Participatory Detection of Language Barriers towards Multilingual Sustainability(ies) in Africa

To cite this version:

HAL Id: halshs-03713876
https://shs.hal.science/halshs-03713876
Submitted on 13 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Review

Participatory Detection of Language Barriers towards Multilingual Sustainability(ies) in Africa

Gabriela Litre 1,2,3,* , Fabrice Hirsch 1, Patrick Caron 4,5, Alexander Andrason 6, Nathalie Bonnardel 7, Valerie Fointiat 8, Wilhelmina Onyothi Nekoto 9, Jade Abbott 9, Cristiana Dobre 3,10, Juliana Dalboni 3,10, Agnès Steuckardt 1, Cristiana Luxardo 1 and Hervé Bohbot 1

1. Introduction

Sustainability science is written in global languages [1] and does not “speak” most of the around 7000 indigenous languages spread around the world. This factor is one often disregarded element that can contribute to explain why, despite the decades-long efforts of sustainability science and related policy and action programs, humanity has not gotten closer to sustainability at a global scale [2]. Global development agendas, such as the 17 United Nations Sustainable Development Goals (SDGs), may be getting lost in translation, from their initial formulation to their final implementation. Sustainability science does not “speak” most of the 2000 languages from Africa, where the lack of indigenous terminology hinders global efforts such as the COVID-19 pandemic fight. Sociolinguistics, social psychology, cognitive ergonomics, geography, environmental sciences, and artificial intelligence (AI) are all relevant disciplinary fields to uncover the “foreign language effect” that hinders the implementation of the SDGs in Africa. We make the case for detecting and addressing language barriers towards multilingual sustainability in Africa by (1) exploring the “foreign language effect” among African decision-makers and recognising their alternative social representations about sustainability; and (2) detecting Western language stereotypes about sustainability. We propose rethinking SDG-related scientific notions through participatory natural language processing (NLP) and the study of African social representations of sustainability, thus enabling a more inclusive and efficient approach to “sustainability(ies)”.

Keywords: sustainable development goals; multilingualism; Africa; foreign language effect; social psychology; sociolinguistics; natural language processing

Abstract: After decades of political, economic, and scientific efforts, humanity has not gotten any closer to global sustainability. With less than a decade to reach the UN Sustainable Development Goals (SDGs) deadline of the 2030 Agenda, we show that global development agendas may be getting lost in translation, from their initial formulation to their final implementation. Sustainability science does not “speak” most of the 2000 languages from Africa, where the lack of indigenous terminology hinders global efforts such as the COVID-19 pandemic fight. Sociolinguistics, social psychology, cognitive ergonomics, geography, environmental sciences, and artificial intelligence (AI) are all relevant disciplinary fields to uncover the “foreign language effect” that hinders the implementation of the SDGs in Africa. We make the case for detecting and addressing language barriers towards multilingual sustainability in Africa by (1) exploring the “foreign language effect” among African decision-makers and recognising their alternative social representations about sustainability; and (2) detecting Western language stereotypes about sustainability. We propose rethinking SDG-related scientific notions through participatory natural language processing (NLP) and the study of African social representations of sustainability, thus enabling a more inclusive and efficient approach to “sustainability(ies)”.

Keywords: sustainable development goals; multilingualism; Africa; foreign language effect; social psychology; sociolinguistics; natural language processing
so by exploring how the use of foreign languages (that we define as languages learned later in life), as well as the imposition of Western social representations, impact African sustainability decision-making.

In terms of global languages, Ammon and Coupland have considered that what determines whether a language is global is its use for global communication, particularly between people who do not share it as a native language, with use as a lingua franca—that is, communication in which only a few of the participants are native—carrying the most weight. English is, for instance, the lingua franca of science. The authors point out a set of “globality indicators”, developed to determine whether a language can be defined as a global language [1]. The number of non-native speakers is the most important of these indicators, which is why we use the term “global languages” as a synonym for foreign languages (languages learned later in life) in our article. Another factor for globality is the number of native speakers. While this factor is not a definitional criterion, it experimentally does correlate positively with globality and may indirectly impact it by making the language more appealing. Economic strength (measured as the GDP of native speakers), the number of nations that use the language as an official language, as well as the geographical distribution of those countries, international business use, and prevalence in scientific publications and on the internet are additional globality indicators.

