

Assessment of Online Translation Tools for Indigenous Languages in the Ocean Domain in the Arctic

Sigmund Kluckner, Maria van Veldhuizen, Jay Pearlman
IEEE France Section, as part of the project CAPARDUS



This effort is part of the CAPARDUS project.
The CAPARDUS project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No. 869673.



Welcome to the last presentation in the last session on the last day

Show of hands:

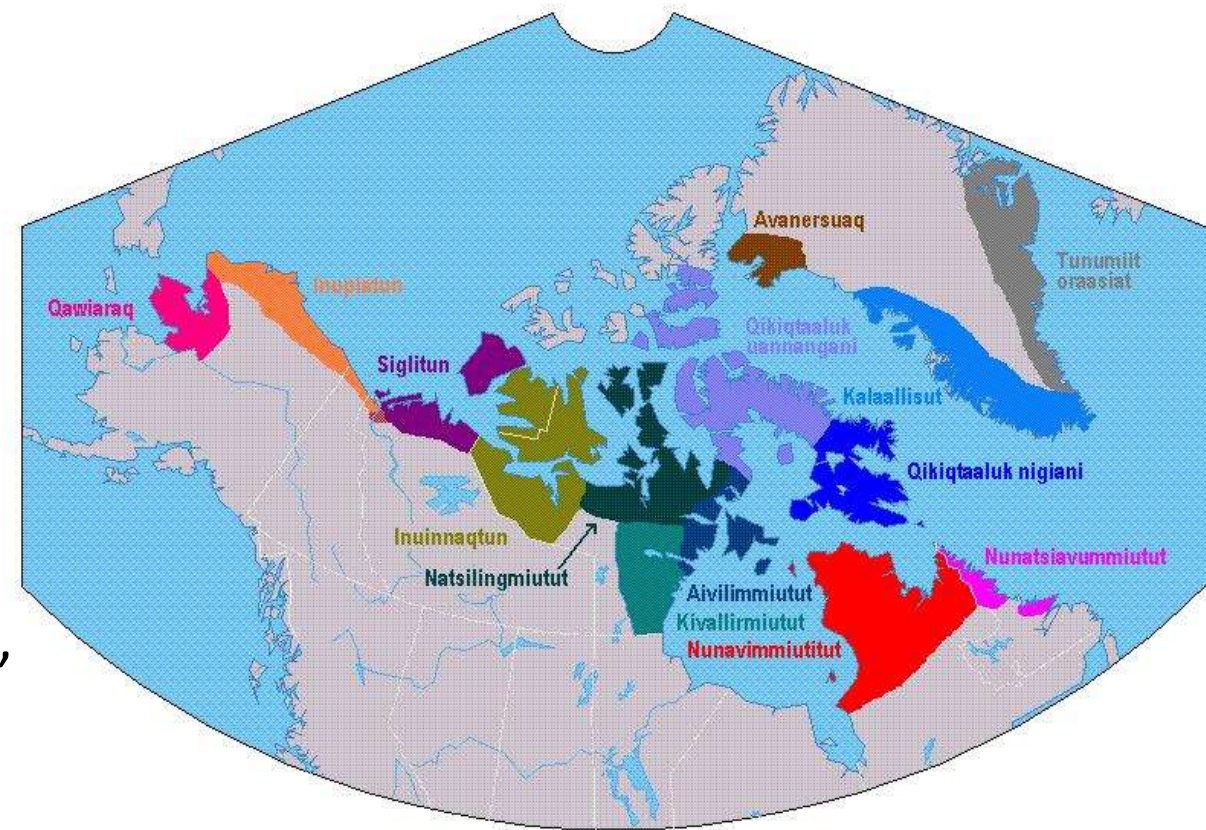
- Who's mother tongue is not English?
- Except for English and your mother tongue, who speaks a third/fourth/... language?
- Who can explain Climate Change in that language (or any third language)?
- Who of you speaks any languages of the Arctic (including Kalaallisut)?
- Now imagine, that third language is only spoken by a few thousand people:
 - What are the odds of correctly communicating scientific/topical knowledge?

Background – Why did we do this?

- Knowledge/information published on Arctic and Ocean topics
- Information could be useful to local and indigenous communities
- BUT: much of it is published in English or other major language, **adding an access barrier**
- Can automated (machine) translation help?

