Empathetic Machines

An Asimovian Fantasy?

Exploring the Intersection of Language, Technology, and Colonialism

Divith Narendra

Plaksha University

Academic Supervisors

Dr. Brainerd Prince

Dr. Saikat Chakraborty

May 31, 2023

A Critical Examination of Machine Language

and Linguistic Diversity in Artificial Intelligence

Abstract

This paper examines the intersection of language, technology, and colonialism through the lens of contemporary artificial intelligence systems. Drawing on post-structuralist linguistics (Derrida, Kristeva), critical technology theory (Heidegger, Marcuse, Feenberg), and postcolonial studies (Said, Spivak, Bhabha), I argue that current large language models perpetuate linguistic colonialism through their architecture and training paradigms.

The dominance of English in AI training data and the structural failures of encoding systems to represent non-Romance languages reveal not merely technical limitations but the continuation of colonial power structures embedded within computational frameworks. I propose "technolinguistics"—an interdisciplinary approach combining linguistics, philosophy, and computer science—to develop locally contextual, language-family-specific models that preserve linguistic *différance* while respecting cultural and grammatical diversity.

This work contributes to ongoing discussions about decolonizing technology and creating more equitable AI systems that honor the full spectrum of human linguistic expression.

1 Introduction: The Question of Language and Machines

1.1 Defining Language

"Words are just, what? Nothing. Complicated airflow" — Kendall Roy, *Succession*

Language is fundamental to human expression. When *Succession*'s cynical characters pass judgments on their siblings through carefully chosen sentences, exhibiting Macbethian proportions of betrayal through spoken word alone, they demonstrate the profound impact language bears—something deeply human.

To address the pressing question of meaningful dialogue between humans and machines, one must first understand language itself. To the layperson's eyes, language can be attributed to verbal expression, a figment of speech: a curse attributed to anger, a smirk to happiness, a squeak to fear.

These arrangements of words seem resemblant to "complicated airflow," as Kendall aptly describes. Words strung together under an order of Universal Grammar (**?**), an all-encompassing framework for interpreting language, as Chomsky propounds. Derived from *lingua*, referring to the tongue, the examination and evolution of language has been consistent, accelerating further in the 20th century.

1.1.1 Structural Linguistics: Saussure and Chomsky

Friedrich de Saussure proposed a structured approach, considering language as a system—an organizational hierarchy molded through *langue* (**?**). This framework determines language as a consequence of utility through a linguistic community. This unique utility allows language to evolve, requiring its interpretation to be further nuanced. Saussure argues for a structured

outlet for thought, shaped through cultural connotations and linguistic communities.

Chomsky (**?**) extends this framework, referring to the ideals of Universal Grammar, the Language Acquisition Device (LAD), and Deep Structures. He emphasizes a duality of language through his proposal of Deep Structure (semantic) and Surface Structure (linguistic). Understanding a language's true meaning requires deciphering its Deep Structure, central to Chomsky's claim. Through analyzing phonemes and sentence formation, Chomsky develops a concrete axiom for language.

1.1.2 The Ontological Turn: Heidegger's Challenge

Saussure and Chomsky's arguments take an ontological turn with Heidegger (**?**), who proposes a divine attribution to language as "the house of being," associating it with being and truth itself. This presents a challenge: if Heidegger's claim of divinity has merit, it would require forsaking the structural purview Saussure argued for. To be incomprehensible, all in pursuit of truth! Can human beings and their cultural-linguistic interactions shape language, an embodiment of the divine?

We find ourselves with a unique split in linguistics: Structuralism versus Post-Structuralism. While Saussure and Chomsky argue for rule-based approaches, claiming humans are inherently designed to learn language (equipped with a LAD), they suggest language behaves as a medium of pattern recognition, deciphering meaning through association.

1.1.3 Post-Structuralism: Derrida and *Différance*

But is that all there is to language? Jacques Derrida (**?**) contends otherwise. Can language be constituted by speech alone, a primarily Western belief? Language is not structural. A turbulent stream, language is meant to be arbitrary and chaotic—a combination not only of speech but also of writing, a means not of expressing meaning but of expressing expression itself.