Research on the formally or informally imposed use of these global languages can contribute to a more complete understanding of the dynamics of science–policy interaction at different scales. For example, sociolinguistics and social psychology would allow the exploration of how foreign languages may alter moral judgments about sustainable development, as well as risk perceptions. It would also shed some light on the stereotypes and biases embedded in the North-Western concept of “sustainability”, as it is currently formulated in English and French. In fact, many of the terms used in political discourse are controversial or submerged in ambiguity, even in the Northern Hemisphere [3] and need to be rethought. Besides contributing to giving visibility to the foreign language effect on decision-making, we explore some innovative solutions to improve dialogue in different languages through scientific terminology creation (when very common scientific words, such as “fossil” or “virus” do not exist in the native language). More concretely, we analyse how innovative NLP initiatives can change their role from widening the knowledge gap to democratising knowledge through participation and collaborative translation of sustainability scientific articles. Participatory, grassroots NLP can accelerate and scale up the translation of sustainability science into low-resource languages (languages that lack large monolingual or parallel corpora and/or manually crafted linguistic structures to facilitate translation, as most indigenous languages from the Global South are) [4,5].

Finally, we illustrate the need to go beyond the “information gap” model of generating more sustainability scientific data with no particular attention to its legitimacy, ownership and actionability among local actors. In effect, Western sustainability researchers frequently show confidence about the relevance and the actionability of their research while attributing most of the responsibility of global inertia to external, structural factors such as institutional fragility or corruption, or to the alleged insensitivity of local decision-makers [6]. Rather to the contrary, our participation in knowledge co-production processes for climate change adaptation in the Brazilian Amazon and Semiarid biomes [7], as well as our interaction with citizen scientists (or community researchers) in those biomes, shows that development goals will only be successfully implemented if co-created with local actors [8]. In all cases, the lack of terminology to horizontally exchange sustainability solutions [9] remains a challenge.

2. This Study

We argue that the UN SDGs use several Western-oriented definitions of science, sustainability, development, society, and nature, among others, which, in some cases, conflict with indigenous narratives and social representations [10]. New studies are needed to shorten, not increase, distances between worldviews to achieve global sustainability. As stated by Bruno Latour [11], it is dangerous to establish sharp oppositions between
people who impose modernisation and those who resist it. We focus our discussion on Africa for several reasons. To start with, African values, where being itself is a relational category of recognition, rights, and responsibilities, cut across cultures and are frequently shared by both Western and non-Western societies. Those values can become a bridge to a more fruitful dialogue [12]. Secondly, given Africa’s ancestral multicultural and multilingual tradition, translation is not new in Africa: it existed before colonial agents arrived [13] and practised after they left [14–18]. Based on this fertile multilingual tradition, sustainability terminology creation could benefit from cross-evaluations in a “wiki” style. There are already successful terminology creation experiences in participatory workshops in other continents [19], showing that when translators identify a scientific term that they feel they cannot translate (especially in oral languages), terminology experts can handle these separately, facilitating contact with scientific paper authors to disambiguate complex terminology. We suggest that this exercise can be replicated through interdisciplinary teams and also transdisciplinary, that is, in collaboration with sustainable development and humanitarian aid workers who may find it difficult to operationalise and translate concepts in indigenous languages. Last but not least, Africa is pioneering in the development of grassroots NLP research communities such as Masakhane (https://www.masakhane.io/ongoing-projects/masakhane-mt-decolonise-science (accessed on 22 April 2022). Which, based on African values, uses the principles of horizontality, openness, and solidarity to allow researchers to improve translation, terminology creation, and other NLP processes by iterating faster and more effectively.

In this context, our study introduces and discusses the novel concept of “multilingual sustainability(ies)” in Africa through which we seek to show the relevance of multilingualism in broader social and political debates, while bringing realism vis-a-vis the grand expectations generated by the SDGs [20,21].

We share our collective reflections as an interdisciplinary team of scholars who analyse how to better detect and address language barriers to sustainable transformations through: (1) the exploration of the “foreign language effect” among African decision-makers; (2) the identification of existing Western language stereotypes about the concept and meaning of sustainability through participatory, African led NLP; (3) the creation of scientific terminology in indigenous languages to allow meaningful and updated discussions about urgent socio-environmental challenges, including global equity, sustainable development and the health crisis.

Our discussion on multilingual sustainability(ies) revolves around the following three hypotheses: (1) the use of global languages in sustainability agendas in Africa alters moral judgements and risk perceptions while increasing psychological distance among local actors. In other words, Western sustainability science and norms are seldom “psychologically salient” to engage local actors in global principles and inspire action; (2) global languages, which are of mainstream use in global sustainability narratives, frequently embed North-Western biases about key global concepts such as “sustainability” and “development”. Those biases risk replicating intellectual colonialism while disregarding local representations and indigenous knowledge; (3) artificial intelligence (AI), which has mostly ignored thousands of low resources languages, can become part of the solution by scaling up the translation of sustainability science through participatory, grassroots NLP processes.

