Large-scale Paraphrasing for Natural Language Generation

Chris Callison-Burch
November 21, 2014

with Juri Ganitkevitch, Benjamin Van Durme, Ellie Pavlick, Wei Xu, Courtney Napoles, Xuchen Yao, Peter Clark, Jonny Weese, Matt Post, Tsz Ping Chan, Rui Wang, Trevor Cohn, Mirella Lapata and Colin Bannard
Paraphrases

Differing **textual** expressions of the same meaning:

\[
\begin{align*}
\text{cup} & \iff \text{mug} \\
\text{the king’s speech} & \iff \text{His Majesty’s address} \\
X_1 \text{ devours } X_2 & \iff X_2 \text{ is eaten by } X_1 \\
on \text{ JJ instance of NP} & \iff \text{ a JJ case of NP}
\end{align*}
\]
Paraphrasing in NLP

Recognition or generation of paraphrases plays a part in...

...information extraction, question answering, entailment recognition, summarization, translation, compression, simplification, automatic evaluation of translation or summaries, natural language generation, etc.
Data-Driven Paraphrasing

Monolingual parallel: English – English

Monolingual comparable: English ~ English

Plain monolingual: English

Bilingual parallel: English – French

What a scene! Seized by the tentacle and glued to its suckers, the unfortunate man was swinging in the air at the mercy of this enormous appendage. He gasped, he choked, he yelled: "Help! Help!" I'll hear his harrowing plea the rest of my life! The poor fellow was done for.

What a scene! The unhappy man, seized by the tentacle and fixed to its suckers, was balanced in the air at the caprice of this enormous trunk. He rattled in his throat, he was stifled, he cried, "Help! help!" That heart-rending cry! I shall hear it all my life. The unfortunate man was lost.
Paraphrasing with parallel monolingual data

Barzilay and McKeown (2001) identify paraphrases using identical contexts in aligned sentences:

| Emma burst into tears and he tried to comfort her, saying things to make her smile. |
| Emma cried and he tried to console her, adorning his words with puns. |

burst into tears = cried and comfort = console
Paraphrasing with comparable texts

Dolan, Quirk, and Brockett (2004) extract sentential paraphrases from newspaper articles published on the same topic and date:

<table>
<thead>
<tr>
<th>On its way to an extended mission at Saturn, the Cassini probe on Friday makes its closest rendezvous with Saturn's dark moon Phoebe.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Cassini spacecraft, which is en route to Saturn, is about to make a close pass of the ringed planet's mysterious moon Phoebe.</td>
</tr>
</tbody>
</table>
If we consider oculist and eye-doctor we find that, as our corpus of utterances grows, these two occur in almost the same environments. In contrast, there are many sentence environments in which oculist occurs but lawyer does not...

It is a question of the relative frequency of such environments, and of what we will obtain if we ask an informant to substitute any word he wishes for oculist (not asking what words have the same meaning).

These and similar tests all measure the probability of particular environments occurring with particular elements... If A and B have almost identical environments we say that they are synonyms.

–Zellig Harris (1954)
Lin and Panel (2001) operationalize the Distributional Hypothesis using dependency relationships to define similar environments.

Duty and responsibility share a similar set of dependency contexts in large volumes of text:

<table>
<thead>
<tr>
<th>modified by adjectives</th>
<th>objects of verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>additional, administrative, assigned, assumed, collective, congressional, constitutional ...</td>
<td>assert, assign, assume, attend to, avoid, become, breach ...</td>
</tr>
</tbody>
</table>
My focus:
Paraphrasing & Translation

Translation is re-writing a text using words in a different language.

Paraphrasing is translation into the same language.
Inspiration from Statistical Machine Translation

We reuse & adapt:

Training data + alignment algorithms
Models + feature functions
Parameter estimation
Decoder
Bilingual Data

Sentence-aligned parallel corpora in English and any foreign language

Available in large quantities

Strong meaning equivalence signal

... but different languages.
... 5 farmers were thrown into jail in Ireland ...

