Natural Language Understanding, Generation, and Machine Translation

Lecture 22 and 23: Paraphrase Acquisition and Generation

Laura Perez-Beltrachini

(Slides by Mirella Lapata)

School of Informatics
University of Edinburgh
mlap@inf.ed.ac.uk

What Is Paraphrasing?

What Is Paraphrasing?

Sentences or phrases that convey approximately the same meaning using different words (Bhagat and Hovy, 2012).

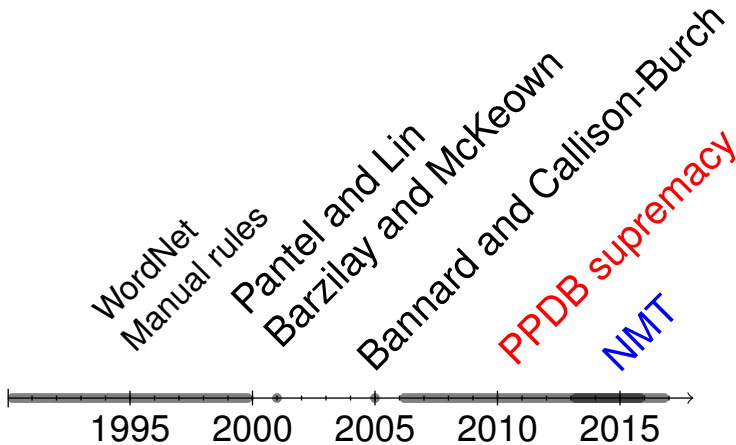| Words	Phrases	Sentences
fortunate	for example	The parrot is dead
lucky	one such example	It's not pinin,
favourable	an istance	it's passed on!
advantageous	in particular	This parrot is no more!
opportune	including	It has ceased to be!
timely	specifically	It's gone to meet its maker!
convenient	however	It's a late parrot!
felicitous	more concretely	It's a stiff!

Why Should we Care?

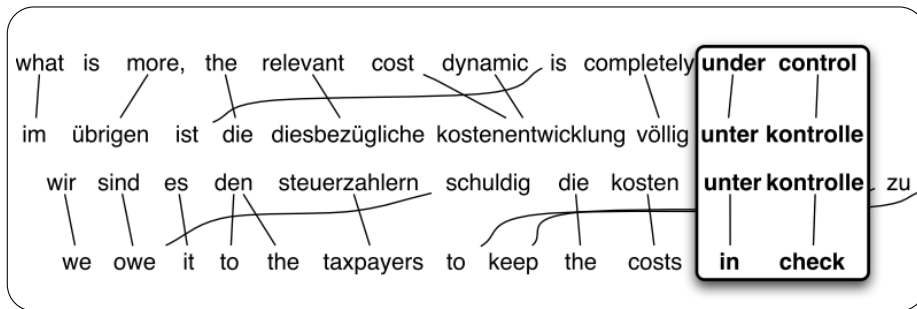Recognizing and generating paraphrases is an important component in many natural language processing applications:

- Question Answering (Fader et al., 2013)
- MT Evaluation (Denkowski and Lavie, 2014)
- Semantic Parsing (Berant and Liang, 2014)
- Simplification (Xu et al., 2016)
- Summarization (Woodesend and Lapata, 2012)
- Style transfer (Xu et al., 2012; Pavlick and Nenkova, 2015)
- Grammar correction (Dahlmeier and Tou Ng, 2011)

The Past

A Very Brief History of Paraphrasing

WordNet
Manual rules
Pantel and Lin
Barzilay and McKeown
Bannard and Callison-Burch
PPDB supremacy
NMT

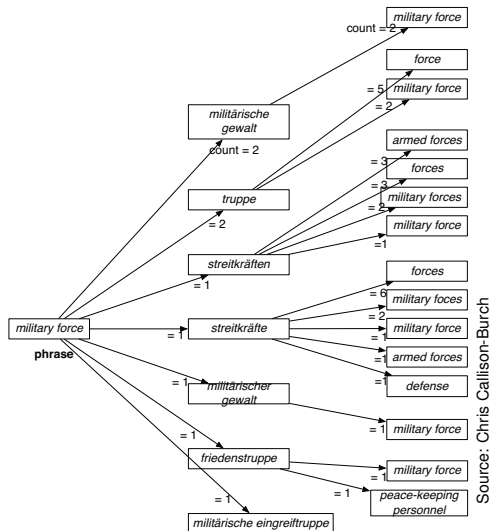1995 2000 2005 2010 2015

Bilingual Pivoting: An Example

Source: Bannard and Callison-Burch (2005)