NLP is an AI field that aims to create human-like communication with computers. NLP “success” is mostly achieved for global languages which have a text corpora of hundreds of millions of words, such as English and a few others. The problem is that those global languages (in which sustainability science is mainly written) represents only about 20 of the world’s 7000 languages [5]. In fact, most human languages are low-resource languages—those that lack large monolingual or parallel corpora and/or manually crafted linguistic structures. They require extra tools and resources to overcome the digital gap and the resource barrier, allowing NLP to provide more inclusive and widespread benefits globally.
Nevertheless, the challenge of promoting multilingual sustainability(ies) goes well beyond translation and NLP. Despite the collaborative process of the SDGs [21,22] and previous efforts to translate the SDGs into African languages [23], most global goals are still crafted in the specialised parlance (and priorities) of North-Western development experts. The consequences of this are numerous, but only recently explored, as we show in the next section.

3. What We Know (Literature Review)

Consistent research shows that the perception of risks (that influences the decisions we make) may be affected by the language in which we learn about them. Depending on the situation, the language we use can increase or decrease our perception of something as dangerous or risky [24]. This may have both positive and negative consequences in the development field. In the health field (SDG 3—health and well-being) for instance, a study by Geipel et al. [25] examined COVID-19 vaccine uptake among Hong-Kong participants. They found increased willingness to vaccinate when the advantages and disadvantages or risks of vaccination were described in a foreign language—English—rather than in their native language, Chinese. Other studies show that certain hazards—such as climate change or even taking a plane—are perceived as less risky and more beneficial when printed in a foreign language rather than in the native one [26–28]. Since communication in a foreign language seems to promote a more positive affective impression of a hazard, this may decrease the impetus toward corrective measures [29]. If confirmed at a larger scale, this specific “foreign language effect” on risk perceptions and decision-making can be relevant to the study of (in)action towards sustainable transformations. Orality also plays a role in risk perceptions, and this is especially relevant for the 700 million illiterate people worldwide. Research results suggest that people perceive novel technologies as less risky/more beneficial when they hear about their advantages and disadvantages than when they read about them. In other words, the modality in which information is conveyed—orally or via printed text—could also affect judgment, decision, and behaviour [25].

Not only languages, but also communication styles make a difference. Even if the importance of generating empathy and trust with partners may sound obvious, many international organisations and donors seldom understand African communication styles, negatively impacting on the “self-esteem” and assertiveness of African decision-makers. Gulere [30] mentions the example of the Basoga people from Eastern Uganda, who systematically manifest a lack of access to sustainable solutions and development opportunities due to existing stereotypes among Western donors about both their language, Lusaga, and their communication style. Gulere describes how foreigners frequently define the Basoga people’s way of communicating as either “belligerent” (“abempwitu”) or “insincere” due to their use, as a sign of politeness, of riddling [31], metaphors, proverbs, or indirect expressions instead of talking “straight to the point” [30]. Many groups speaking an indigenous language that does not have an official status end up being excluded from participating in project design and providing feedback on the performance of aid workers. These issues are well known among international cooperation agencies, especially after Oxfam was accused of covering up claims that staff had sexually exploited victims of the 2010 earthquake. The Oxfam sexual abuse scandal showed the difficulties of the most vulnerable to express themselves and be understood in their indigenous languages and/or communication styles [32].

Laboratory experiences confirm that even when the content of a message is the same, engagement and decisions depend on whether the information is delivered in a native or foreign language [33–35]. This contradicts the common-sense assumption, also prevalent in sustainability science and development, that if people understand their options, their choices should be language-independent [33]. Indeed, social psychology and cognitive linguistics show that using a foreign language impacts our inferences, respect for norms, honesty, and morality while leading to reduced emotion, psychological distance, increased deliberation to maximise profit, and utilitarianism [29,35,36].
Regarding SDG 4 (inclusive and equitable quality education), research confirms that performance in science, technology, engineering, and mathematics increases if students are taught in their native languages [37], a fact that also has critical implications for African youth [34,38]. Despite punctual success experiences, the statistics from the African Youth Survey 2022, carried out in 15 countries among more than 4500 young Africans aged 18 to 24, are discouraging. The survey shows that 52% of young Africans were contemplating emigrating in the next few years. Most young Africans interviewed in a prior edition of the same survey (done before the COVID-19 pandemic) had stated the contrary: they did not want to migrate, and they expressed they wanted to stay in their home country to make a life for themselves there. But in the 2022 survey, more than half of young Africans state that they just want to emigrate, citing the lack of quality education, as well as economic difficulties and a lack of jobs as the main reasons to leave their countries. It is in the global interest to provide equal chances for all Africans, and especially to youth, to develop their capacities and achieve their dreams in their countries, especially as they will account for 42% of the world’s young people by 2030 [39]. While global goals may positively orient action in the long term, “In the long run, we are all dead”, as Keynes said:

