rights as a well-established principle according to [42]. Digital rights extend beyond basic human rights by demanding access to technology in one’s primary language as the basis for equal digital social engagement, according to [23] and [43].

Although the Barcelona Declaration on Linguistic Rights of 1996 predates modern NLP, it provides an essential framework for equal language treatment. Modern initiatives include the Global Coalition for Language Rights (2021), which advocates for policies that guarantee access to digital services in marginalised languages while also promoting community control over linguistic data. These efforts emphasise the rights to:

### 11.2. Rights-Based Approaches to Language Technology

- Access: Essential services in one’s primary language.
- Participation: Inclusion of language communities in technology design.
- Data Sovereignty: Community governance over linguistic data [3].

The implementation of these principles needs governance structures in place. For example:

- Researchers such as [38] suggest that involving communities in dataset creation and model evaluation helps to avoid extractive practices through Participatory Design Frameworks.
- Ethical Impact Assessments: Inspired by [9] “datasheets for datasets,” such tools could evaluate NLP systems’ societal impacts on marginalised languages.
- The Local Contexts initiative (2019) has implemented traditional knowledge licenses that grant Indigenous communities the authority to control how their data is used, and this model can be adapted for NLP use.

By framing resource asymmetry as a justice issue, the NLP community can align technical advancements with equity and self-determination. The direction for change comes from ethical frameworks, yet practical strategies must be implemented to address real-world limitations while upholding equity principles.

## 12. Implementation Challenges and Mitigation Strategies

Multiple obstacles prevent the execution of proposed solutions to address resource asymmetry:

**1) Computational Resource Disparities:** Institutions located in areas where few people speak the language do not have enough computational equipment. A 2023 Stanford HAI survey reveals AI research laboratories in Africa possess only

1/18th of the computational resources that European institutions maintain at comparable levels. The large difference between computing resources prevents even perfectly designed models from being trained locally.

- **Mitigation:** Cloud computing grants and regional computing centres, as demonstrated by the successful NVIDIA-funded AI research centre at Makerere University.

**2) Funding Sustainability:** The periodic funding patterns lead institutions to experience fluctuating patterns of resource development. The Masakhane initiative demonstrated [18] that it experienced difficulties in sustaining its operations after the initial grant funding period because 40% of projects became inactive after the first year.

- **Mitigation:** Public and private funding sources should combine with community resources to develop sustainable local revenue generation systems.

**3) Expertise Circulation:** The departure of trained local experts through brain drain causes disruptions to the process of building capacity. Research conducted by [40] showed that 79% of African AI researchers currently work at institutions located outside Africa.

- **Mitigation:** Developing local research facilities with remote working solutions will attract scientists while preserving their presence in the region.

**4) Data Ownership Tension:** Data collection commercial interests frequently oppose community sovereignty, especially when the languages in question are indigenous. Reference [3] showed that community-developed corpora suffered from unauthorised appropriation that deprived their authors of proper recognition and profit distribution.

- **Mitigation:** Formal data sharing agreements with explicit terms for attribution, usage limitations, and benefit-sharing.

**5) Evaluation Misalignment:** The current evaluation methods fail to measure language-specific accomplishments because standard benchmarks do not suit these particular languages. Research by [5] demonstrated that 78% of benchmark tasks designed for low-resource languages consisted of direct English translations instead of culturally relevant evaluation methods.

- **Mitigation:** Local communities need to take charge of developing evaluation frameworks to create task definitions which match how their languages operate in real-life situations.

A combination of technical and social, and institutional approaches will address the current barriers to success. The Estonian case demonstrates how long-term policy commitment can overcome many of these barriers, while the Masakhane initiative shows how distributed collaboration models can build resilience against individual points of failure.

## 13. Conclusions: Towards Equitable Language Technology Futures

Resource asymmetry in multilingual NLP represents a multifaceted challenge that

spans technical, economic, social, and ethical dimensions. This comprehensive analysis has examined the historical evolution, contributing factors, current approaches, and ethical implications of language resource disparities. Several key insights emerge from this examination.

- **Beyond Technical Solutions:** Historical patterns show that technological advances alone fail to resolve resource disparities and may exacerbate inequities without deliberate equity-focused interventions [1].
- **Economic Drivers:** Market dynamics systematically prioritise dominant languages, necessitating alternative funding models to support underrepresented languages [14]-[23].
- **Interdisciplinary Context:** Resource asymmetry is intertwined with sociolinguistic hierarchies and educational inequities, demanding collaboration with fields like anthropology and pedagogy [4].
- **Measurement Frameworks:** Granular metrics (e.g., data availability, tool coverage, community participation) enable targeted interventions.
- **Case Study Lessons:** Context-specific strategies such as community partnerships [44] or policy-driven initiatives [18] demonstrate replicable success models.
- **Ethical Imperative:** Language technologies mediate access to critical services, making equity a justice issue [38] [39].