... fünf Landwirte festgenommen, weil ...
Large, diverse sets of bilingual training data

- French-English: $10^9$ word webcrawl
- 2 languages @ 250M each (DARPA GALE Program)
- 21 languages @ 50-80M each (European Parliament)
Wide range of paraphrases

thrown into jail

arrested detained imprisoned incarcerated jailed locked up taken into custody thrown into prison

be thrown in prison been thrown into jail being arrested in jail in prison put in prison for were thrown into jail who are held in detention

arrest cases custody maltreated owners protection thrown
Paraphrase Probability

\[ p(e_2|e_1) = \sum_f p(e_2, f|e_1) \]

\[ = \sum_f p(e_2|f, e_1)p(f|e_1) \]

\[ \approx \sum_f p(e_2|f)p(f|e_1) \]
military force

force

armed forces

forces

military forces

military force

peace-keeping personnel

defense
Syntactic constraints

thrown into jail

arrested  
detained  
imprisoned  
incarcerated  
jailed  
locked up  
taken into custody  
thrown into prison

be thrown in prison  
been thrown into jail  
being arrested  
in-jail  
in-prison  
put in prison for  
were thrown into jail  
who are held in detention

arrest  
cases  
custody  
maltreated  
owners  
protection  
thrown

Sentential paraphrases from bitexts?

Bilingual parallel corpora provide an excellent source of lexical and phrasal paraphrases.

Sentential | structural paraphrases are more obviously learned from English-English sentence pairs.

Can we learn structural paraphrases from bitexts? How should we represent them?
Syntactic MT in the Joshua Decoder

• Synchronous context free grammars generate pairs of corresponding strings
• Can be used to describe translation and re-ordering between languages
• Because Joshua uses SCFGs, it translates sentences by parsing them

http://joshua-decoder.org
**Example SCFG for translation**

<table>
<thead>
<tr>
<th>Urdu</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>S → NP① VP②</td>
<td>NP① VP②</td>
</tr>
<tr>
<td>VP→ PP① VP②</td>
<td>VP② PP①</td>
</tr>
<tr>
<td>VP→ V① AUX②</td>
<td>AUX② V①</td>
</tr>
<tr>
<td>PP → NP① P②</td>
<td>P② NP①</td>
</tr>
<tr>
<td>NP → hamd ansary</td>
<td>Hamid Ansari</td>
</tr>
<tr>
<td>NP → nab sdr</td>
<td>Vice President</td>
</tr>
<tr>
<td>V → namzd</td>
<td>nominated</td>
</tr>
<tr>
<td>P → kylye</td>
<td>for</td>
</tr>
<tr>
<td>AUX → taa</td>
<td>was</td>
</tr>
</tbody>
</table>
Hamid Ansari was nominated for Vice President.
Hamid Ansari was nominated for Vice President.
Hamid Ansari

for

Vice President nominated

was
Hamid Ansari

Vice President

for

Vice President

was

nominated
Hamid Ansari was nominated for Vice President.
SCFGs via Pivoting

- Adapting our syntactic MT models, we learn structural transformations, like the English possessive rule

\[
\begin{align*}
\text{NP} & \rightarrow \text{NP }'s \text{ NN} \mid \text{le NN de NP} \\
\text{NP} & \rightarrow \text{the NN of NP} \mid \text{le NN de NP} \\
\end{align*}
\]