Background (2)

- Languages in the Arctic
 - Partly share same roots
 - Not mutually intelligible
 - Considered „Low Resource Languages”
 - Few thousand speakers
 - Few written artifacts
 - Few documents with direct translation to a major language
- Ocean Information
 - Might not reach "remote" communities



What we did

- Desk research on existing tools
 - Focus: Neural Machine Translation Systems
 - Support for sample of 15 relevant languages
 - Feature comparison
- General Quality assessment of six tools
 - Using the same test document
 - Target languages: DE, NL, FR, DA (limitation: team members/associates)
- Translation capacity for Arctic languages
 - Which tools are available?
 - Quality assessment with multi-lingual documents for comparison

Wed May 22
In the fluorometric method, without using any membrane diffusion, the sample is combined with a working reagent made up of OPA, sodium sulfite, a borate buffer, and then heat to 75 °C. Fluorescence proportional to the ammonium concentration is measured at 460 nm following excitation at 370 nm. For the membrane diffusion method NH₄⁺ ions in the sample are converted to NH₃ gas with subsequent diffusion across a Teflon membrane into a stream of OPA. The product is fluorometrically measured at 460 nm following excitation at 370nm. This method is for nanomolar analysis (Jones, 1991).

2. Relatively academic language, including acronyms.

Pearlman, J. et al. (2021). Evolving and Sustaining Ocean Best Practices to Enable Interoperability in the UN Decade of Ocean Science for Sustainable Development. *Front. Mar. Sci.* 8:619685. doi: 10.3389/fmars.2021.619685

Thus far, we referenced the UNESCO Intergovernmental Oceanographic Commission (IOC) Ocean Best Practices System (OBPS) as an example of marine methodology management system (MMS), but there are many other ocean-relevant methodology collections published in differing formats by, for example, the International Council for the Exploration of the Sea (see also ICES, 2019), the World Meteorological Organization, the National Ecological Observatory Network (NEON) and multiple research institutes around the world. Unfortunately, this immense pool of knowledge is typically not interoperable, coordinated, or effectively digitized. When collections are accessible online, ocean practitioners still cannot search across these resources without knowing they exist and using multiple portals with drastically different user experiences.

Table 1: The UN Ocean Decade outcomes and how they can each be supported by systematic handling of methodological content (methods, standards, policies, etc.) and identification of best practices.

Ocean Decade outcome	Relationship to a coordinated ocean methodology and best practices system
1. A clean ocean where sources of pollution are identified and reduced or removed	Methodologies associated with a context-sensitive registry or ad hoc list of pollutants (physical, chemical, biological) can be rapidly identified and triaged for suitability for monitoring and remediation
2. A healthy and resilient ocean where marine ecosystems are	Methodologies can be indexed and accessed by their

What we found

- 43 tools catalogued and assessed:
 - Barely any supported Kalaallisut, Inuktitut
 - No support: Iñupiaq, Saami
- Quality Assessment
 - DeepL and Google Translate scored highest
 - Tools struggled with
 - Consistency
 - Symbols (e.g. taking over correct formatting)
 - Diagrams with Labels (little context)
 - All translated documents found in link:
 - <https://drive.google.com/drive/folders/1Z4HKpdzMBQgPrVodXkZlIS6JHIWUerKp>



An “outside” test on uptake

- Machine translate endorsed practices on OBPS
 - Thanks to Pauline Simpson for support on this!