- No red direct path from en → en; indirect path from en → de and de → en
- If two English phrases are aligned with the same phrase in another language they are likely to be a paraphrase:

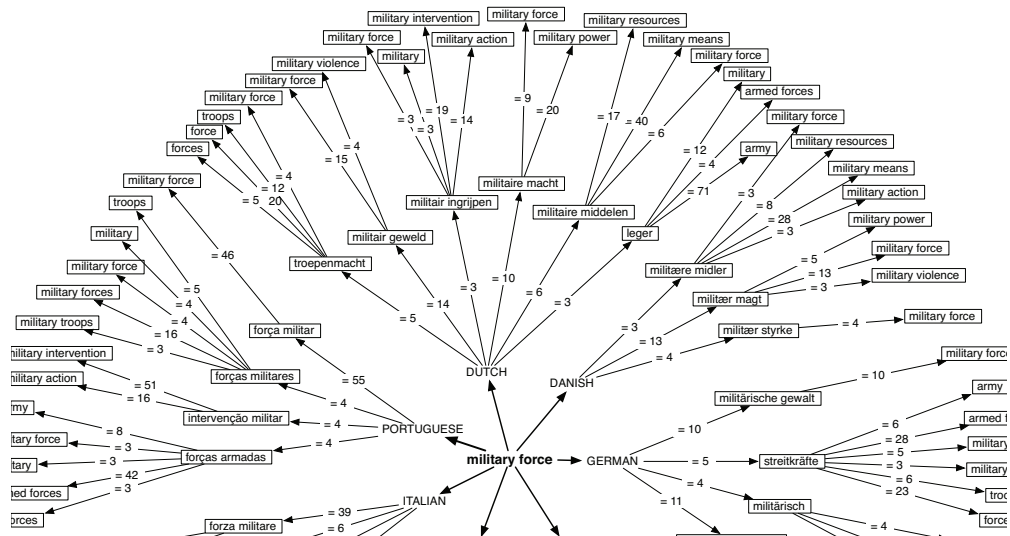$$P(e_2|e_1) = \sum_F P(e_2|f)P(f|e_1)$$

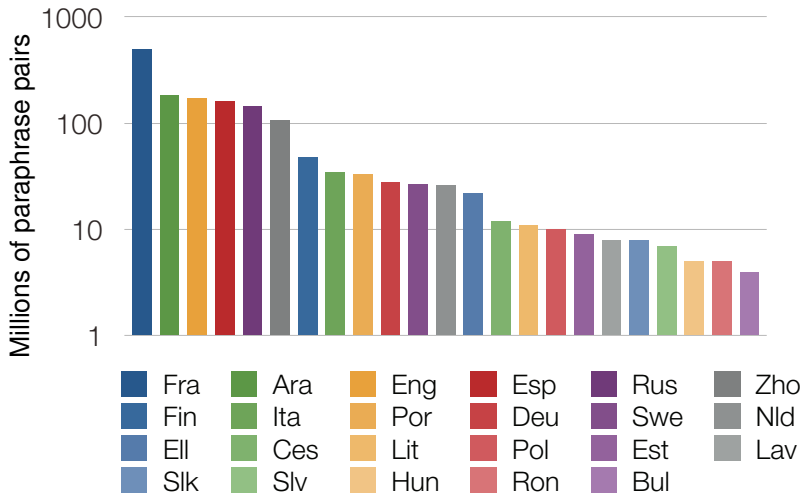# Bilingual Pivoting: An Example

Source: Chris Callison-Burch

- paraphrases extracted from phrase table of SMT system
- lexical and syntactic (aka rules)
- enriched with annotations for specific tasks (e.g., entailment, simplification)
- http://paraphrase.org

Bilingual Pivoting: An Example

Source: Chris Callison-Burch

PPDB: The Paraphrase Database

Source: Ganitkevitch and Callison-Burch, 2014

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

The Present

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

Neural Machine Translation

Source: Synced Review

Time	Company	NMT Implementation	Framework	Characteristic
2016.09	Google	GNMT	Tensorflow	State of the art industrial implementation of the "attentional encoder-decoder networks" model
2016.11	Microsoft	No technical details disclosed		
2017.05	Facebook	Fairseq	Torch	Used CNN to replace RNN
2017.06	Google	Transformer	Tensorflow/Tensor2Tensor	Solely attention based NMT
2017.07	Amazon	Sockeye	MXNet	