“The long run is a misleading guide to current affairs. In the long run we are all dead. Economists set themselves too easy, too useless a task if in tempestuous seasons they can only tell us that when the storm is past the ocean is flat again”. [40]

In spite of all this evidence, investment in language learning continues to have a low priority among international donors: research done by UK FCDO, formerly DFID, shows that officials assume that NGOs have sufficient language capacity to communicate with local communities, but few NGOs have formally created language learning strategies for their staff [41]. This hinders the possibility for local actors to communicate their priorities in meaningful ways [19].

Last but not least, research shows that people make sense of reality and its challenges depending on their specific narratives and social representations [42–44]. Those narratives include a “confirmation bias”, where people tend to (1) look for information that confirms their own values and beliefs and disregard the ones that conflict with them; and (2) resist changing those values and beliefs [44–46]. A direct consequence of this is that a sustainability narrative is only compelling and engaging if it is relevant to the audience’s beliefs and values [47,48]. At the household level, for instance, research shows that social representations contribute to anticipating sustainability-related decision-making, such as recycling programme dropouts [49].

4. What We Do Not Know (and Why We Should)

Although the literature mentioned above is an invaluable resource, we propose that the following points remain open and/or unaddressed by science and could contribute to shedding light on the hidden factors hindering the articulation of global principles with local priorities in Africa. Questions that remain open and should be explored interdisciplinarily to overcome the global language “blah blah blah” that Greta Thunberg complained about [50] include:

4.1. We Still Do Not Know Much about the Foreign Language Effect on Policy Implementation

Science and international organisations have largely attributed slowness in policy implementation to external, structural barriers including the lack of information transparency; lack of financial/technical resources; and institutional/governance and legal fragilities [6,51]; as well as limited human resources [52]. Social psychology, ergonomics, and sociolinguistics can contribute to explore the foreign language effect on decision-making as one key still often disregarded intrinsic factor blocking effective policy implementation at different levels, from local agents to international organisations.

One open question is whether the use of foreign languages reduces or increases moral concerns when deciding on socio-environmental dilemmas. This issue remains controversial. Most studies seeking to respond to this question use experiments involving sacrificial
dilemmas, including the question “Would you kill one to save five lives?” that pit a deontological action, for example, “you should not violate the rights of an individual” with a utilitarian action, including “you should do what’s best for the majority”. These studies show that the use of a foreign language promotes utilitarian moral judgments, which, in this context, have been associated with analytic reasoning rather than intuitive/emotional reasoning, something known as the “dual process theory for moral judgment” [53]. But the connection between increased utilitarian judgment and reduced moral engagement and their impact on decision-making need to be further explored.

Other studies use vignettes that involve different types of violations such as purity violations (e.g., cutting one’s national flag into small pieces and using the pieces to clean the house, eating one’s dog after the dog was killed in an accident) [54] and found that presenting the vignettes in a foreign language increased the permissibility of these actions. The condemnation of such actions is attributed to intuitive processes, and so these results suggest that foreign language use promotes less intuitive moral judgments [55]. In the same line, societal infractions such as “Cut in line when in a hurry” and “Sell someone a defective car” use increased judgments of moral permissibility for such actions when formulated in a foreign language [54].

These findings may suggest that speaking a foreign language separates people from sociocultural–moral norms, maybe because such norms were learned in a linguistic setting in which the native language was spoken instead of, for instance, a classroom, and so a native language prompts them more firmly. A study that supports this viewpoint found that when scenarios featuring bad-luck or good-luck aspects are presented in a foreign language, both good-luck and bad-luck superstitions had less power [24]. Superstitious beliefs and behaviours, like moral norms, are learned in a linguistic setting where the original, native language is spoken.