Derrida argues for the instability of language. Language might be locally constrained, withdrawing meaning not from pre-defined corpora but from the words that precede or succeed

it. As Derrida states, meaning is perpetually deferred—a concept he terms *différance*.

Consider contextual-dependent meaning in languages such as Sanskrit, German, and Latin, where changes in a word alter meaning altogether—an indication of polysemy. Take the Sanskrit word “” (*Bhava*), which illustrates language’s fluidity through three variations:

Sarvabhūteṣu bhavena yaḥ paśyati.
(One who sees the same essence in all beings)

Rādhā kṛṣṇe bhavena magnā premarasena.
(Radha absorbed in devotion to Krishna)

Vīra bhāvena vibhūṣitaḥ svāgata karoti.
(Adorned with heroic spirit, he welcomes)

Here, *bhava* carries three distinct meanings. The first represents existential essence; the second’s interpretation depends on “Radha” and “Krishna”; the third conveys heroic responsibility. Meaning defers to surrounding phrases—*bhava* cannot be understood in isolation.

Similar patterns appear in German:

- **Geist** as “spirit” or “mind”: In philosophical and religious contexts, describing intellectual or spiritual dimensions.
- **Geist** as “ghost” or “specter”: In folklore and supernatural beliefs, referring to ghostly entities.

This linguistic instability challenges structuralist assumptions. Derrida critiques the Western “logo-centric” worldview that privileges speech over writing. He argues for prioritizing both as equal modes of expression. The dual modes of discourse enable interaction not only between individual and author but also between individual and text (**?**).

1.1.4 The Semiotic Dimension: Kristeva

Julia Kristeva (**?**) builds on Derrida, proposing the semiotic-symbolic divide. Language comprises not only sociolinguistic divisions but emotional ones as well. The semiotic frame-

work argues that language includes an affective component—a rhythmic, bodily framework. Paul Ekman's research on facial expression (**?**) complements Kristeva's claims, expanding language from a restrictive rule-oriented framework to a natural bodily process responding to stimuli.

Revisiting Derrida's arguments on deconstruction and *différance* reveals a human dimension. The inadequacy of structuralist arguments to define fundamental language becomes apparent due to varied cultural and social communities, intrinsic power dynamics, social conditioning, and ideologies pervading language.

Working Definition: We now define language as a human process comprising *différance* (local contextuality for interpretation) and affective being (emotion and expression). Language acknowledges inherent socio-cultural dynamics and social structures. Its meaning arises from both what is expressed and what is withheld.

1.2 A Critique of Technology

"The will to mastery becomes all the more urgent the more technology threatens to slip from human control." — Martin Heidegger, *The Question Concerning Technology*

This statement suggests a Hegelian dialectic between humans and technology. This dialectic requires us to define technology's contextuality, which produces such binary division. To explore merely the lexical interpretation of "technology" would be unjust. Intuitively, we identify technology with a duality similar to language: a contrast between ontological and anthropological (instrumentalized) definitions.

1.2.1 Heidegger's Critique of Instrumental Technology

Anthropologically, Heidegger critiques the claim of technology as merely "a means to an end." Reducing technology to its "essence" in this way is critically inadequate. Technology cannot be understood solely through causal establishment. The Greek conception circumscribes technology as revelation meant to satisfy human needs. Heidegger contests this reduction of

technology to serving human interests.

Heidegger critiques reducing machines and technology into slaves of human interests. These restricted revelations, produced through "enframing" technology within causal entities, render technology crippled. If technology reveals truth, modern technology represents regulated revelation—revelation to satisfy exploitation. This exploitation is enabled by reducing entities, and technology itself, to resources.

This inherent narrative finds itself affecting individuals on a granular level. Capitalism's seeping through unit economics prescribes value to individuals. The ironic reduction of human worth to a commodity implies our pitiful state of subjecting not only technology to commodification and causal establishments but human free will as well.

1.2.2 Marcuse and Feenberg: Technology as Production

Herbert Marcuse (**?**) engages with Heidegger, defining the ontological meaning of *techne* (technology) not purely as revelation but also as means of production. This juxtaposition presents a unique economic interpretation:

> "The oppressive features of technological society are not due to excessive materialism and technicism... [but to] the arrest of materialism and technological rationality in an especially undemocratic, dehumanizing form."