Other companies took part in the NMT R&D include: IBM, NVIDIA, SYSTRAN

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

NMT and Multiple Pivots

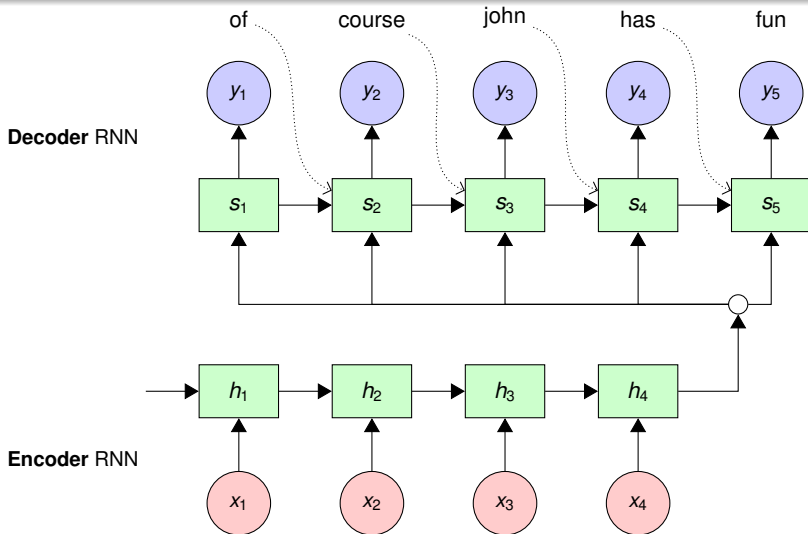## Neural machine translation

- context-aware, places greater emphasis on semantics;
- allows for multiple pivots with multiple models;
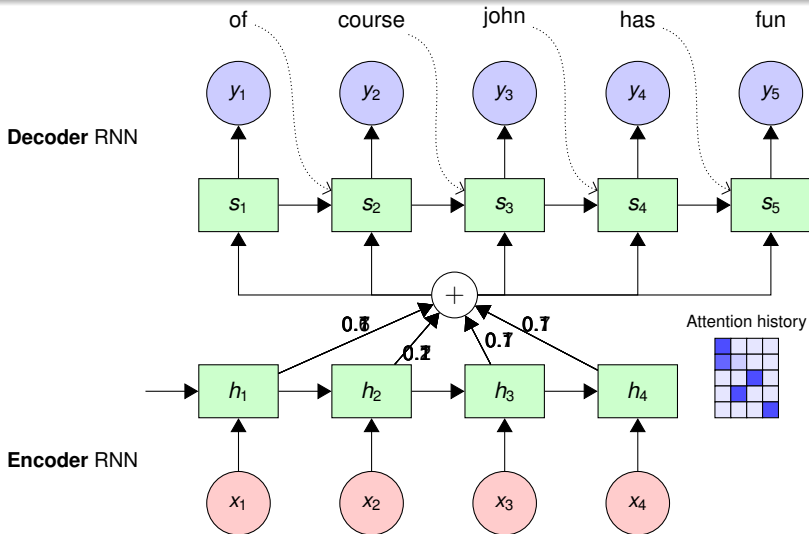- assigns a probability to arbitrary sequences.

Relevance to paraphrases

- represent paraphrases in continuous space;
- NMT learns embeddings for words and phrases naturally;
- estimate semantic relatedness between text segments of arbitrary length;
- generate candidate paraphrases for any source input.

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

Encoder-Decoder Architecture

Decoder RNN

Encoder RNN

Encoder-Decoder with Attention

What is Paraphrasing
The Past
The Present

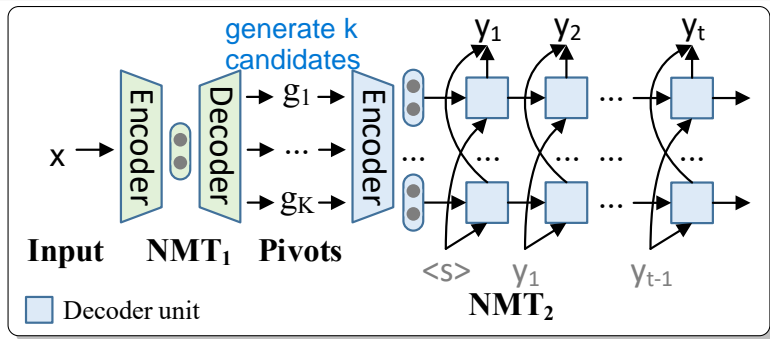**NMT and PARANET**
Sentence Compression
Question Answering