On the other hand, research employing scenarios that pitted a self-serving action (e.g., failing to offer a recommendation for a rival, thus gaining fellowship) against an ethical action (e.g., failing to deliver a recommendation for a rival, thus winning fellowship (e.g., deliver the recommendation for the rival) confirmed that the use of a foreign language was observed to promote actions linked with self-profit [56]. This seems to occur because making the proper decision is linked to emotional moral cues elicited by the story. The notion is that simply digesting information in a foreign language causes emotional signals—people become upset while reading literature in a foreign language—and as a result, emotional signals connected to an action’s ethicality are lost/masked.

The issue of honesty [57–59] in foreign languages also remains controversial. For example, in several experiments, participants were given tasks that allowed them to earn more money by lying (e.g., they rolled a die and would get some reward based on what they rolled; participants rolled the die in private, which gave them the opportunity to lie to receive more money). Bereby-Mayer and colleagues [59] discovered that when the game was played in a foreign language, participants were more honest (or had more difficulty with lying) in reporting the outcome. In short, only recently have we started to discover more about how using a foreign language influences honesty [57–59], utilitarian actions, and moral engagement, and there is much room to explore how this is linked to concrete decision-making in the sustainability and development fields.

4.2. Alternative Social Representations and Narratives about Sustainability(ies) Remain Largely Ignored by Global Decision-Makers and EVEN researchers

Much of the research on narratives (including sustainability narratives) has focused on their internal structure, while not sufficiently considering their social nature [60,61]. Exploring the social nature of sustainability narratives and their relationship with indigenous social representations [62] could contribute to developing and testing what we propose to call multilingual sustainability(ies): a bottom-up (re) conceptualisation of sustainability goals in the light of indigenous epistemologies, values, and social representations, as reflected in their local languages. In that regard, the concept of multilingual sustainability(ies)
could open the way to the recognition and value of African epistemologies, like Ubuntu, in which economy and human beings cannot be separated from what has been referred to as “the environment” [12].

As put into evidence by the COFOR or “Common Frame of Reference” model [63] proposed in cognitive ergonomics, some shared knowledge and mental representations are necessary to construct a “common frame of reference” (or “operational referent”) and, on this basis, to develop cooperation between different stakeholders to achieve a shared objective, here, global sustainability.

More precisely, the COFOR model points out which conditions are required, at a meta-cooperation level, for actors involved in exchanging information with mental representations and thus construct an “operational referent”. On this basis, a common communication code appears as necessary, as well as a compatible knowledge system and a shared model of the tasks to be performed. In all cases, stakeholders need a shared representation of what they intend to achieve. This issue becomes difficult when not only languages but also concepts are not shared. Just one of many examples: only naturalist (Western) society describes a boundary between self and other, based on a dichotomy between nature and culture that does not exist in most indigenous communities [64–66]. This Western distinction, a result of a particular history and trajectory, reflected, for instance, in European dictionaries currently studied diachronically (Bohbot et al. [67]), is non-existent in many other societies [68,69] and need to be further investigated.

4.3. Western Language Biases and Embedded Stereotypes about Sustainability Need to Be Better Acknowledged

The fact that most sustainability agendas, including the UN SDGs, which have been conceived in English, has strong epistemological consequences. As Ngugi wa Thiong’o [70] reminds us “the choice of language and the use of language is central to people’s definition of themselves with their relation to the natural and social environment, and the entire universe”. Additionally, translations are also inherently political [13] because they “always imply an unstable balance between the power one culture can exert over another” [71]. Indeed, in translation, even apparently “neutral” actions, such as the choice of text, “equivalent” words, and even language, are highly political. In the same line, Christiane Nord [72] wrote that “translation is an intentional interaction intending to change an existing state of affairs. As such, translation carries the translator’s own political views and those of the editor, which includes biases, stereotypes, and power tensions. These biases and power tensions disseminated through translations were first made evident by the role of colonial missionary translations, not only in African politics but also in other former colonies such as India [73], where the missionaries engaged with politics via their translations [13]. The corpora of religious data generated by missionaries (and replicating Western religious views and values) remains, in many cases, the only parallel sources to translate indigenous languages, and is currently affecting NLP in Africa. Detecting the resulting embedded biases in Western sustainability narratives has been mostly done on more minor scales, usually through classical lexicon quantification and subjective, manual content and discourse analysis [74]. As we will see in the next item, NLP and word vectors can contribute to scaling up measuring biases both in international organisations’ discourses and in scientific papers. This allows much-needed reliability, scalability (vast number of internet available data/corpora/texts) and comparability [5,75,76].