combine to

\[
\begin{align*}
\text{NP} & \rightarrow \text{NP }'s \text{ NN} \mid \text{the NN of NP} \\
\end{align*}
\]
<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Possessive rule</strong></td>
<td>NP → the NN of the NNP</td>
</tr>
<tr>
<td></td>
<td>NP → the NNS made by the NNS</td>
</tr>
<tr>
<td><strong>Dative shift</strong></td>
<td>VP → give NN to NP</td>
</tr>
<tr>
<td></td>
<td>VP → provide NP to NP</td>
</tr>
<tr>
<td>**Adv.</td>
<td>adj. phrase move**</td>
</tr>
<tr>
<td></td>
<td>S → it is ADJP VP</td>
</tr>
<tr>
<td><strong>Verb particle shift</strong></td>
<td>VP → VB NP up</td>
</tr>
<tr>
<td><strong>Reduced relative clause</strong></td>
<td>SBAR</td>
</tr>
<tr>
<td></td>
<td>ADJP → very JJ that S</td>
</tr>
<tr>
<td><strong>Partitive constructions</strong></td>
<td>NP → CD of the NN</td>
</tr>
<tr>
<td></td>
<td>NP → all DT\NP</td>
</tr>
<tr>
<td><strong>Topicalization</strong></td>
<td>S → NP, VP</td>
</tr>
<tr>
<td><strong>Passivization</strong></td>
<td>SBAR → that NP had VBN</td>
</tr>
<tr>
<td><strong>Light verbs</strong></td>
<td>VP → take action ADVP</td>
</tr>
<tr>
<td></td>
<td>VP → to make a decision PP</td>
</tr>
</tbody>
</table>

Learning Sentential Paraphrases from Bilingual Parallel Corpora for Text-to-Text Generation.
Juri Ganitkevitch, Chris Callison-Burch, Courtney Napoles, and Benjamin Van Durme. EMNLP 2011.
Text-to-Text Generation

T2T involves generating meaning-equivalent text that is subject to some constraints:

- sentence compression, shorter
- simplification, easier to understand
- poetry from prose, rhyme and meter
Sentence Compression

Reduce length of a sentence (#chars) while retaining the meaning

Compression ratio: \[ \varphi = \frac{\text{length}_{\text{compression}}}{\text{length}_{\text{original}}} \]

Paraphrasing as a task and problem is of paramount importance to a multitude of applications in the field of NLP.
Sentence Compression

Reduce length of a sentence (#chars) while retaining the meaning

Compression ratio: \[ \varphi = \frac{\text{length}_{\text{compression}}}{\text{length}_{\text{original}}} \]

Paraphrasing as a task and problem is of paramount importance to a multitude of applications in the field of NLP.
### Paraphrase Grammar

<table>
<thead>
<tr>
<th>English</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>S → NP(^1) were VBD by NP(^2)</td>
<td>NP(^2) VBD NP(^1)</td>
</tr>
<tr>
<td>NP → NP that VP</td>
<td>NP VP</td>
</tr>
<tr>
<td>VP → are JJ to NP</td>
<td>JJ NP</td>
</tr>
<tr>
<td>NP → CD of the NNS</td>
<td>CD NNS</td>
</tr>
<tr>
<td>CD → twelve</td>
<td>12</td>
</tr>
<tr>
<td>NNS → cartoons</td>
<td>comics</td>
</tr>
<tr>
<td>JJ → offensive</td>
<td>insulting</td>
</tr>
<tr>
<td>NP → the islamic prophet</td>
<td>mohammed</td>
</tr>
<tr>
<td>VBD → sparked</td>
<td>caused</td>
</tr>
</tbody>
</table>
riots were sparked by twelve of the cartoons that are offensive to the Islamic prophet Muhammad that were caused by insulting comics.
riots were sparked by twelve of the cartoons that are offensive to the Islamic prophet.

riots caused 12 comics insulting mohammad
riots were sparked by twelve of the cartoons that are offensive to the Islamic prophet Mohammad.
Text-to-Text Applications

Claim:

Paraphrasing is suitable to tackle sentential text-to-text tasks, and we can re-use SMT machinery for T2T.