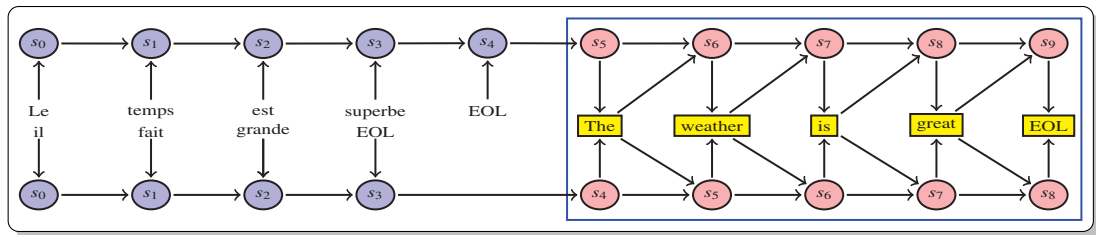# PARANET

generate k candidates

source to pivots to source

Source is translated into *K*-best list of foreign pivots $\mathcal{G}_x = \{g_1, \ldots, g_K\}$.

$$p\left(x'|\mathcal{G}_x\right) = \prod_{t=1}^{|x'|} p\left(y_t|y_{<t}, \mathcal{G}_x\right) = \prod_{t=1}^{|x'|} \sum_{k=1}^{K} p\left(g_k|x\right) p\left(y_t|y_{<t}, g_k\right)$$

PARANET: Multiple Pivots

- **Late-weighted combination** (Firhat et al., 2016)
- Pivot sentences are encoded separately by two encoders
- At each time step, two decoders produce a probability distribution over all words, which are then combined (in the yellow square)
- From this combined distribution a word is chosen, which is then given as input to each decoder.

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

Pivoting over Multiple Languages

- Pivot over sentences from multiple language pairs
- Translate x into K-best German $\mathcal{G}^{\mathcal{DE}}$ and French $\mathcal{G}^{FR}$ sentences
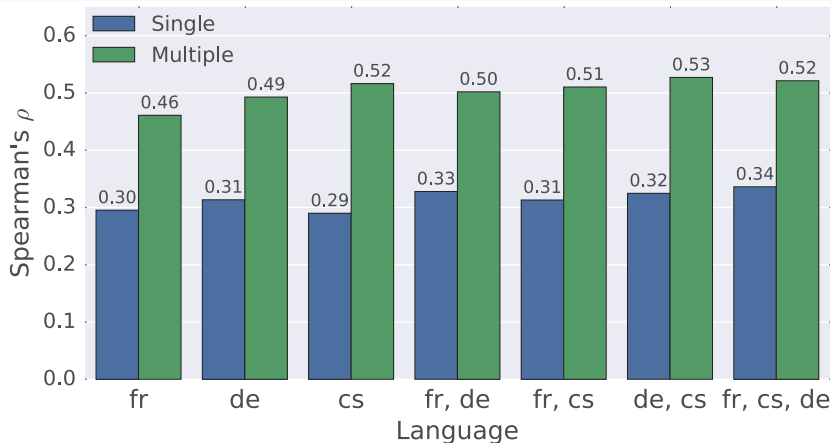- Then average distributions

$$P(x'|\mathcal{G}^{DE}, \mathcal{G}^{FR}) = \prod_{t=1}^{|x'|} p(y_t|y_{<t}, \mathcal{G}^{DE}, \mathcal{G}^{FR})$$

$$p(y_t|y_{<t}, \mathcal{G}^{DE}) = \sum_{k=1}^{K} p\left(g_k^{\mathcal{DE}}|x\right) p\left(y_t|y_{<t}, g_k^{\mathcal{DE}}\right)$$

$$p(y_t|y_{<t}, \mathcal{G}^{FR}) = \sum_{k=1}^{K} p\left(g_k^{\mathcal{FR}}|x\right) p\left(y_t|y_{<t}, g_k^{\mathcal{FR}}\right)$$

$$p(y_t|y_{<t}, \mathcal{G}^{DE}, \mathcal{G}^{FR}) = \lambda_1 \sum_{k=1}^{K} p\left(g_k^{\mathcal{DE}}|x\right) p\left(y_t|y_{<t}, g_k^{\mathcal{DE}}\right) + \lambda_2 \sum_{k=1}^{K} p\left(g_k^{\mathcal{FR}}|x\right) p\left(y_t|y_{<t}, g_k^{\mathcal{FR}}\right)$$