4.4. We Need to Understand the Potentialities, but Also the Limitations, of Using Participatory NLP to Detect and Address Language Biases

As Gagnon defines it, any text is political “if it involves power or resistance” or if it “contain[s] some form of power struggle” [77,78]. Language biases condense many of those power struggles. In parallel, the “low-resourced”-ness of indigenous languages in which texts are written is a complex problem beyond data availability and reflects systemic power asymmetries in society. Significant African, Latin American, and Asian populations have spoken local languages for thousands of years, creating rich oral history traditions that
have served communities by bringing alive ancestral stories and historical perspectives and passing down knowledge and morals [79]. Local decision-makers seeking to be legitimised by these populations make every effort to articulate their narratives to their potential voters’ languages and priorities.

On the opposite side, technology does not “speak” most of the around 2000 languages and varieties used by Africans. Research in NLP lacks geographic, gender, and cultural diversity [4,5]. Mobile technologies like Apple’s Siri, Google Assistant, and Amazon’s Alexa collectively use zero African languages [79], while the AI field continues to largely set aside multilingualism among Silicon Valley innovators [80]. The implications of this are not only ethical, but also practical: by not engaging diverse knowledge systems and low resource languages, researchers and practitioners limit their own knowledge and the impact of their work in times of planetary crisis. For example, many key scientific words related to SDG 3 (human health and well-being) do not exist in African languages and during the COVID-19 pandemic, some African governments did not perform health surveys or communicate cutting-edge research in the most widespread languages because of the difficulty in translating words such as “virus” [4,81].

The challenge of the lack of scientific terminology in local languages is, of course, not exclusive of Africa. Departing from a series of case studies from China and the global Chinese diaspora, Piller, Zhang & Li explain how, also during the COVID-19 pandemic, multilingual crisis communication emerged as a global challenge. The authors state that widespread exclusion of linguistic minorities from timely, high-quality information characterizes global public health communication. In other words, the severe limitations of multilingual crisis communication exposed by the COVID-19 crisis were the result of the dominance of English-centric global mass communication, the longstanding devaluation of minority languages, and the failure to recognise the importance of multilingual repertoires in building trust and resilient communities [82]. As solutions, the authors suggest including local knowledges and grassroots practices not only as objects of investigation between sociolinguists, but also in its epistemologies to diversify its knowledge base and the academic voices producing that knowledge base, and to re-enter dialogue with policymakers and activists [82].

To engage diverse epistemologies, and besides creating scientific terminology, one crucial step is to detect and acknowledge current biases in Western sustainability narratives. This can be done through current NLP models which, as the global languages they are trained in, show much more than algorithms. They mirror the historical values, inequities, and prejudices of society. One potentiality of NLP to detect language biases is word embeddings. Word embeddings are an increasingly popular application of neural networks wherein enormous text corpora are taken as input, and words are mapped to a vector in some high dimensional space. These word vector representations estimate the similarity between words based on the context of their nearby text or predict the likelihood of seeing words in another context. Their use is now standard in training complex language models.

At the same time, word embeddings are prone to express the bias inherent in the data it is extracted from [83]. Bolukbasi et al., [75] found out, for example, that the term “man” is very frequently associated with “computer programmer”, while the word “woman” appears to be related to “homemaker”. The use of word vectors is rather new, but its ultimate goal is at least half a century old: Firth [84] already pointed out in the fifties that “a word is characterised by the company it keeps”. Any attentive reader can detect that several mainstream Western notions of “sustainability” appear as systematically related to economic expressions such as “growth”, natural “resource”, or ecosystem “services”—all quantifiable terms [62]. The detection of global language biases can be accelerated and scaled through NLP [85]. Grassroots African NLP research communities such as Masakhane (https://www.masakhane.io/ongoing-projects/masakhane-mt-decolonise-science (accessed on 22 April 2022)), which uses a participatory approach to allow researchers to improve the NLP processes by iterating faster and more effectively [4], can contribute to closing the digital divide [86].
5. Discussion

In this review article, we argue that while indigenous languages are spread around all the continents populated by humans, sustainability goals and solutions keep being formulated in a few global languages. Global goals are being lost in translation all along the way: from their formulation in Western experts’ parlance to their final implementation in the humanitarian and disaster response fields.