However:

Naive application of MT techniques will not work, need to adapt them
Task Adaptation

<table>
<thead>
<tr>
<th>SMT</th>
<th>T2T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive application of the MT machinery to the task</td>
<td>Task-specific adaptations</td>
</tr>
</tbody>
</table>

- Development data
- Objective function
- Feature set
- Grammar augmentations
## Development Data

<table>
<thead>
<tr>
<th>SMT</th>
<th>T2T</th>
</tr>
</thead>
<tbody>
<tr>
<td>English reference translations that are used to calculate BLEU for SMT.</td>
<td>Selected pairs of reference translations that significantly differ in length.</td>
</tr>
</tbody>
</table>

and he said that the project will cover the needs of the region in the long term.  

he said the project includes all the district's long-term needs. 

compression ratio = 0.79
## Objective Function

### SMT


### T2T

Add a “verbosity penalty” to BLEU that allows a target compression ratio to be set.

---

### Graph

- **X-axis:** actual CR \& target CR
- **Y-axis:** penalty term
- **Legend:**
  - BLEU
  - PRÉCIS

---

*Note:* The graph illustrates the relationship between actual compression ratio (CR) and target compression ratio, with penalty terms for BLEU and PRÉCIS scoring.
Features

<table>
<thead>
<tr>
<th>SMT</th>
<th>T2T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrasal and lexical probabilities quantify general paraphrase quality.</td>
<td>Features counting number of source and target words and the difference between them.</td>
</tr>
</tbody>
</table>

\[ p(e_1|e_2) = 0.1 \]
\[ c_{e_1} = 14 \]
\[ c_{e_2} = 5 \]
\[ c_{\text{diff}} = -9 \]

\[ \log CR = \log \frac{c_{e_1}}{c_{e_2}} \]

VP → NP was eaten by NN | NN ate NP
Augmentations

<table>
<thead>
<tr>
<th>SMT</th>
<th>T2T</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is not typical for additional task-specific rules to be added in the standard SMT pipeline.</td>
<td>Augment the grammar with deletion rules for specific POS (JJ, RB, DT) allowing for shorter compressions.</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
JJ & \rightarrow \text{superfluous} \mid \varepsilon \\
RB & \rightarrow \text{redundantly} \mid \varepsilon \\
DT & \rightarrow \text{the} \mid \varepsilon
\end{align*}
\]
Monolingually-derived Features

<table>
<thead>
<tr>
<th></th>
<th>SMT</th>
<th>T2T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All features, aside from the LM, are bilingually derived.</td>
<td>Calculate distributional similarity of paraphrase pairs from monolingual data</td>
</tr>
</tbody>
</table>

Orthogonal signal to bilingual pivoting

Even more data available

Incorporated as features in T2T model
Distributional Similarity

Idea: similar words occur in similar contexts.

Characterize words by their contexts

Contexts represented by co-occurrence vectors, similarity quantified by cosine

“Are these paraphrases substitutable?”
Similarity

Easy for lexical & phrasal paraphrases

More involved for syntactic paraphrases

..sip from a **cup** of cocoa..
..a **cup** of coffee.

..sip from a **mug** of cocoa..
..a **mug** of coffee.

cup  \iff  mug

..anxiously awaiting the king’s speech..

..anxiously awaiting His Majesty’s address..

the king’s speech  \iff  His Majesty’s address

?  one JJ instance of NP  \iff  a JJ case of NP
Syntactic Paraphrase Similarity

NN 's NP in the long term

NP \rightarrow

the long-term NP of NN

Syntactic Paraphrase Similarity

Syntactic Paraphrase Similarity

Syntactic Paraphrase Similarity

$sim(r) = \frac{1}{2} \left( sim\left( \text{the long-term} \right) + sim\left( 's \right) \right)$

$\mathbf{\tilde{\operatorname{sign}}}_{ngram}(\text{the long-term}) = \begin{pmatrix} \text{L-achieve} = 25 \\ \text{L-revise} = 43 \\ \text{L-confirmed} = 64 \end{pmatrix} = \begin{pmatrix} \text{R-plans} = 97 \\ \text{R-goals} = 23 \\ \text{R-investment} = 10 \end{pmatrix}$
syntactic context

\[
\tilde{\sigma}_{\text{syntax}}(\text{the long-term}) = \begin{pmatrix}
\text{lex-R-investment} \\
\text{lex-L-on-to} \\
\text{pos-L-IN-TO} \\
\text{pos-L-TO} \\
\text{lex-L-to} \\
\text{dep-det-R-investment} \\
\text{pos-R-NN} \\
\text{dep-amod-R-investment} \\
\text{dep-det-R-NN} \\
\text{dep-amod-R-NN} \\
\text{syn-gov-NP} \\
\text{syn-miss-L-NN}
\end{pmatrix}
\]
Large Monolingual Data Sets