What is Paraphrasing
The Past
The Present

NMT and PARANET
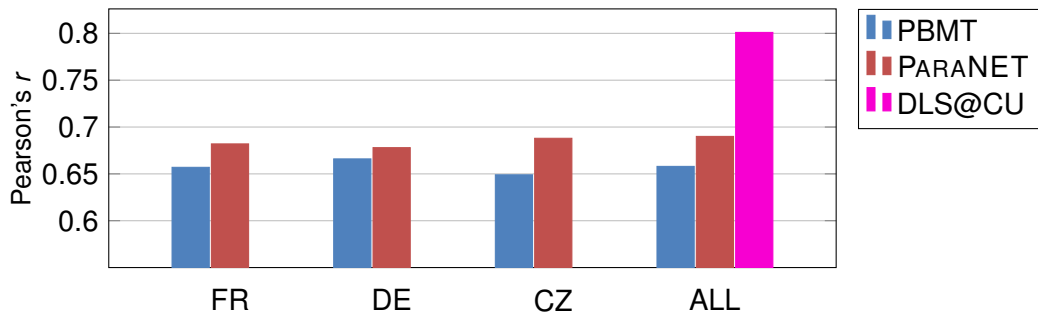Sentence Compression
Question Answering

Results: One vs Many Pivots

- PPDB 2.0 dataset: humans rate (1–5) paraphrase pairs (Pavlick et al. 2015);
- Spearman's ρ between ratings and PARANET's $P(e_2|e_1)$.

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

Results: How Good are the Paraphrases?
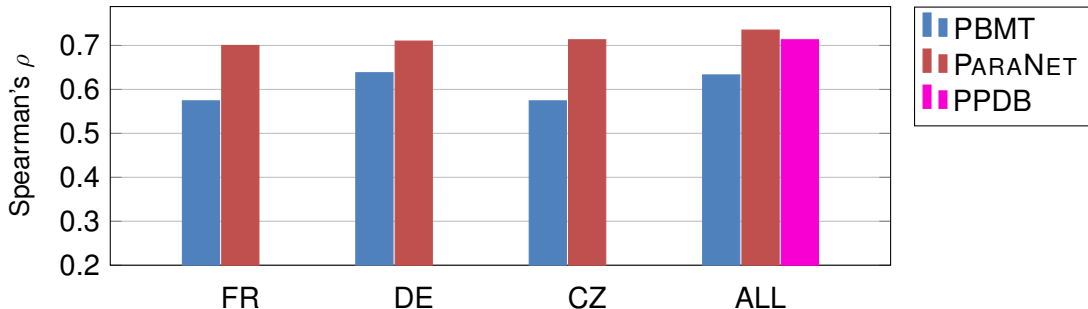
- Supervised scoring model (Pavlick et al., 2015)
- Features: sentence length,1-4 gram string similarity, paraphrase probability $p(e_2|e_1)$, the language model score, cosine distance of paraphrase vectors, attention scores.

What is Paraphrasing
The Past
The Present
NMT and PARANET
Sentence Compression
Question Answering

Results: How Good are the Embeddings?

- SemEval-2015 Task 2: Semantic Textual Similarity
- Predict similarity between two sentences on a 1–5 scale
- News headlines, image captions, student answers, answers to question in public forums and sentences expressing commited belief

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

What about Generation?

WikiAnswers (Fader et al., 2013)

Beauty is not in the eye of the beholder.
Beauty is not in the mind of the viewer.

What is the importance of employee satisfaction in an organization?
What is the significance of staff satisfaction at an organisation?

2000 Leagues under the Sea (Barzilay and McKeown, 2001)

"I owed myself this revenge!" Said the Captain to the Canadian.
"I am indebted to this revenge!" the captain told the Canadian.

"That's what I've been telling you Ned"
"That's what I said, Ned".

Results: What about Generation?
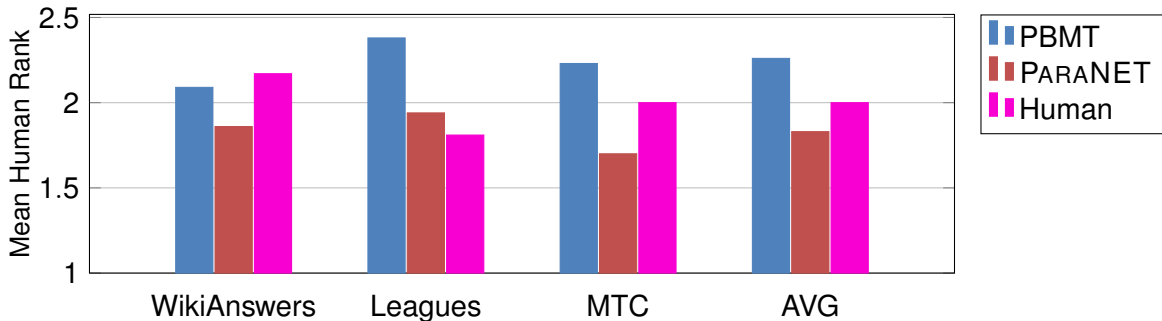
Multiple Translation Chinese (MTC) corpus

We will accelerate the drafting of telecommunications legalization, amend the law of post and the regulations governing wireless telecommunications.