Our literature review confirms, in principle, our three initial hypothesis: (1) that Western sustainability science and norms formulated in foreign languages are seldom “psychologically salient” to engage local actors; (2) that global languages, which are of mainstream use in global sustainability narratives, embed North-Western biases which could replicate intellectual colonialism; (3) that AI can become part of the solution by scaling up and accelerating the translation of sustainability science (including through terminology creation) through participatory NLP processes.

We show that, despite Africa’s rich tradition with translation, the multilinguisation of sustainability(ies) concepts and norms continue to be an immense challenge. In effect, global languages continue to dominate sustainability narratives, that are, in turn, restricted in Africa to groups of international scholars with significant degrees of exposition to this mainstream and global “jargon”.

Such global jargon is frequently promoted, through local language policies, as an environment-neutral “operative language” [84], efficient for communication and cooperation among experts in the domain but “hermetical” for non-global experts and insensitive to local cultural differences.

This cultural insensitivity puts much needed language policies into a vacuum, ignoring the complex set of environmental conditions that surround human life. Spolsky argues against this ‘linguicentrism’, or the belief that languages are independent of their environment. Rather to the contrary, Spolsky points out, a number of relevant non-linguistic features—physical geography, which includes characteristics that encourage human settlement such as agricultural and mineral resources as well as those that constrain it such as mountains, seas, and climate, demographic forces such as the intensity and diversity of ethnic settlements and movements, technological progress, events such as modernisation and globalisation, civil and external wars and epidemics, and even climate change, they are all relevant in the design of effective and inclusive language policies [85–88].

In parallel, the publication of African-led scientific research still struggles to reach influential global journals, which usually demand a perfect command of written English, or request professional English revisions of their manuscripts (at prohibiting costs for many Global South authors). In the meantime, Western scientists and international organisations continue by talking more to themselves than the communities they claim to aid, especially under the mandate of the “publish or perish” principle, the need to quantify and measure “impact” to get tenure and professional stability, or to fund new projects. Myriads of international scientific and development projects claiming to implement “bottom-up” processes continue to trivialise effective local participation while offering “perfect” impact reports to donors written in global languages.

In parallel, and in spite of significant efforts towards language inclusivity, many African university structures continue to resist, explicitly or tacitly, most manifestations of diversity [89,90].

As the Masakhane research community [4], among AI initiatives [79], shows, NLP can promote knowledge inclusiveness for specific categories and groups, specifically by detecting language biases and by promoting gender sensitivity, customisation to the needs of impaired users, valorisation of cultural and territorial differences, and communication in local low resource languages.

Initiatives like this can make the UN SDGs, including SDG 16 (peace, justice and strong institutions), target 16.7 (“ensuring responsive, inclusive, participatory, and representative decision-making at all levels”) more realistic [20].
But even if promising, using participatory NLP to promote multilingual sustainability(ies) in Africa also poses significant challenges. Some of them are:

5.1. Sustainability Science Translation and Terminology Creation Is Non-Trivial

There is a risk that translation and terminology creation is too difficult to implement in a place as large, intricate, and diverse as Africa, where, as we saw, around 2000 local languages and variations are spoken (and many of them are unwritten). Additionally, no NLP model can replace human-performed translation and sophistication [91]. Besides this, the “place effect” problem (that is, the limitations of an NLP model trained on translations rooted in one cultural context) risks to fall in the “linguacentrism”, or the belief that languages are independent of their environment, a trait already detected by Spolsky when questioning some language policies [85,88]. Rivero mentions the example of translations from US news archives into low resource languages. US news is full of local references such as American food, customs, and place names. NLP models trained in the US specific context will perform significantly worse when used in a clashing cultural context [92].

We consider that these difficulties need to be systematically registered and addressed, but they should not deter the task from advancing. Success stories exist, as shown by Hasler et al. [19] with East Cree and Innu communities in the Americas and of the San Development Trust in Zimbabwe [93]. The “Africa Arrive” team of Masakhane is tackling this problem by directly contacting scientific paper authors to disambiguate complex scientific terminology. This puts translators, NLP specialists, local languages speakers and scientists on the same discussion table (See, for instance, https://www.masakhane.io/ (accessed on 22 April 2022)).

5.2. The Lack of Technical and Scientific Background of Translators Is Frequent

Translators working in low-resourced languages can not only have problems with linguistic and societal knowledge that would enable them to form effective interactions; they may also lack the necessary scientific and technical resources, knowledge, connections, and incentives [4]. Therefore, participatory NLP needs to involve open dialogue with stakeholders in every part of the scientific process (as, for example, Masakhane is doing through a very active Slack forum) to identify and quickly address problems for low-resourced languages.