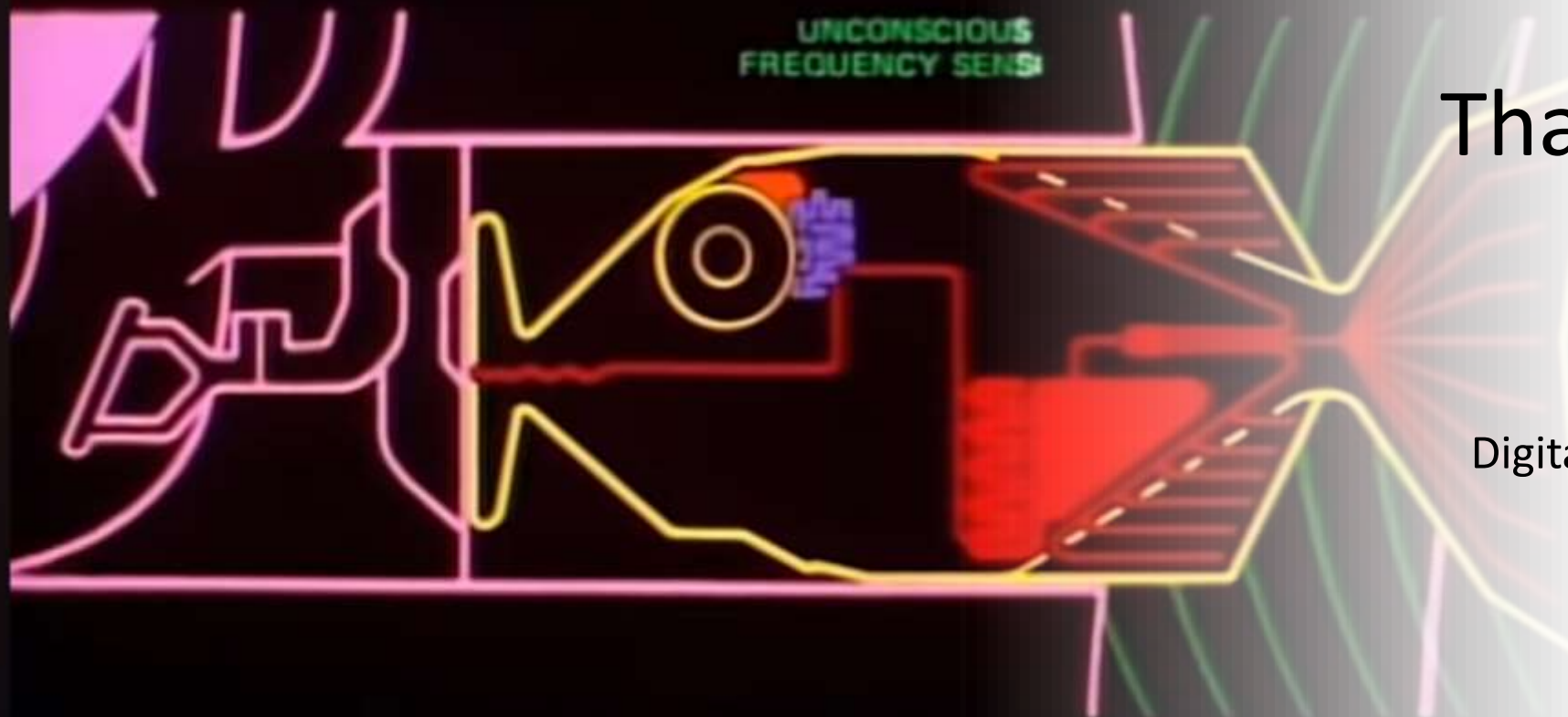
GOOS ENDORSED PRACTICES IN OBPS	DEPOSITED	DOWNLOADS ENGLISH	DOWNLOADS TRANSLATIONS (from: July 2023)
Halewood ,E., Opalk, K., Custals, L., Carey, M., Hansell, D.A. and Carlson, C.A. (2022) Determination of dissolved,organic carbon and total dissolved nitrogen in seawater using High Temperature Combustion Analysis. Frontiers in Marine Science, 9:1061646, 15pp. DOI: https://doi.org/10.3389/fmars.2022.1061646	2023-05-23	2722	DE -596 ES -805 FR -663 PT - 239
OceanSITES (2020) OceanSITES Data Format Reference Manual NetCDF Conventions and Reference Tables. Version 1.4 July 16, 2020. Geneva, Switzerland, OceanSITES, JCOMMOPS, 36pp. DOI: http://dx.doi.org/10.25607/OBP-421.2	2019-03-08	532	DE - 206 ES - 475 FR - 1140 PT -660
Parks, Justine, et al (2021) XBT Operational Best Practices for Quality Assurance, Version 1.0. San Diego, CA, Scripps Institution of Oceanography, Climate, Atmospheric Sciences, and Physical Oceanography, University of California, 32pp. DOI: http://dx.doi.org/10.25607/OBP-1720	2022-02-18	6280	DE - 185 ES - 295 FR - 80 PT - 122

Discussion & Conclusion

- Machine Translation can be a useful tool (for major languages)
- Translations good enough to convey the gist of the message
- Translations still need human quality checks

- No high quality models available (yet) for low resource languages
 - Some projects working on this challenge (e.g. Meta “No Language Left Behind”)
- “Generative AI” Models are getting better at translating
 - See OpenAI’s recent demo of ChatGPT 4-o
 - Other tools like videos with lip sync into other languages

- Automatic Translations can help reach underserved populations and regions



Thank you

Sigmund Kluckner

Digital Innovation & Sustainability

sigmund@kluckner.eu

Sigmund.Kluckner@ieee.org

THE PRACTICAL UPSHOT OF WHICH IS THAT IF YOU
STICK ONE IN YOUR EAR, YOU CAN INSTANTLY
UNDERSTAND ANYTHING SAID TO YOU IN ANY FORM
OF LANGUAGE: THE SPEECH YOU HEAR DECODES THE

Also Me:

Iliad – Standardization for Digital
Twins of the Earth & Ocean (via
IEEE)