#### **Priority Actions for the NLP Community:**

- **Governance:** Adoption of participatory frameworks [40] to involve communities in research prioritisation. Additionally, adopt the Masakhane model of community review boards for dataset approval.
- **Funding:** Develop sustainable models (e.g., public-private partnerships) for long-term support [1].
- **Capacity Building:** Expand training initiatives (e.g., Masakhane) for underrepresented languages.
- **Evaluation:** Design metrics reflecting community needs, not just technical benchmarks [41].
- **Collaboration:** Integrate linguistic and cultural expertise into NLP pipelines.

By addressing these dimensions holistically, the field can advance linguistic equity and sustain global diversity.

## **Acknowledgements**

Doyin Akindotuni thanks Dr. Kakia Chatsiou and Adnan Ez-zizi for their invaluable guidance and mentorship during my postgraduate thesis, which laid the foundation for my research journey. I am grateful to the University of Suffolk for their support and encouragement throughout my program, fostering an environment that sparked my passion for research.

I also extend my appreciation to Rob Davidson, my former colleague at Signify, for his insightful nudge toward the field of Natural Language Processing, which ultimately shaped the direction of this work. His encouragement proved instru-

mental in my exploration of this fascinating research area.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- [1] Joshi, P., Santy, S., Budhiraja, A., Bali, K. and Choudhury, M. (2020) The State and Fate of Linguistic Diversity and Inclusion in the NLP World. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6-8 July 2020, 6282-6293. <https://doi.org/10.18653/v1/2020.acl-main.560>
- [2] Mielke, S.J., Cotterell, R., Gorman, K., Roark, B. and Eisner, J. (2019) What Kind of Language Is Hard to Language-Model? *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, 28 July-2 August 2019, 4975-4989. <https://doi.org/10.18653/v1/p19-1491>
- [3] Bird, S. (2020) Decolonising Speech and Language Technology. *Proceedings of COLING 2020*, Barcelona, 8-13 December 2020, 3504-3519. <https://aclanthology.org/2020.coling-main.313/>
- [4] Blasi, D., Anastasopoulos, A. and Neubig, G. (2022) Systematic Inequalities in Language Technology Performance across the World's Languages. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Dublin, 22-27 May 2022, 5486-5505. <https://doi.org/10.18653/v1/2022.acl-long.376>
- [5] Ruder, S., Constant, N., Botha, J., Siddhant, A., Firat, O., Fu, J., *et al.* (2021) XTREME-R: Towards More Challenging and Nuanced Multilingual Evaluation. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Punta Cana, 7-11 November 2021, 10215-10245. <https://doi.org/10.18653/v1/2021.emnlp-main.802>
- [6] Costa-Jussa, M.R., *et al.* (2022) No Language Left Behind: Scaling Human-Centered Machine Translation. arXiv: 2207.04672. <https://doi.org/10.48550/arXiv.2207.04672>
- [7] Mufwene, S.S. (2017) Language Vitality: The Weak Theoretical Underpinnings of What Can Be an Exciting Research Area. *Language*, **93**, e202-e223. <https://doi.org/10.1353/lan.2017.0065>
- [8] Collart, A. (2024) A Decade of Language Processing Research: Which Place for Linguistic Diversity? *Glossa Psycholinguistics*, **3**, Article 11. <https://doi.org/10.5070/g60111432>
- [9] Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J.W., Wallach, H., H.D., *et al.* (2021) Datasheets for Datasets. *Communications of the ACM*, **64**, 86-92. <https://doi.org/10.1145/3458723>
- [10] Kunchukuttan, A., Kakwani, D., Golla, S., Gokul, N.C., Bhattacharyya, A., Khapra, M.M., & Kumar, P. (2020) AI4Bharat-IndicNLP Corpus: Monolingual Corpora and Word Embeddings for Indic Languages. arXiv: 2005.00085. <https://arxiv.org/abs/2005.00085>
- [11] Wu, S. and Dredze, M. (2020) Are All Languages Created Equal in Multilingual Bert? *Proceedings of the 5th Workshop on Representation Learning for NLP*, 9 July 2020, 120-130. <https://doi.org/10.18653/v1/2020.repl4nlp-1.16>
- [12] Conneau, A., *et al.* (2020) Unsupervised Cross-Lingual Representation Learning at Scale. *Proceedings of ACL 2020*, 5-10 July 2020, 8440-8451. <https://aclanthology.org/2020.acl-main.747/>

