# Towards Automated Translation of Poetry

Michael A.

Exeter Mathematics School

March 30, 2017

# What is this?

- Translation is hard — for humans and computers
- Poetry is awesome, but difficult to move across languages
- What if a computer could translate poetry for us?
- EMC: hypotheses for improving existing methods
- A look at literary and computational approaches to poetry

# What is poetry?

> If I read a book and it makes my whole body so cold no fire can warm me I know that is poetry. If I feel physically as if the top of my head were taken off, I know that is poetry. These are the only way I know it. Is there any other way?
>
> — Emily Dickinson

If I read a book and it makes my whole body so cold no fire can warm me I know that is poetry. If I feel physically as if the top of my head were taken off, I know that is poetry. These are the only way I know it. Is there any other way?

— Emily Dickinson

- Quite a qualitative description of poetry
- As an expression of feeling, emotion — just a human thing?
- How can machines translate it if it's exclusive to humans?
- Does this exclude certain kinds of poetry? What about poems that eg. only play with the language, as opposed to seeking to be as ground-breaking as Dickinson suggests?

# What is poetry?

Poetry is that which is lost in translation.

— Robert Frost

Poetry is that which is lost in translation.

— Robert Frost

- Different idea of poetry — focus on aesthetic and rhythmic form
- Emotional response to a poem closely linked to cultural context around it — cannot translate that. Form, on the other hand...
- More what most people would associate poetry with — rhythm, rhyme, alliteration, form...
- EMC was looking at how we can translate these kinds of form, rather than emotion

# What is MT?

- **Machine translation** (MT) is translation using computers
- Warren Weaver's theories:
  1. Looking at *n* surrounding words to disambiguate word meaning
  2. "Deducing any legitimate conclusion from a finite set of premises"
  3. Cryptographic process
  4. Descend "down to the common base of human communication"

# Why translate poetry?

1. Translated poems contribute massively to English literature
2. Translating to/from languages with little or no translation communities
3. Translated poems worth reading in their own right
4. Computational, quantitative treatment of poetry

# Nabokov

> The hack who has never read the original, and does not know its language, praises an imitation as readable because easy platitudes have replaced in it the intricacies of which he is unaware.
>
> — Vladimir Nabokov

# Nabokov's theories

- **Paraphrastic** A free version of the original
- **Lexical** Basic translation of words and their order
- **Literal** Preserving exact contextual meaning

To reproduce the rhymes and yet translate the entire poem is computationally impossible.

— Vladimir Nabokov

Nabokov's theories

- **Paraphrastic** A free version of the original
- **Lexical** Basic translation of words and their order
- **Literal** Preserving exact contextual meaning

To reproduce the rhymes and yet translate the entire poem is computationally impossible.

— Vladimir Nabokov

- Nabokov wrote a translation of Eugene Onegin, which was literal — didn't translate rhyme scheme, but otherwise completely faithful to original
- However literal impossible to do computationally — would need general AI (cultural context, emotion, etc)
- Lexical good for getting general idea of poem, easiest to do computationally
- Paraphrastic not great - doesn't preserve original poem. But doable computationally

# Goethe: original

Über allen Gipfeln
Ist Ruh,
In allen Wipfeln
Spürest du
Kaum einen Hauch;
Die Vögelein schweigen im Walde.
Warte nur, balde
Ruhest du auch.

Johann Wolfgang von Goethe

Goethe: original

Über allen Gipfeln
Ist Ruh,
In allen Wipfeln
Spürest du
Kaum einen Hauch;
Die Vögelein schweigen im Walde.
Warte nur, balde
Ruhest du auch.

Johann Wolfgang von Goethe

- Widely considered most beautiful poem in German language
- Contrast between man — restless, uncomfortable in the silence of the forest — while nature is united in silence
- Scale of poem — large scale (summits), middle distance (treetops), immediate surroundings (forest), then finally man
- Encompassing all the universe
- Stylistically — note rhyme scheme, meter

# Goethe: lexical

Over all peaks/summits/tops
Is rest/peace/silence,
In all treetops
You (informal) sense/feel
Hardly a breath/breeze;
The little birds remain silent/keep still in the wood/forest.
Just wait, soon
You rest/repose too/also.

TU Chemnitz Dictionary

Goethe: lexical

Over all peaks/summits/tops
Is rest/peace/silence,
In all treetops
You (informal) sense/feel
Hardly a breath/breeze;
The little birds remain silent/keep still in the wood/forest.
Just wait, soon
You rest/repose too/also.

TU Chemnitz Dictionary

- I've let alternative translations be represented by slashes - clearly, many options on how to translate each word
- Purposely sacrificing ease of reading, beauty, etc for completeness of translation

# Goethe: paraphrastic

Over all of the hills
Peace comes anew,
The woodland stills
All through;
The birds make no sound on the bough.
Wait a while,
Soon now
Peace comes to you.

John Whaley

Goethe: paraphrastic

Over all of the hills
Peace comes anew,
The woodland stills
All through;
The birds make no sound on the bough.
Wait a while,
Soon now
Peace comes to you.

