Translating a low-resource language using GPT-3 and a human-readable dictionary

Micha Elsner and Jordan Needle

THE OHIO STATE UNIVERSITY

Inuktitut: a polysynthetic language

Long words composed of many content and functional morphemes:

saviggirunnaqtutit

savik- ggiq -runnaq -tutit

snow.knife -bring -ABIL -2S

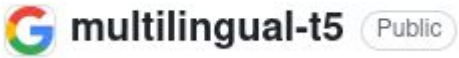
“You can bring your knife.”

snow-knife: Natl. Museum of the American Indian

An inexperienced speaker cannot easily use individual morpheme definitions, or even the gloss, to understand such a word.

Resource development: NMT

Large corpus



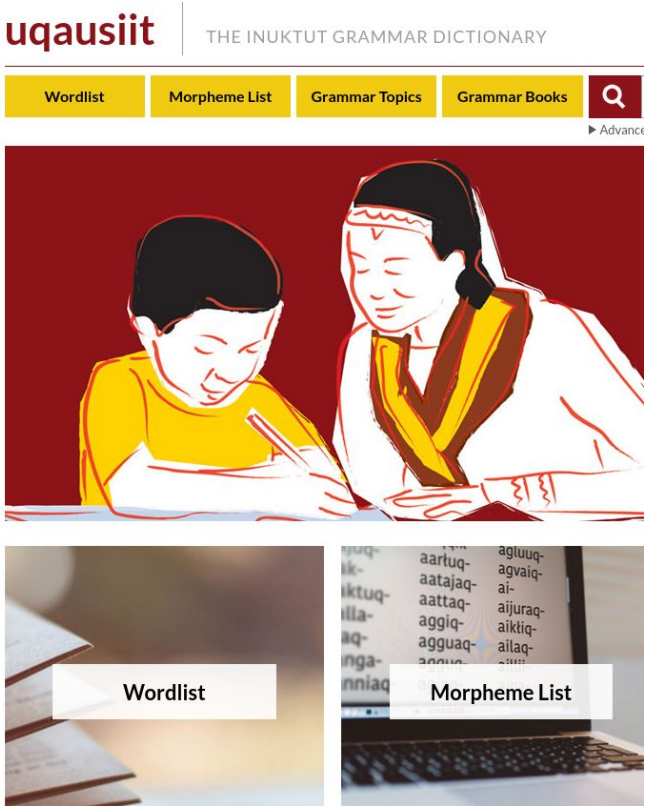
Fine-tuned NMT

saviggirunnaqtutit →
you can bring your knife

Translator app

But we may need a **lot** of data!

Resource development: human-focused



Example

saviggirunnaqtutit

« You can bring your knife. »

Related Morpheme:

savik

« (1) metal; steel; iron (2) snow-knife »

-ggiq-

« to bring someone or something along »

-runnaq-

« to be able to perform a certain action »

-tutit

« you »

This work: translation with human-readable materials

Communities often want human-readable grammars and dictionaries, and may invest significant effort in building them.

They provide material for learners and inexperienced speakers, written in a well-resourced language.

Can we use the **monolingual text-to-text** abilities of large LLMs to leverage this material?

Previous work

Focus on adding dictionary resources to NMT (Luong et al. 2015, Zhang et al. 2021, Niehues 2021, Ghazvininejad et al. 2023).

Our contributions:

1. Do this **without** a base NMT system
2. Target a polysynthetic language.
3. Also map from whole-word translations to suggested definitions (the reverse task)

Translate the following sentence to English: *Pada dasarnya, hal tersebut terbagi ke dalam dua kategori: Anda bekerja sambil mengadakan perjalanan atau mencoba mencoba atau membatasi pengeluaran Anda. Artikel ini berfokus pada hal yang terakhir.*
In this context, the word "sambil" means "while"; the word "membatasi" means "limiting", "restrict", "limit".
The full translation to English is: *Basically, they fall into two categories: Either work while you travel or try and limit your expenses. This article is focused on the latter.*

Translate the following sentence to English: *Ia melakukan pembuatan bel pintu dengan teknologi WiFi, katanya.*
In this context, the word "pembuatan" means "creation"; the word "bel" means "buzzer", "bell"; the word "pintu" means "door", "doors".
The full translation to English is:

Ghazvininejad et al. 2023

Dataset

Inuit Uqausinginnik Taiguusiliuqtiit
ᐃᓄᐃᑦ ᐅᖅᑲᐅᓯᖕᒋᓐᓂᒃ ᑕᐃᒍᓯᓕᐅᖅᑏᑦ

Dictionary example words with English definitions scraped from Uqausiit dictionary:
saviggirunnaqtutit → you can bring your knife

Segmented into canonical morphemes using dictionary-provided structure, filling gaps with FST segmenter (Farley 2012)
saviggirunnaqtutit : savik ggiq runnaq tutit

	Segmented words with translations
Dev	22
Test	219

Substitute into template to generate GPT-3 prompt

The Inuktitut word saviggirunnaqtutit is made up of the following parts:
savik: (1) metal; steel; iron (2) snow-knife; or, to be fitted with a metal point (harpoon; spear)
ggiq: to bring someone or something along, as in "Don't bring your gun along."
runnaq: to be able to perform a certain action, as in "Could you find that out if he/she arrived"
tutit: you, as in "you sleep"
In English, saviggirunnaqtutit means roughly: you are able to bring a snow-knife

Prompt tuning: questions

Baselines:

- Translation without the definitions (does GPT-3 know Inuktitut already?)
- Concatenate the definitions in order

To what extent does GPT-3 benefit from:

- Dictionary usage examples
- Different ways to elicit multiple answers when morphemes have multiple definitions
- A tiny “grammar lesson” explaining some compositional phenomena
- Chain-of-thought prompting

Baselines

	BLEURT (Sellam et al. 2020)	Translation
No definitions	13	We are thankful.
Concatenate	43	(1) metal; steel; iron (2) snow-knife, to be fitted with a metal point (harpoon; spear) to bring someone or something along to be able to perform a certain action you

GPT-3 does **not** know Inuktitut.

Text-to-text generation **is** required.

Systems		
	BLEURT (Sellam et al. 2020)	Translation
Definition only	48	are you able to bring a snow-knife?
Definition + usage example	51	you are able to bring a snow-knife
Ask for multiple answers (average)	46	you are able to bring metal along
Ask for multiple answers (oracle best)	59	you are able to bring metal along
Grammar lesson	48	you are able to bring a snow-knife
Chain-of-thought	43	you can make it

Definitions help!

Asking for multiple answers is useful if the user has contextual cues for which one is right.

GPT-3 is not great at applying grammatical information to concrete examples.

Definition induction

The Inuktitut word niriligiaqtut is made up of the following parts:

niri: (currently unknown)

giaq: to start to do something; to try to do something

liq: an action that is underway or starting; also marks a change from one state to another

tut: they (3+)

In English, niriligiaqtut means roughly: "they have started to eat"

"niri" probably means "eat".

Results are not as good as for translation, but show promise for root morphemes.

GPT-3 can usefully translate polysynthetic words by composing English definitions.

Potentially useful for communities which are investing their time in dictionaries and grammars rather than corpora.

Future work? Better prompts, even more powerful LLMs, incorporation of segmentation and disambiguation within the few-shot framework.

See the paper: more examples, definition induction results, BLEU scores, confidence intervals, ethics discussion.

Thank you / qujannamiik!