We will speed up the design of telecommunications, change the law and regulations governing wireless telecommunication.

(London, AP) The British government is working on resolving the increasingly serious problems of street crimes and will strengthen patrolling police.

London, AP The British government is working to resolve the increasingly serious problems of street crime and will strengthen patrols.

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

Results: What about Generation?

- Human evaluation using Mechanical Turk
- Participants see source sentence and rank paraphrases produced by PBMT, PARANET, and Humans.
- 100 sentences per domain, 5 judgments per item

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

Text Rewriting, No Problem!

Sentence compression: produces a summary of a single sentence by using less words, preserving the most important information, and remaining grammatical.
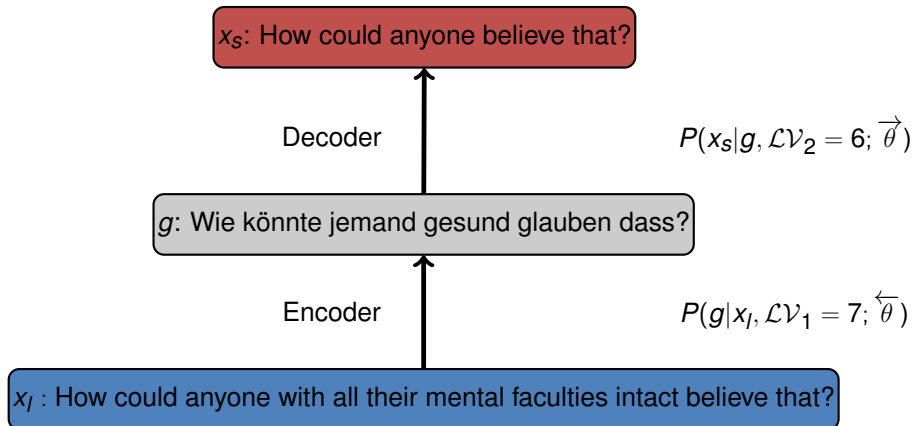
Sentence compression: main idea

- use PARANET as back-bone for rewriting;
- Pivot through multiple languages extracting latent compression information;
- Incorporate length information into the decoder.
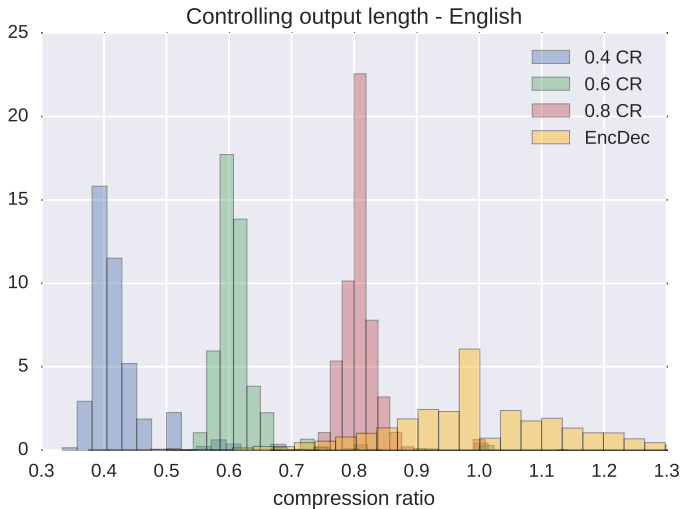
Length embedding (Kikuchi et al., 2016)

Initialise the decoder, not only with the encoder's hidden state but also a length embedding, which is then scaled according to the target length.

Pivoting and Compression

$$P(x_s|x_l, g, \mathcal{LV}_1, \mathcal{LV}_2) \triangleq P(g|x_l, \mathcal{LV}_1)P(x_s|g, \mathcal{LV}_2)$$

x_s: How could anyone believe that?

Decoder $\qquad P(x_s|g, \mathcal{LV}_2 = 6; \overrightarrow{\theta})$

g: Wie könnte jemand gesund glauben dass?