John Whaley

- What it preserves — rhyme scheme, in a way
- Reordering of lines, change of tense to force rhymes
- Loses largest scale (summits to hills, again to force a rhyme) — but progression of scale change is preserved
- But as a poem on its own, it reads well
- Completely doable to do a translation in this style computationally

# Modeling language

- Cognitive models — Chomsky
- Statistical models as an approximation
- Monolingual and bilingual corpuses

- Humans are good at learning languages, but why?
- Chomsky's theory — 'universal grammar', innate language faculty that 'knows' the rules of language
- But exactly *what* it is eludes us, so can't model it computationally
- Instead, approximate it — mimic it in such a way that it seems close to indistinguishable from the real thing
- Best to do this using a statistical model — given a large sample of text, should be able to estimate the probability of a given word or sentence, and choose the highest probability option
- In MT, need to estimate both the probability of words in one language, and probability of words being translated from one language to another

# Statistical MT

By Bayesian decomposition:

$$\Pr(e|f) = \frac{\Pr(e)\Pr(f|e)}{\Pr(f)} \tag{1}$$

$\Pr(f)$ remains constant, so:

$$\Pr(e|f) = \Pr(e)\Pr(f|e) \tag{2}$$

**Language model** $\Pr(e)$
**Alignment model** $\Pr(f|e)$

Statistical MT

By Bayesian decomposition:

$$\Pr(e|f) = \frac{\Pr(e)\Pr(f|e)}{\Pr(f)} \quad (1)$$

$\Pr(f)$ remains constant, so:

$$\Pr(e|f) = \Pr(e)\Pr(f|e) \quad (2)$$

**Language model** $\Pr(e)$
**Alignment model** $\Pr(f|e)$

- Want to translate a foreign sentence $f$ into an English sentence $e$
- Want to maximize the probability that a chosen $e$ is a correct translation of $f$ (LHS)
- Intuitively — probability that a translator will produce $e$ given $f$
- LHS is hard to model — use Bayes' rule to decompose into RHS. Denominator is same for all $e$, so simplify equation
- Gives two new probability distributions to model — language model and alignment model
- Language model makes output sound fluent
- Alignment model makes sure foreign words are translated to correct English words

# Hypotheses

## 1. Language model

Training the language and alignment models on a poetic corpus improves poetic qualities of output translation.

## 2. Sentence alignment

Altering sentence alignment for poetic works will improve line-by-line translation quality.

## 3. Post-processing

Poetic characteristics can be recovered by post-processing of the output translation.

# Language model

- Estimate the probability of a sequence of words $W = w_1, w_2, \ldots w_n$ by Markov assumption:

$$\Pr(w_n | w_1 \ldots w_{n-1}) \approx \Pr(w_n | w_{n-m} \ldots w_{n-1})$$

- Hypothesis: look to choose more poetic corpus

Language model

- Estimate the probability of a sequence of words $W = w_1, w_2, \ldots w_n$ by Markov assumption:

$$\Pr(w_n|w_1 \ldots w_{n-1}) \approx \Pr(w_n|w_{n-m} \ldots w_{n-1})$$

- Hypothesis: look to choose more poetic corpus

- Language model used to make output sound fluent — want most statistically likely output
- To estimate probability of a sequence of words: Markov assumption, probability of last word is only affected by probability of some history of previous words
- If we want poetic sounding output, makes sense that we'd need poetic sounding input — for Euro languages, EuroParl tends to be used. Not poetic!
- Alignment model and language model could both be used here — but we need bilingual corpus for alignment model, which is hard to get
- So... I built a monolingual corpus myself for the language model. Scrape Project Gutenberg.

# Sentence alignment

one's not half two. It's two are halves of one:
which halves reintegrating,shall occur
no death and any quantity;but than
all numerable mosts the actual more

ee cummings

# Sentence alignment

Sunlight pouring across your skin, your shadow
flat on the wall.
The dawn was breaking the bones of your heart like twigs.
You had not expected this,
the bedroom gone white, the astronomical light
pummeling you in a stream of fists.

Richard Siken

Sentence alignment

Sunlight pouring across your skin, your shadow
flat on the wall.
The dawn was breaking the bones of your heart like twigs.
You had not expected this,
the bedroom gone white, the astronomical light
pummeling you in a stream of fists.

Richard Siken

- Sentence alignment needed for statistical MT — generally sentences will be translated roughly to each other
- Sentence boundaries in prose clearly defined — full stops, question marks, etc
- But in poetry, line breaks can indicate end of sentence, too
- Hypothesis: split into two types of poems, and align like that

# Post-processing

- Recover poetic characteristics by post-processing
- Non-poetic $\rightarrow$ poetic English translation system
- But how to identify poetic characteristics to preserve?

- Genzel et al. used hypothesis constraints to limit output translations to those with eg. certain meter
- Solution could be simpler — implement MT system from non-poetic to poetic
- Identifying characteristics: some work on finding stress from dictionaries and speech synthesis, rhyme from unsupervised learning

# Where to next?

- Implementation of hypotheses — Moses? GIZA?
- Applying to other MT systems — neural networks (Google Translate)
- Preserving contextual content, emotions

Poetry is that which is lost in translation,
Unless we use a computational calculation.

— Soon?

# Want to know more?

- Report and poster in Einstein