- [13] Lin, X.V., *et al.* (2022) Few-Shot Learning with Multilingual Generative Language Models. *Proceedings of EMNLP 2022*, Abu Dhabi, 7-11 December 2022, 9019-9052. <https://aclanthology.org/2022.emnlp-main.616/>
- [14] Bender, E.M. (2019) The #BenderRule: On Naming the Languages We Study and Why It Matters. The Gradient. <https://thegradient.pub/the-benderrule-on-naming-the-languages-we-study-and-why-it-matters/>
- [15] Kreutzer, J., Caswell, I., Wang, L., Wahab, A., van Esch, D., Ulzii-Orshikh, N., *et al.* (2022) Quality at a Glance: An Audit of Web-Crawled Multilingual Datasets. *Transactions of the Association for Computational Linguistics*, **10**, 50-72. <https://doi.org/10.1162/tacl.a.00447>
- [16] Wu, C.J., *et al.* (2021) Sustainable AI: Environmental Implications, Challenges, and Opportunities. arXiv: 2111.00364.
- [17] Rehm, G., *et al.* (2020) European Language Grid: An Overview. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, Marseille, 11-16 May 2020, 3366-3380. <https://aclanthology.org/2020.lrec-1.413/>
- [18] Nekoto, W., *et al.* (2020) Participatory Research for Low-Resource Machine Translation: A Case Study in African Languages. *Findings of EMNLP 2020*, 16-20 November 2020, 2144-2160. <https://aclanthology.org/2020.findings-emnlp.195>
- [19] Scao, T.L., *et al.* (2022) BLOOM: A 176B-Parameter Open-Access Multilingual Language Model. arXiv: 2211.05100.
- [20] (2016) Tools and Resources for CEF Automated Translation. <https://digital-strategy.ec.europa.eu/en/funding/tools-and-resources-cef-automated-translation>
- [21] Martin, G., Mswahili, M.E., Jeong, Y. and Young-Seob, J. (2022) SwahBERT: Language Model of Swahili. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Seattle, 10-15 July 2022, 303-313. <https://doi.org/10.18653/v1/2022.naacl-main.23>
- [22] Rehm, G. and Uszkoreit, H. (2012) META-NET. In: Rehm, G. and Uszkoreit, H., Eds., *The Polish Language in the Digital Age*, Springer, 35-35. [https://doi.org/10.1007/978-3-642-30811-6\\_5](https://doi.org/10.1007/978-3-642-30811-6_5)
- [23] Kornai, A. (2013) Digital Language Death. *PLOS ONE*, **8**, e77056. <https://doi.org/10.1371/journal.pone.0077056>
- [24] Zaugg, I.A., Hossain, A. and Molloy, B. (2022) Digitally-Disadvantaged Languages. *Internet Policy Review*, **11**, 1-11. <https://doi.org/10.14763/2022.2.1654>
- [25] Agrawal, R., Biswas, G. and Hall, M. (2020) Intelligent Tutoring Systems for Second Language Acquisition: A Systematic Review and Meta-Analysis. *Computers & Education*, **144**, Article ID: 103707.
- [26] Pfeiffer, J., Vulić, I., Gurevych, I. and Ruder, S. (2021) Unks Everywhere: Adapting Multilingual Language Models to New Scripts. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Punta Cana, 7-11 November 2021, 10186-10203. <https://doi.org/10.18653/v1/2021.emnlp-main.800>
- [27] Artetxe, M., Ruder, S. and Yogatama, D. (2020) On the Cross-Lingual Transferability of Monolingual Representations. *Proceedings of ACL 2020*, 5-10 July 2020, 4623-4637. <https://aclanthology.org/2020.acl-main.421/>
- [28] Lauscher, A., Ravishankar, V., Vulić, I. and Glavaš, G. (2020) From Zero to Hero: On the Limitations of Zero-Shot Language Transfer with Multilingual Transformers.

- Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 16-20 November 2020, 4482-4499.  
<https://doi.org/10.18653/v1/2020.emnlp-main.363>
- [29] Chen, X., Sun, Y., Athiwaratkun, B., Cardie, C. and Weinberger, K. (2018) Adversarial Deep Averaging Networks for Cross-Lingual Sentiment Classification. *Transactions of the Association for Computational Linguistics*, **6**, 557-570.  
[https://doi.org/10.1162/tacl\\_a\\_00039](https://doi.org/10.1162/tacl_a_00039)
- [30] Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H.W., Sutton, C., Gehrmann, S. and Schuh, P. (2023) Palm: Scaling Language Modeling with Pathways. *Journal of Machine Learning Research*, **24**, 1-113.
- [31] Qin, L., Chen, Q., Wei, F., Huang, S. and Che, W. (2023) Cross-Lingual Prompting: Improving Zero-Shot Chain-Of-Thought Reasoning across Languages. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, Singapore, 6-10 December 2023, 2695-2709.  
<https://doi.org/10.18653/v1/2023.emnlp-main.163>
- [32] Latif, E., Fang, L., Ma, P. and Zhai, X. (2024) Knowledge Distillation of LLMs for Automatic Scoring of Science Assessments. In: Olney, A.M., Chounta, I.A., Liu, Z., Santos, O.C. and Bittencourt, I.I., Eds., *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky. AIED2024*, Springer, 166-174. [https://doi.org/10.1007/978-3-031-64312-5\\_20](https://doi.org/10.1007/978-3-031-64312-5_20)
- [33] Nie, E., Liang, S., Schmid, H. and Schütze, H. (2023) Cross-Lingual Retrieval Augmented Prompt for Low-Resource Languages. *Findings of the Association for Computational Linguistics: ACL 2023*, Toronto, 9-14 July 2023, 8320-8340.
- [34] Bender, E.M. (2011) On Achieving and Evaluating Language-Independence in NLP. *Linguistic Issues in Language Technology*, **6**, 1-26.  
<https://doi.org/10.33011/lilt.v6i.1239>
- [35] Ponti, E.M., O’Horan, H., Berzak, Y., Vulić, I., Reichart, R., Poibeau, T., *et al.* (2019) Modeling Language Variation and Universals: A Survey on Typological Linguistics for Natural Language Processing. *Computational Linguistics*, **45**, 559-601.  
[https://doi.org/10.1162/coli\\_a\\_00357](https://doi.org/10.1162/coli_a_00357)
- [36] Carpuat, M., Marton, Y. and Habash, N. (2010) Improving Arabic-to-English Statistical Machine Translation by Reordering Post-Verbal Subjects for Alignment. *Proceedings of the ACL 2010 Conference Short Papers*, Uppsala, 11-16 July 2010, 178-183.
- [37] Bender, E.M. and Friedman, B. (2018) Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics*, **6**, 587-604.  
[https://doi.org/10.1162/tacl\\_a\\_00041](https://doi.org/10.1162/tacl_a_00041)
- [38] Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R. and Bao, M. (2022) The Values Encoded in Machine Learning Research. *2022 ACM Conference on Fairness, Accountability and Transparency*, Seoul, 21-24 June 2022, 173-184.  
<https://doi.org/10.1145/3531146.3533083>
- [39] Mohamed, S., Png, M. and Isaac, W. (2020) Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, **33**, 659-684. <https://doi.org/10.1007/s13347-020-00405-8>
- [40] Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T. and Prabhakaran, V. (2021) Re-imagining Algorithmic Fairness in India and Beyond. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 3-10 March 2021,

- 
- 315-328. <https://doi.org/10.1145/3442188.3445896>
- [41] Blodgett, S.L., Barocas, S., Daumé III, H. and Wallach, H. (2020) Language (Technology) Is Power: A Critical Survey of “Bias” in NLP. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5-10 July 2020, 5454-5476. <https://doi.org/10.18653/v1/2020.acl-main.485>
- [42] Skutnabb-Kangas, T. (2000) *Linguistic Genocide in Education—Or Worldwide Diversity and Human Rights?* Routledge. <https://doi.org/10.4324/9781410605191>
- [43] Danet, B. and Herring, S.C. (2007) *The Multilingual Internet: Language, Culture, and Communication Online*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195304794.001.0001>
- [44] Adelani, D., Alabi, J., Fan, A., Kreutzer, J., Shen, X., Reid, M., *et al.* (2022) A Few Thousand Translations Go a Long Way! Leveraging Pre-Trained Models for African News Translation. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Seattle, 10-15 July 2022, 3053-3070. <https://doi.org/10.18653/v1/2022.naacl-main.223>