Encoder $\qquad P(g|x_l, \mathcal{LV}_1 = 7; \overleftarrow{\theta})$

x_l : How could anyone with all their mental faculties intact believe that?

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

Parameterizing via Length

Controlling output length - English

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

Parameterizing via Length

Controlling output length - German

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

Parameterizing via Length

Controlling output length - French

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

Compression Output in English

It was through him that the scandal broke in 2011, in an in-depth investigation into corruption related to road construction contracts in Quebec, to which the liberal Prime Minister at the time, Jean Charest, had consented only reluctantly. (39 tokens)

He met him. (4 tokens)

The scandal was revealed by him due to an in-depth investigation into the Quebec Treaties. (16 tokens)

The scandal was revealed by him due to an in-depth investigation into the contracts related to the contracts for roadworks contracts. (22 tokens)

The scandal was revealed by him due to an in-depth investigation into the roadworks contracts in Quebec, which the then Prime Minister reluctantly agreed to. (27 tokens)

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

Compression Output in German

Wie Sie sicher aus der Presse und dem Fernsehen wissen, gab es in Sri Lanka mehrere Bombenexplosionen mit zahlreichen Toten. (21 tokens)

Sie wissen aus der Presse und dem Fernsehen, dass es in Sri Lanka eine Reihe von Bombenexplosionen und -Morden gegeben hat. (23 tokens)

Sie wissen aus Presse und Fernsehen, dass es in Sri Lanka zahlreiche Bombenanschläge und Morde gegeben hat. (19 tokens)

Die Presse und das Fernsehen haben zahlreiche Bombenexplosionen und Morde in Sri Lanka. (13 tokens)

In Sri Lanka sind zahlreiche Bombenexplosionen aufgetreten. (8 tokens)
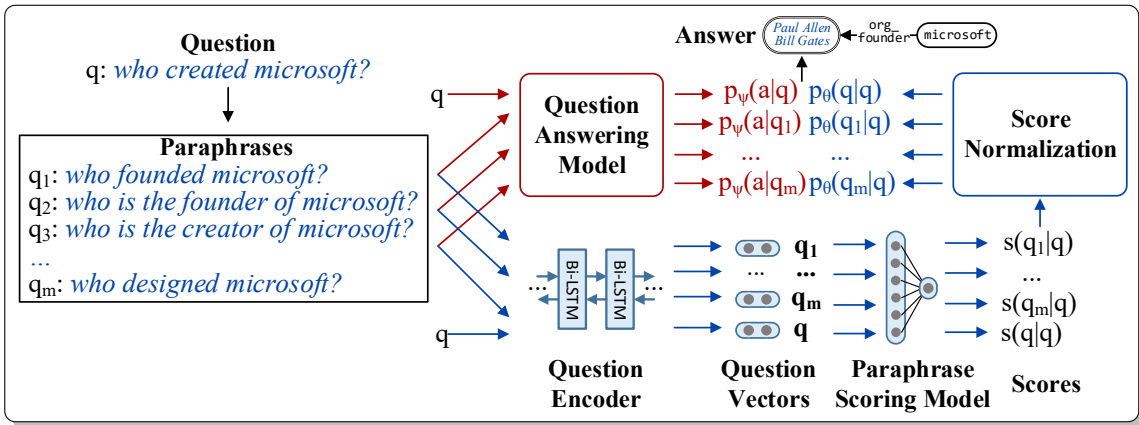
Paraphrases for Question Answering

An Old Idea: paraphrase question, submit paraphrases to QA module
(Duboue and Chu-Carroll, 2006; Narayan et al., 2016; Chen et al., 2016;
Fader et al, 2013; Berant et al., 2014; Bordes et al., 2014a,b; Dong et
al., 2015).

A Common Problem: generated paraphrases often contain inappro-
priate candidates; scoring takes place indepen-
dently of QA module.

A New Solution: train paraphrase model jointly with QA-model for
end tasks.

Problem Formulation

Problem Formulation

$$p\left(a|q\right) = \sum_{q' \in H_q \cup \{q\}} \underbrace{p_\psi\left(a|q'\right)}_{\text{QA Model}} \quad \underbrace{p_\theta\left(q'|q\right)}_{\text{Paraphrase Model}}$$

H_q is the set of paraphrases for question q; ψ, θ parameters of QA and paraphrase scoring model.