Google n-grams
Collection of 1 trillion tokens with counts
Based on vast amounts of text

Annotated Gigaword (AKBC-WEKEX ’12)
Collection of 4 billion words, parsed and tagged

Task-based Evaluation

Evaluated paraphrases in the context of a T2T compression task.

Compared against a state of the art system.

Human assessment (5-point scale):

How well do these sentences retain the meaning of original?

How grammatical is the resulting sentence?
Compression Quality

![Chart showing compression quality for different methods: Ref., ILP, and Random. The x-axis represents different compression quality levels, and the y-axis shows the quality score. The chart compares grammar and meaning quality. Ref. has the highest quality in both categories, followed by ILP, and then Random.]
Input: Hala speaks Arabic most of the time with her son, taking into consideration that he can speak English with others.

Hala speaks to her son mostly in Arabic as he can speak English to others.

Hala speaks Arabic most of the time, taking into consideration that he can speak English with others.

Hala speaks Arabic most of the time with her son, considering that he can speak English with others.
# Adaptation in 5 easy steps

<table>
<thead>
<tr>
<th>Step</th>
<th>SMT to T2T Adaptation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Dev data:</strong> Collect a set of sentence pairs that reflects the task that you are trying to model</td>
</tr>
<tr>
<td>2</td>
<td><strong>Objective function:</strong> Create a new objective function that indicates how well the system output the constraints of your task</td>
</tr>
<tr>
<td>3</td>
<td><strong>Task-specific features:</strong> Add new features to the model that will allow it to score its own output for the task</td>
</tr>
<tr>
<td>4</td>
<td><strong>Augment the grammar:</strong> Use your domain knowledge to add any rules that would not normally be contained in a paraphrase grammar.</td>
</tr>
<tr>
<td>5</td>
<td><strong>Other features:</strong> Take advantage of the English to English to add other features that model grammaticality more generally.</td>
</tr>
</tbody>
</table>
Resources
Joshua Decoder

• An open source decoder that synchronous context free grammars to translate
• Implements all algorithms needed for translating with SCFGs
  – grammar extraction
  – chart-parsing
  – n-gram LM integration

http://joshua-decoder.org
Machine Translation Class

- Developed w/ Adam Lopez, Matt Post and Chris Dyer
- Project based class
- Students solve real open research problems in MT
- Projects are automatically gradable, MOOC ready

http://mt-class.org
PPDB: The Paraphrase Database

- A huge collection of paraphrases
- Extracted from 106 million sentence pairs, 2 billion English words, 22 pivot languages

<table>
<thead>
<tr>
<th>Type</th>
<th>Paraphrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical</td>
<td>7.6 M</td>
</tr>
<tr>
<td>Phrasal</td>
<td>68.4 M</td>
</tr>
<tr>
<td>Syntactic</td>
<td>93.6 M</td>
</tr>
<tr>
<td>Total</td>
<td>169.6 M</td>
</tr>
</tbody>
</table>

http://paraphrase.org
large amount

1. enormous amount
   Noun phrase missing determiner on the left

2. tremendous amount
   Noun phrase missing determiner on the left

3. huge sum
   Noun phrase missing determiner on the left

4. enormous number
   Noun phrase missing determiner on the left

5. huge number
   Noun phrase missing determiner on the left

6. awful lot
   Noun phrase missing determiner on the left

7. massive amount
   Noun phrase missing determiner on the left

Total search results: 129
Do the Scores Work?