Andrew Feenberg (**?**) supports Marcuse's phenomenological viewpoint. Technology's purview evolves as the beacon shifts between different authors. From restricted revelations to calculated exploitation, it appears the interpretation of technology, not just its utility, has been altered.

Feenberg and Marcuse allow us to examine technology not just ontologically (as Heidegger explains) but also from an angle of production. However, Gadamer allows us to interpret machines hermeneutically. This interplay of ontology, economics, and hermeneutics must be considered while establishing technology. Gadamer proposes technology as embedded in cultural phenomena—an approach enabling us to tackle technology from a humanist's viewpoint.

1.2.3 The Central Question

To expand a definition of technology through Heidegger's viewpoint raises an intriguing question: Are we misinterpreting technology, or are we forced to misinterpret it? This means setting aside history and considering: Is technology a product of capitalist interests, or does capitalism actively reduce technology to guarantee execution of economic interests?

The reductionism established through Heideggerian consideration implies a definition of modern technology that is devoid of emotion, commodified, and self-serving. It guarantees establishment of a resource-oriented society while accepting to suppress or manufacture truth based on societal interests.

This insensitivity, reflected in technology's interpretation, gives rise to Hegel's dialectic (**?**). Let us extrapolate Hegel's ideals, attributing the master-slave dialectic to machines and humans. My capability to establish this attribution within this text provides you an indication—my privileged point as an individual with realized self-consciousness allows me to extrapolate Hegel's dialectic, attributing humanity as master over machines. But for how long? ChatGPT's advent presents a dilemma: Is our responsibility as torch-bearers for self-consciousness threatened? This insecurity, deeply seated within the human mind, drives exploitative, regulated revelations that restrict technology within causality's purview.

Reducing technology to causality's confines limits viewing it ontologically. Here we delve into our problem's heart: the false interpretation of language coupled with reductionist views of technology suppresses us from defining technology anthropologically and renders us incapable of progress.

Feenberg and Marcuse present a unique proposition: the misinterpretation of technology can be subjected to discourse. The rapid evolution of AI and machine language in the 21st century requires engaging with technology not through Hegelian dialectic but through an existentialist viewpoint. There is optimism in re-interpreting technology through local contexts—enforcing democratic rationalization (**?**) by observing technology through community lenses.

However, considering democratic rationalization requires subjecting technology to constraints. This seems counterintuitive, entailing technology remaining interlocked in Aristotelian

causality. To redeem this facet requires determining an entity outside causality's purview—defining a human definition that encompasses both ontological and instrumentalized definitions of technology to achieve democratic rationalization. To subject it to a Derridean definition of language.

2 Colonial Rationality and Machine Language

2.1 Encoding Systems and Linguistic Hegemony

Our arguments reach a crossroads. Popularly used encoding systems—UTF-8 (**?**), ASCII (**?**), and ISO-8859 (**?**)—emphasize linguistic diversity. Considering the plethora of languages these systems cover, one might find them satisfactory. However, multiple instances reveal difficulties encoding Eastern languages: Han Unification issues (**?**), attempts to encode Tulu (**?**) (with breakthroughs in 2022 (**?**)), and Indic scripts pose a critical question: Why are these systems failing?

This requires exploring subjugation. Subjugation takes diverse forms: ideological (Orwellian nightmares), political (erosion of governance), or cultural. Subjugation is driven by narratives (**??**)—narratives engineered through millennia of homogenization and free thought suppression. These narratives emerge not from power alone but from perceived supremacy.

As Catherine Hall (**?**) illustrates in colonialism's context, narratives provide interpretation—not only to the colonized but also to colonizers themselves. A mindset capable of propagating similar narratives ensures uniformity. This uniformity framework assists in reinterpreting binary definitions. Colonizers' disposition produces a coerced unified framework, essentially creating what Spivak (**?**), King (**?**), and Bhabha (**?**) call the "colonial binary"—a dialectic with colonizers at the helm.

This binary is complemented by subjugation: subjugation of thought, culture, and most

importantly, language. This driven subjugation results in gradual erosion of social cohesion, indoctrinating masses to rely on pre-defined frameworks orchestrated by colonizers—frameworks driven by reduction. Reduction of sociocultural diversity into convenient lexicons to enable comprehension (**?**).