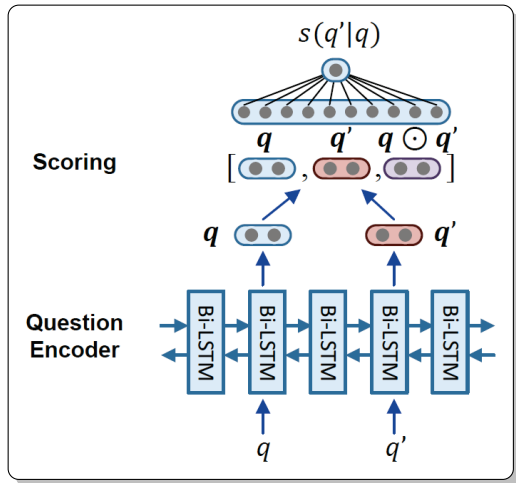**Training**

$$\max \sum_{(q,a) \in \mathcal{D}} \log p(a|q)$$

Inference

$$\hat{a} = \arg\max_{a' \in C_q} p(a'|q)$$

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

Paraphrase Model

- We need to generate paraphrases for every question.
- Various options (rules, WordNet, PPDB).
- We turn to neural Machine Translation.
- Any NMT engine will do, we use `http://opennmt.net/`

English Paraphrases via Different Pivots

What is the zip code of the largest car manufacturer?

What is the zip code of the largest vehicle manufacturer?

What is the zip code of the largest car producer?

What's the postal code of the biggest automobile manufacturer?

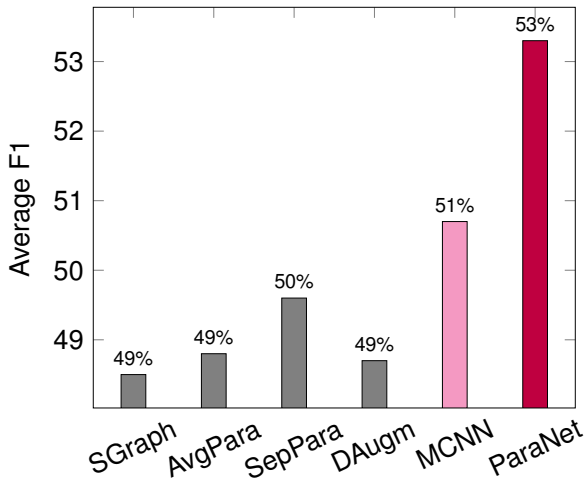**What is the postcode of the biggest car manufacturer?**

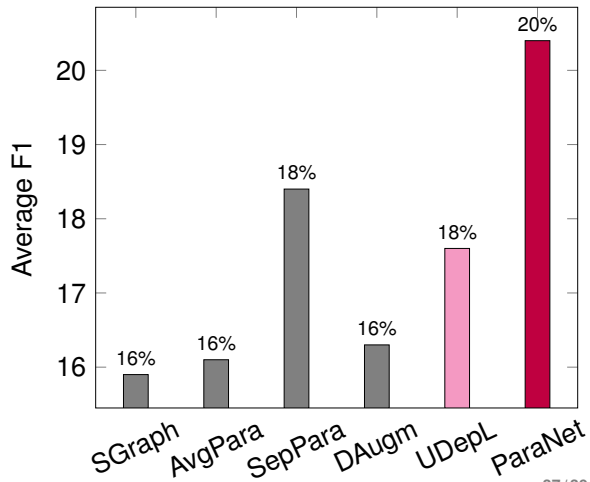**What is the largest car manufacturer's postal code?**

Zip code of the largest car manufacturer

Experimental Results

WEBQUESTIONS

GRAPHQUESTIONS

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

Conclusions

I ❤ Neural Machine Translation!

- There's tons of data in multiple languages!
- New models can utilize this data for specific tasks.
- Sentence compression: NMT is the compression model.
- QA: the model uses paraphrases without NMT.
- As NMT gets better our paraphrasing models will get better too.

What is Paraphrasing
The Past
The Present

NMT and PARANET
Sentence Compression
Question Answering

Ways to Say I Love You

French

Je vous aime.
J'adore vous.
Je t'aime beaucoup.
Je vous aime
beaucoup.
Je vous aime bien.
Je vous remercie.
Je t'aime bien.

German

Ich liebe Dich.
Ich liebe Euch.
Ich liebe Sie.
Ich liebe Dich selbst.
Ich liebe Dich sehr.
Ich liebe dich
jedenfalls.
Ich liebe dich doch.

English

I love you guys.
I do love you.
I just love you.
Me love you.
I love thee.
I really love you.
I love you too.
Well , I love you.