Human Score

great

terrible

PPDB Score

expect | harbour

expect | hope

high recall

high precision
Fun PPDB Examples

munchies ||| hungry

Summary

Extraction & Representation

Extended large-scale paraphrase acquisition from bitexts to syntactic paraphrases

Generation

Introduced a straightforward and effective adaptation framework

Extensions beyond SMT

Improved performance by using monolingual information
Current directions

Domain-specific paraphrasing

What if we want to generate paraphrases for specific domains like biology? Do they vary? How do we ensure ours are appropriate?

Polysemy of paraphrases

Our method sometimes groups paraphrases that correspond to different senses of the input phrase. How can we partition them into sets?

Paraphrase recognition and entailment

The RTE problem diverges in interesting ways from paraphrasing. We are combining natural language inference and data-driven paraphrasing.
Word Sense

bug

- microbe, virus, bacterium, germ, parasite
- insect, beetle, pest, mosquito, fly
- bother, annoy, pester
- glitch, error, malfunction, fault, failure
- microphone, tracker, mic, wire, earpiece, cookie
- squealer, snitch, rat, mole
Textual Inference

hypothesis

twelve insulting illustrations

the prophet

caused unrest

riots in Denmark were sparked by 12 offensive editorial cartoons that were to Muhammad

text
## Attaching a Semantics

<table>
<thead>
<tr>
<th>twelve</th>
<th>12</th>
<th>equivalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>cartoons</td>
<td>illustrations</td>
<td>forward entailment</td>
</tr>
<tr>
<td>caused</td>
<td>prevented</td>
<td>negation</td>
</tr>
<tr>
<td>Europe</td>
<td>the middle East</td>
<td>alternation</td>
</tr>
</tbody>
</table>

- **Riots in Greece**: caused [*Civil unrest in Europe*] → *Civil unrest in Europe*
- *Civil unrest in Europe* → *Riots in Greece*
Thank you!

diet coke

Thank you for your time
many thanks
here you go anyway, thanks
leave a message gee, thanks
thanks, man you look amazing bless you
don't thank me
keep the change

thank you very much

why, thank you

thank you for your attention
uh, thanks

thank you, frank


ParaMetric: An Automatic Evaluation Metric for Paraphrasing. Chris Callison-Burch, Trevor Cohn and Mirella Lapata. COLING 2008


## Entailment relations

<table>
<thead>
<tr>
<th>Hypernym</th>
<th>Synonym</th>
<th>Antonyms</th>
<th>Alternations</th>
<th>Independent</th>
</tr>
</thead>
<tbody>
<tr>
<td>beetle</td>
<td>insect</td>
<td>icebox</td>
<td>refrigerator</td>
<td>advantage</td>
</tr>
<tr>
<td>honeybee</td>
<td>bee</td>
<td>impasse</td>
<td>deadlock</td>
<td>competence</td>
</tr>
<tr>
<td>fees</td>
<td>spending</td>
<td>infirmary</td>
<td>hospital</td>
<td>continuity</td>
</tr>
<tr>
<td>know-how</td>
<td>knowledge</td>
<td>insurrection</td>
<td>revolt</td>
<td>inflow</td>
</tr>
<tr>
<td>pond</td>
<td>lake</td>
<td>jewel</td>
<td>gem</td>
<td>insanity</td>
</tr>
<tr>
<td>fertilizer</td>
<td>manure</td>
<td>john</td>
<td>lavatory</td>
<td>legitimacy</td>
</tr>
<tr>
<td>actor</td>
<td>entertainer</td>
<td>kale</td>
<td>cabbage</td>
<td>niece</td>
</tr>
<tr>
<td>actor</td>
<td>performer</td>
<td>labyrinth</td>
<td>maze</td>
<td>descendants</td>
</tr>
<tr>
<td>acquisition</td>
<td>buying</td>
<td>laundry</td>
<td>washing</td>
<td>husbands</td>
</tr>
</tbody>
</table>