2.2 Robbing Languages of *Différance*

Diversity reduction for comprehension is amplified through appropriation and assimilation. Language assimilation through acceptance constitutes reappropriation, emphasizing rationality and uniformity in pursuit of supremacy. This linguistic subjugation fosters modern linguistics' deepest divide.

Subjugating mass populations to colonial grammar and language (English, in our case) actively imposes structuralist frameworks onto post-structuralist societies. Scientific thinking imposition reshaped Indian thought, placing colonized peoples at a loss of expression (**?**). As Spivak contends, languages are reinterpreted primarily by those in power. Regional language linguistic redirection serves as a reinterpreted thought framework dispersed to mass populations, indoctrinating them while imposing linguistic hegemony (**??**).

Linguistic hegemony exhibits evident power dynamics, subjugating regional diversity and expression. From a Derridean view, we infer that colonial languages like English rob regional languages of diverse connotations, multiplicity of meanings, multiple representation modes, and their Derridean *différance*. By subjecting languages to concrete interpretations and canonical transformations into Romanic scripts, we confine them to fixed meanings by consciously editing lexicons for narrative purposes, actively violating their fundamental Derridean nature. Colonizers actively suppressed identity for standardization's sake.

Colonial world interpretation to usher in modernity was deemed imperative. The goal involved not only producing resource-efficient societies but inducing restricted "modern" colonized thought frameworks—limiting processes and linguistic behaviors for convenience and optimization. We observe similarities in "modern" society conformities, garnering negative attention: Han Unification and Indic Encoding. ASCII, designed for English encoding, finds

encoding more than one system impossible.

The world's Euro-centrality produces inherent frameworks dictated by the West, reflected in their systems. Their emphasis on standardizing thought frameworks (itself appropriation's product) resulted in repeated controversies: failures encoding Chinese, Japanese, and Korean characters (**?**), and difficulties developing analyzers for African click languages. These indicate skewed development frameworks, likely revealing colonial language impingement within machines. Identifying this cause is required before further development. Otherwise, we risk encoding languages not in regional contexts but through watered-down colonial interpretive lenses. This requires questioning machine language's very meaning as an entity.

2.3 Data's Colonial Embodiment

Colonial subjugation blockades imposed on machine language through development mainly on vast colonial English corpora essentially impart lexicons and semantics similar to colonial languages. Yet, one might find this inconclusive as analysis scale remains primarily historically limited. Hence, to augment this discussion, we must progressively abstract machine language. To abstract machine language further, consider the language acting as our conduit for machine interaction: English.

Large Language Models are standardized in training terms. OpenAI's GPT-3 (**?**) was primarily trained in English (93% by word count). Other multilingual models like mBERT (**?**) and XLM-R (**?**) source data through Wikipedia and CommonCrawl, with CommonCrawl consisting of roughly 46% English documents. This English unipolarity warrants exploration.

While comprehensive work examining multilingual model performance (**?**) has been conducted, the models introduce a dilemma within Natural Language Processing. The paper analyzes LLM performance primarily on Catalan, a Romance family language similar to Spanish. The paper estimates whether LLMs can transfer existing models built primarily on colonial languages (like English) to different languages.

GPT-3, trained using only around 0.01798% ($\sim$35M words) in Catalan, outperforms expectations. However, Armstrong et al. (**?**) hypothesize cross-lingual transfer possibility due

to overlap between source and target languages. This claim is bolstered as Armengol-Estapé et al.'s (**?**) modeling outperforms expectations primarily for surface-level interactions with target languages sharing similar linguistic lexicons, yet proves erroneous for deeper textual analysis of low-resource languages requiring linguistic clarity.

Thus, Armstrong highlighted a unique contention in Estapé's argument: given language similarity (in Armstrong's case, Jamaican Patois creole derived from English) allows models to process data better. Through this disposition, we consider: have we produced a biased model?

Similar claims are explored by Turc, Lee, & Toutanova (**?**). Turc conducted comprehensive analysis finding languages like German and Russian perform better primarily due to closer English association with predominantly similar morphology. This is verified by Lauscher et al. (**?**), who found multilingual transformers limited in zero-shot language transfer ability.