5.3. NLP Risks Perpetuating Language Biases

NLP models are trained on data that shows not only language characteristics, but also cultural norms and social representations which are highly diverse. For example, different communities have their own set of taboo issues and gender or ethnic biases [5]. As NLP techniques become more widespread, it is becoming more important to understand how they influence (and are in turn influenced by) history, social prejudices and preconceptions. Despite their success in modelling a variety of applications, NLP models spread and may even magnify gender or ethnic biases in text corpora. For example, models that have learned social representations that are biased against historically disadvantaged groups (based on race, gender, socio-economic status, nationality, or religious stereotypes) can cause a great deal of harm when those biases surface in downstream tasks or applications, such as automatic summarisation or web search [79,85]. The same applies to key sustainability notions such as well-being, health, peace, freedom, quality education, and sustainability, which are, as we saw, language and culture dependent. While the topic of bias in artificial intelligence is not new, strategies for reducing bias in NLP are still in their infancy [89]. Understanding how machine learning approaches store biases and how humans perceive those biases is critical to effectively debiasing at the source.

Interdisciplinarity is unavoidable to tackle language biases, and NLP researchers developing guidelines for bias identification and correction acknowledge this. Bias detection and mitigation requires the integration of disciplines, including not only engineering and computer sciences, but also the social and human sciences [86,89]. AI ethics advisors also
need to get effectively involved along the process. Interdisciplinary (integrating disciplines) and transdisciplinary (fostering dialogue between academic and non-academic actors) approaches are certainly a challenge, but as the experience of co-writing this review article proved, it is not impossible if participants are open to leaving their comfort zones [94,95], if they value collaboration, and if they are curious and solutions oriented.

6. Conclusions

Unless people’s access to the communication networks relevant to decision-making is specifically addressed, participatory discourses will continue to be empty in many Global South regions. Fed by the creation of terminology/reconceptualisation of fundamental sustainability notions through participatory NLP, the novel concept of multilingual sustainability(ies) could be used to better articulate global principles with local priorities and humanitarian and emergency response, increasing the moral engagement and the psychological salience of the SDGs. The recent experience of the COVID-19 pandemic showed that the widespread exclusion of linguistic minorities from timely, high-quality information characterised global public health communication. The severe limitations of multilingual crisis communication exposed by the COVID-19 crisis were the result of the dominance of global languages mass communication, the longstanding devaluation of minority languages, and the failure to recognise the importance of multilingualism in building empathy and trust.

In Africa, grassroots NLP research communities involving academic and non-academic actors have the potential to allow greater scientific impact and local engagement. They can also contribute to building richer and more meaningful NLP datasets that detect embedded biases and stereotypes, while including the priorities of stakeholders.

Much must be done to develop a rigorous theoretical approach toward understanding how the use of global languages and narratives affects sustainability transitions in Africa. The multilingualisation of sustainability(ies) is an intersectional challenge that can only be dealt with through interdisciplinary and transdisciplinary approaches. The integration of different disciplines—e.g., social psychology, cognitive ergonomics, environmental sciences, sociolinguistics, and artificial intelligence’s natural language processing (NLP)—can collaboratively contribute to shed new light on this and on several academic fields of study in which sustainable development remains a contested issue.

Author Contributions: Writing—original draft preparation, G.L. (Gabriela Litre), P.C., A.A. & F.H.; writing—review and editing, G.L. (Gabriela Litre), P.C., A.A., F.H., N.B., V.F., W.O.N., J.A., C.D., J.D., A.S., G.L. (Giancarlo Luxardo) and H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by INCT/Odisseia-Observatory of socio-environmental dynamics: sustainability and adaptation to climate, environmental and demographic changes under the National Institutes of Science and Technology Program (Call INCT—MCTI/CNPq/CAPES/FAPs No. 16/2014), with financial support from Coordination for the Improvement of Higher Education Personnel (Capes): Grant 23038.000776/2017-54; National Council for Scientific and Technological Development (CNPq): Grant 465483/2014-3; Research Support Foundation of the Federal District, (FAP-DF): Grant 193.001.264/2017).


Informed Consent Statement: Not Applicable.

Data Availability Statement: Not Applicable.

Acknowledgments: The authors thank the Montpellier Advanced Knowledge Institute on Transitions (MAKT) of the University of Montpellier and the research units Praxiling (CNRS-UPVM) and SENS (CIRAD) for offering their support and infrastructure to perform this research in France.

Conflicts of Interest: The authors declare no conflict of interest.