

Primum non nocere:

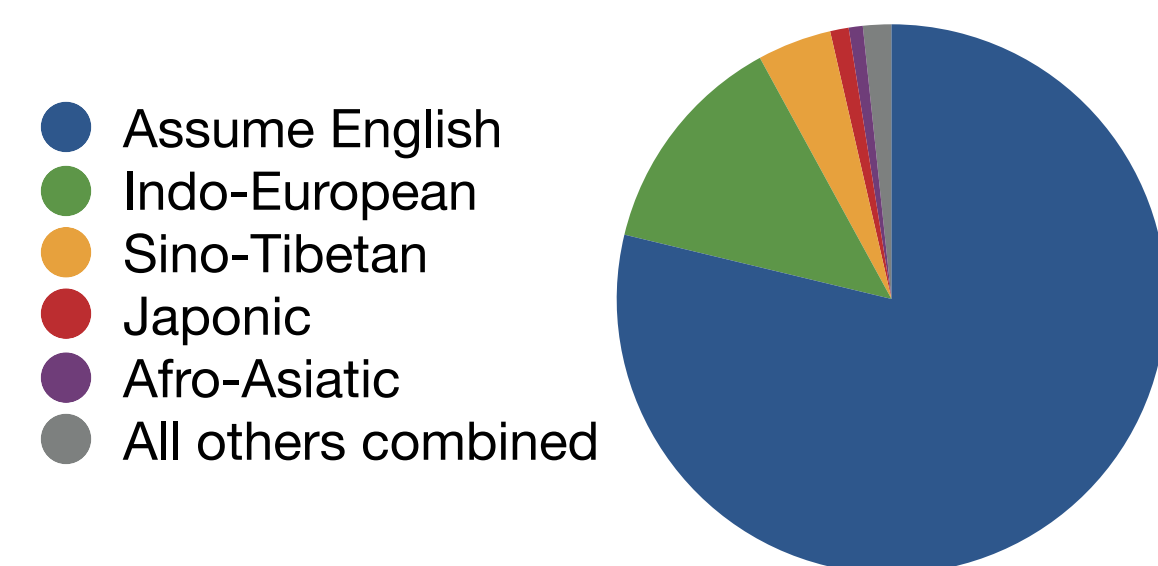
Before working with Indigenous data, the ACL must confront ongoing colonialism

Lane Schwartz, Associate Professor, University of Alaska — loschwartz@alaska.edu

Recent *ACL papers lack language diversity

Of 9602 *ACL abstracts (2013–2021), percentage and number of abstracts that explicitly mention at least one language from the language family.

83.26%	7995	Implicitly assume English
13.70%	1315	Indo-European (incl. English)
4.50%	432	Sino-Tibetan
1.12%	108	Japonic
0.85%	82	Afro-Asiatic
0.41%	39	Turkic
0.26%	25	Koreanic
0.25%	24	Austroasiatic
0.24%	23	Dravidian
0.22%	21	Uralic
0.21%	20	Austronesian
0.09%	9	Basque
0.09%	9	Atlantic-Congo
0.07%	7	Na-Dene
0.05%	5	Kra-Dai
0.02%	2	Arnhem
0.02%	2	Iroquoian
0.02%	2	Inuit-Yupik-Unangan
0.01%	1	Sumerian



Over 83% of *ACL papers fail the #BenderRule:
“Always name the languages you work with.”

Linguistic homogeneity at ACL is a symptom of a much larger problem, namely that our research paradigms are deeply rooted in a Western scientific tradition that is inextricably intertwined with colonialism.

Framing expresses power

The act of failing to name a language, of assuming English as the default language of study in NLP, is a practice that expresses colonial power.

Research scientists rarely consider the philosophy of science on which our research is predicated:

- Ontology
- Epistemology
- Methodology
- Axiology

These surface as unacknowledged assumptions that science is value-neutral and empirical observations and logical reasoning fully and completely define the nature of science and reality.

The first step in enacting decolonial ethical practices is acknowledging that we hold these assumptions and recognizing that there are other Indigenous philosophies of science that are equally valid and are rooted in fundamentally distinct worldviews that center relationality.

Recommended readings



Prerequisite Obligations for Ethical Research involving Indigenous Languages and Indigenous Peoples

Obligation 0: **Above all, do no harm**

Colonization and colonial practices have inflicted substantial and often genocide-scale harm on Indigenous communities over the past five centuries, harm that is ongoing and is often perpetuated by modern research practices.

The **obligation of non-maleficence** therefore mandates that above all else, we do no harm to Indigenous people and Indigenous communities. In practical terms, this means that researchers seeking to engage with Indigenous data critically examine the harmful ramifications of proposed work well before it is conducted. If we can do good through our research without doing harm, that is well, but it is better to not engage than to cause harm.

Obligation 2: **Research must benefit Indigenous people**

The **obligation of beneficence** mandates that we as researchers ensure that our work benefits the Indigenous communities with which we work in ways that those communities recognize as beneficial.

In practical terms, this means that any outside researcher who wants to work with Indigenous data must seek to engage with the relevant Indigenous communities in order to learn about and to meaningfully support priority areas identified by Indigenous governing bodies and decision-making institutions that fall within our respective scopes of expertise.

Put another way, ethical research involving Indigenous data must include concrete deliverables requested by the respective Indigenous community or communities.

ABCs of ethical research:

- **Accountability**
- **Beneficence**
- **Cognizance**
- **Do no harm**

Obligation 1: **Educate yourself about colonial harms**

The colonial political and racial ideas and behaviors that support and enable colonization and oppression are intentionally invented historical creations. Before we engage with Indigenous peoples, let alone work with Indigenous data, we must intentionally make ourselves cognizant of this history. As outside researchers, we stand in a privileged position, and as such have an urgent obligation to educate ourselves about this history and about current practices that perpetuate these systems of oppression.

The **obligation of cognizance** mandates that we as researchers intentionally and thoroughly educate ourselves about colonization of Indigenous communities; about the role that academic researchers have had and continue to play in the exploitation of Indigenous communities, Indigenous languages, Indigenous culture, and Indigenous data; and about Indigenous expectations and ways of being centered on relationality that differ from those we typically encounter in our research.



The **obligation of accountability** mandates that we as researchers develop meaningful relations with the sovereign governing bodies of the Indigenous communities with which we seek to engage, and that we be meaningfully accountable to such bodies in our work involving their data. This relationship-building should take place before the research project begins. This relationship between researcher and sovereign Indigenous institutions can be thought of as highly analogous to the relationship between the researcher and governmental granting agencies such as the U.S. National Science Foundation.