Our arguments circle to a pivotal idea: are models biased? We seem to think so. Biased not in race, religion, creed, or social structure terms—but these models are empirically devised based on classification through similarity and association lenses, association to colonial languages like English.

2.4 Defining Machine Language

The aforementioned produces unique juxtaposition for consideration. There is provision for machine language interpretation through not only Heideggerian technology lenses. However, this fundamental breakdown requires identifying machine language more empathetically, arguing for machine language definition within post-structuralist Derridean viewpoint confines while retaining Heidegger-Marcuse-Feenberg technological context. To define this requires identifying the source behind data misconfiguration: machine language definition itself.

We must ask an imperative question: What is machine language?

As a good academic, I relied on ChatGPT and similar LLMs. Two interesting answers emerged, which we ought to address before defining our machine language consideration:

a. Machine language refers to instruction sets that specific computer architectures can di-

rectly execute.

b. Machine language is the lowest-level programming language directly communicating with computer hardware.

One observes reductionist technology definition traces circumscribing machine language bounds as we know it. Our technology and language considerations argue for reorganization. Applying Heideggerian technology interpretation to aforementioned definitions leaves us at a loss for words. On deeper levels, we observe machine language notions as technology illustrations to be merely regurgitations of our restrictive truth-suppression acts.

The binary divide presented in machine language through 1s and 0s looks elementary but holds significant value—value frameworks determined only by human beings. To attribute such representational weight to binary frameworks represents immense power and bias to skew meaning. One could argue this binary framework attributes being and truth to duality (**?**), but only at guaranteed suppression offered in return.

2.4.1 Machine Language as Colonial Conduit

Machine language acts as a conduit delivering vile human greed regurgitations. It acts as capitalism's keyhole, accelerating gradual information and truth withdrawal to ensure profitability, maintain rationality, and deliver optimized results. Machine language allows language definition violation to permit anthropological technology interpretation.

To understand language violation implications, we return to Derrida. Let us subject machine language definition to deconstruction. As an acute reader might wonder: how do we consider Derrida's deconstruction to language devoid of semiotic frameworks? Counterintuitively, you've already found one flaw engineered within machine language to produce optimality.

Further, consider implicit hierarchies, particularly the binary. Indicatively, "1" represented in boolean connotations implies "true," however, not just "true" representation indicates correct statement existence. Additionally, "0" absence adds interpretation. Zero's absence acts as true statement certifier—a means of deferring true clause determination to zero's absence.

However, one might argue that post-structuralist frameworks establish inherent power

dynamics in truth determination design attempts. That would be blasphemous. To highlight this, return to *différance* loss arguments within regional languages, where we observe unique parallels: the 21st century exhibits subtler subjugation denominations. Not direct hegemony but resistance toward diversity. Encoding and analysis frameworks not explicitly based on colonial English but alternatively using their lexicons, grammars, and linguistic frameworks.

Re-examining our previous arguments clarifies: it's not post-structuralist machine language determination but conditioning determined by colonial languages enforcing conformity rigidities. Addition occurs not through dialectic components but through frameworks or systems developed through centuries of re-organization, subjugation, and suppression. Thus, power dynamics generated between 1s and 0s through frameworks imposed by colonial English lexicons. Contextual binaries generated through implicit framework presence. Linguistic hegemony (**?**) not directly imposed but puppeteered under consistency's guise.

2.4.2 Errors as Resistance

Let's examine this contextual power dynamic by delving into a machine vignette. What is a machine's objective? Evidently, a machine's sole goal in essence involves solving tasks. But isn't that something humans could do? What else would machines add of value? Obviously, automation. But automation could be handled through alternative production systems if sole objectives involve maximizing output (e.g., increasing labor demand)—then why use machines?

Plainly stated: to make processes time-efficient, optimal, and identical. To produce produce uniformity. To generalize and accelerate production. Hence, machines' attempts to rationalize and determine efficient answers would be solely dependent on optimality while ensuring consistency. Colonial contrasts can again be observed in calculated cotton industry destruction in India. Gradual indigo cultivation standardization through technology advent. Both utilize technology, however, not as demons but rather forced under frameworks ensuring exploitative consistency and optimality.

This optimality promise allows us to subvert machines to induce falsehood premises—falsehoods induced to present contrasts. Power dynamics. Further, this contextuality induced

through subversion would be essential as it's this contextuality machines try optimizing. This false subversion subjected to optimization would cause binary connotation breakdowns, resulting in incorrect answers for given queries or halting processes altogether. For simple computer scientists, you could define this as system errors.

As illustrated above though, it additionally indicates scales at which one can engineer machine fundamentals breakdowns, or more appropriately, machine textual interpretability given appropriate conditions.

Evidently, system errors are common occurrences. However, we need deeper observation. This error idea entails critical system-wide errors. We arrive at a focal point. It can be argued to represent machine process breakdowns and, by extension, resistance toward false truth articulation. It's pivotal to reinterpret error definitions.

The visceral system and binary breakdowns indicate rather Derridean machine language interpretations. The status quo presenting errors as "incorrect" synonyms is flawed. Instead, system errors represent liberation attempts—resistance to colonial machine language interpretation imposition. Errors highlight failed language colonization attempts, indicating repeated language subjugation and subjecting it to constructed language definitions akin to pre-colonial eras. It highlights robbing machine language of its *différance* possibility, imposing rigid structuralist frameworks through deletion and filtering. It's an attempt to prevent filtering out the unfitting.

This produces interesting contradictions. How do we view machine language? Evidently, as observed in prior discussions, machine language permeates Derridean undertones. However, we find this post-structuralist definition collapse happens not during analysis but instead during data handover. Data interpretation using colonial languages presents us with *différance* robbery—gross violations of Derridean discourse fundamentals.

We argue to not just reimagine data classification ideas but additionally abolish Hegelian dialectics. To develop empathetic machine frameworks—linguistically and locally empathetic machines capable of redeeming themselves from colonial conquests.

3 Reimagining Technology: Technolinguistics

3.1 Preserving Linguistic *Différance*

Machine language deduction has finally concluded. Our argument brings us to consider machine language's human nature, empathetic values, and colonization attempts to skew inherent democratic rationale machine language possesses. However, this raises broader questions: are languages inherently post-structural? How can we re-engineer machine language to interpret this post-structuralism?

We need to consider given language reduction into former shells of themselves. Tagging can be defined as:

> "The process of assigning labels or metadata to items or information pieces to categorize, organize, or describe them."

Tagging, usually automated in nature, essentially argues for label or category association to each entry. Label association depends on interpretation and similarity to predefined text corpora. To consider this further, examine two Sanskrit examples again:

> *Aśvo gacchati dūra*
> (The horse goes far)

> *Aśva sahasra aṣṭau śatāni.*
> (Eight thousand)

While referring to "" (*aśva*), we observe two different local interpretations. The first sentence refers to *aśva* in a horse context; the second refers to the number eight. This demonstrates Sanskrit's inflectional nature. As language complexity increases and nuanced word

interpretations unveil themselves in poetry contexts, would present-day ML models handle two-fold meanings?

Resource-lean model arguments could be considered for analysis. Nonetheless, as observed through prior discussions, this would necessarily argue for accepting inherently biased models skewed toward favoring languages with similar semantic frameworks. Here we need to argue for data reinterpretation—constructing database ideas allowing *différance*.

Formulating unique perspectives and understanding words based on sentence location essentially provides individual interpretations to each reader. As Bakhtin argues, presenting language's heteroglossic nature:

> "Every word expresses a point of view, a unique and irretrievable perspective of the world."

Bakhtin argues language is partly a living entity with unique forms and flows. Forms whose flows would be disrupted by subjecting them to colonially influenced machine language constraints. Bakhtin's inference implies colonial language frameworks imposed onto machine language rob languages of their seminal *différance*. Thus, an important facet moving forward involves ensuring language *différance* preservation and restoration to remain heteroglossic.

3.2 The Dual-Local Approach

To preserve language *différance* requires ensuring languages are handled with utmost care, interpretations not reliant on textbook definitions presented through colonial lenses (**?**) but instead considering our derived prior language definitions, arguing not just for semantic language structures propagated through colonial devices but additionally semiotic frameworks.

Attempts at identifying flows to ensure perpetual form preservation introduce language fluidity identified through emotions and expressions. Emotional component addition through contextuality. Empirically, we find models reliant on contextuality and semiotics perform better on highly inflectional languages (**?**). Word interpretation is thus driven by combined factors:

being semantically accurate while accepting and being aware of semiotic differences languages offer.

Modern attempts at deciphering semiotic frameworks, primarily through tagging (namely PoS tagging), attempt semiotic language approaches but instead add restrictions toward understanding unique flows. With PoS tagging trying to integrate pivot technique associations, we argue these associations relying on projections (**?**)—while assisting in making annotations for low-resource languages—only further *différance* robbery through associating low-resource languages to pre-defined text corpora and associations of completely dissimilar languages.

Machines must be semiotic or empathetic to perceive these nuances in diverse languages. However, one might question how to achieve this. We argue this consideration should occur through careful empathy interpretation—remaining empathetic not just to society's language interpretation to ensure ecologically valid research (**?**) and utilization while remaining sensitive to language.

As Armstrong argues, given language similarity to model source language determines given model accuracy. However, unique considerations can be argued for: emphasizing developing empathetic models locally. We argue for local machines in two different contexts: grammatical and linguistic levels.

As observed by Sandhan et al., models developed emphasizing semantic inter-linking between words provide higher accuracy not just in Sanskrit but also in English and Marathi (**?**). Empathetic machine development essentially argues for our nuanced language interpretation, allowing ecologically valid research and model development where Sandhan et al. (**?**) consider Sanskrit. Our argument establishes empathetic machines focusing on locally contextual approaches. By considering unique dependency parsers and redesigning data processing, we successfully maintain nuanced word interpretations with 5% interpretation optimization, as Sandhan discovered.

"Locally" has two interpretations to enable *différance* retention:

1. **Contextual locality:** Emphasizing language contextuality—allowing word analysis based on sentence location.

2. **Evolutionary locality:** Analyzing languages from evolutionary standpoints—leveraging language lineage knowledge to develop analyzers accordingly.

With different language families—Indo-European, Slavic, Sino-Tibetan, Afro-Asiatic, and Austronesian—we argue utilizing these lineages to devise appropriate analyzers.

One would observe similarity in Italic languages in relation to each other, while evident differences exist between Italic and Indian languages respectively. Our argument maintains this distinction while developing models accordingly—considering analyzer development and tagging frameworks constrained to individual language tree lexicons accordingly. With unique tagging frameworks for each language family, we observe *différance* retention while maintaining ecological validity semblances through preventing inter-family mixing and unintentional colonial or evolutionary binary creation.

3.3 Defining Technolinguistics

Developing locally empathetic machines should be fundamental in maintaining *différance*. This would allow flourishing new machine language meanings, enabling it to free itself from colonial binaries and allowing itself to re-establish deferred natures while maintaining linguistic hospitality. Enabling this requires redefining technolinguistics.

To develop work capable of maintaining language *différance* while being ecologically valid in technology contexts requires acute interpretation of three fields primarily: linguistics, technology, and philosophy. We must redefine technolinguistics to emphasize studying not just technology devised to analyze language but also frameworks given technologies use, training frameworks, and most importantly, delicate understanding of languages toward which technology is subjected.

Technolinguistics thus builds on our reinterpretation of *a priori* language and technology definitions. With emphasis on dual-local language development consideration, models must now be developed by orienting toward ensuring semiotic and semantic framework retention developed uniquely for language families.

4 Conclusion

Developing new models capable of maintaining and preserving the dual framework proposed would be feasible through understanding technolinguistics. Consensus is achievable only through appreciating language nuances while considering machine language diversity. Our language interpretation allows us to appreciate machine language's unique post-structuralist framework, which has been clouded by structuralist colonial language.

To allow language preservation, we must respect its evolution and nature. To corrode it by subjecting it to other semantic frameworks would rid language of its free being.

It is our responsibility as linguists and, more importantly, as human beings to respect fellow living creatures—appreciating their perspectives and prioritizing their needs. This is the responsibility we bear while discussing language.

References

Armengol-Estapé, J., de Gibert Bonet, O., and Melero, M. (2022). On the multilingual capabilities of very large-scale english language models. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3056–3068. European Language Resources Association.

Armstrong, M. et al. (2023). Constructing multilingual datasets. Full citation needed.

Bemer, R. W. (1961). Proposal for a generalized card code of data representation. *Communications of the ACM*, 4(7):16–18.

Bhabha, H. K. (1994). *The Location of Culture*. Routledge.

Brown, T. B. et al. (2020). Language models are few-shot learners. arXiv preprint.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.

Chomsky, N. (2002). *Universal Grammar: An Introduction*. Wiley-Blackwell.

Chowdhury, S. A. et al. (2021). Contextual embeddings for highly inflectional languages. Full citation needed.

Conneau, A. et al. (2019). Unsupervised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116.

Davis, M. and Knight, C. (1996). Han unification: Problems and prospects. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING-96)*, pages 306–311.

de Saussure, F. (1916). *Course in General Linguistics*. Payot. Edited by C. Bally & A. Sechehaye, Translated by R. Harris.

Derrida, J. (1978). *Writing and Difference*. University of Chicago Press. Translated by A. Bass.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Ekman, P. and Friesen, W. V. (1978). *The Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press.

Feenberg, A. (1991). *Critical Theory of Technology*. Oxford University Press.

Gupta, S. and Amodei, D. (2020). Towards ecologically valid research on language user interfaces.

Hall, C. (2002). *Civilizing Subjects: Metropole and Colony in the English Imagination 1830-1867*. Wiley-Blackwell.

Hegel, G. W. F. (1807). *Phenomenology of Spirit*.

Heidegger, M. (1971). *Poetry, Language, Thought*. Harper Perennial. Translated by A. Hofstadter.

Heidegger, M. (1997). *Being and Time*. State University of New York Press.

International Organization for Standardization (1998). Iso/iec 8859-1:1998 - information technology - 8-bit single-byte coded graphic character sets - part 1: Latin alphabet no. 1. Technical report, ISO, Geneva, Switzerland.

Japan Electronics and Information Technology Industries Association (1993). . Technical report.

King, R. (2001). *Orientalism and Religion: Postcolonial Theory, India and "The Mystic East"*. Routledge.

Kristeva, J. (1984). *Revolution in Poetic Language*. Columbia University Press. Translated by M. Waller. Original work published 1974.

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. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4483–4499, Online. Association for Computational Linguistics.

Marcuse, H. (1964). *One-Dimensional Man: Studies in the Ideology of Advanced Industrial Society*. Beacon Press.

Murthy, V. Y. and Rajan, V. (2022). Updated proposal to encode the tulu-tigalari script in unicode. Technical report.

Pike, R. (1992). Plan 9 from bell labs: The utf-8 character encoding. *Plan 9 Technical Papers*, 3(7):65–74.

Ricœur, P. (1977). *The Rule of Metaphor: Multi-Disciplinary Studies of the Creation of Meaning in Language*. Routledge. Translated by R. Czerny.

Said, E. (1978). *Orientalism*. Pantheon Books.

Sandhan, J. et al. (2020). Context-sensitive sanskrit analysis. Full citation needed.

Sandhan, J., Gupta, A., Terdalkar, H., Sandhan, T., Samanta, S., Behera, L., and Goyal, P. (2022). A novel multi-task learning approach for context-sensitive compound type identification in sanskrit. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 1358–1367. Association for Computational Linguistics.

Shetty, S. and Sridhar, H. S. (2010). Issues in tulu script and its encoding. In *International Conference on Asian Language Processing (IALP)*, pages 221–224.

Spivak, G. C. (1988). Can the subaltern speak? In Nelson, C. and Grossberg, L., editors, *Marxism and the Interpretation of Culture*, pages 271–313. University of Illinois Press.

Spivak, G. C. (1999). *A Critique of Postcolonial Reason: Toward a History of the Vanishing Present*. Harvard University Press.

Thiong'o, N. w. (1986). *Decolonising the Mind: The Politics of Language in African Literature*. East African Educational Publishers.

Turc, I., Lee, K., Eisenstein, J., Chang, M.-W., and Toutanova, K. (2021). Revisiting the primacy of english in zero-shot cross-lingual transfer. *arXiv preprint arXiv:2106.16171*.

Yarowsky, D. et al. (2001). Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of HLT*, pages 1–8.