“Automatic Language Translation Generation Help Needs Badly”

or:

“Can a computer compress a text file without knowing what a verb is?”

Kevin Knight

Information Sciences Institute
University of Southern California

Computer Science Department
University of Southern California

UCNLG+MT, Copenhagen, September, 2007
Gobbledygook

• MT systems output gobbledygook

• Wikipedia:
  – **Gobbledygook** is an *English* term used to describe *nonsensical* language, sound that resembles language but has no meaning.
  – The term was coined in **1944** by Maury *Maverick*, chairman of the United States Smaller War Plants Corporation.
  – Maverick’s inspiration, he said, was the *turkey*, “always gobbledy gobbling and strutting with ludicrous pomposity. At the end of his gobble, there was a sort of gook.”
MT Samples
(best Chinese/English statistical MT system)

• The export of high-tech products, and is an important contribution to the growth of foreign trade in Guangdong.

• South Korea hopes to resume talks will be communicated to the DPRK’s information.

• The South Korean government has conveyed this message to the Indonesian authorities through diplomatic channels.
MT Samples

• She said that based on this situation, United Nations relief agencies have begun to help victims rebuild their homes.

• Iraq made the above remarks at a hearing held by the European Parliament in Brussels.

• On Turkey’s position on the Middle East issue, he said that the Middle East is of important strategic significance to Turkey, and Turkey’s hopes for peace and stability in the region.
MT Samples

- This fighting on the morning of the 28th of last month, the United Nations team investigating the circumstances of the Liberian government to trade diamonds for weapons and imposed sanctions against Liberia.

- The German government on the 4th, three pieces of the Berlin Wall, the symbol of the East-West Cold War presented to the United Nations.
Amazing...

• In translation, *there are so many right answers*...
• All we have to do is hit one of them!
What’s Going Wrong?

• Unknown words?
• Insufficient translation dictionaries?
• Not enough target language knowledge.
  – My native language is English
  – I get along in Spanish
  – I can comprehend both
  – I can translate Spanish → English well
  – I can’t translate English → Spanish well
• The main problem is generation.
MT Pyramid

interlingua

semantics

syntax

words

SOURCE

TARGET
MT Pyramid

- **SOURCE**
  - words
  - syntax
  - semantics
  - woman(x) & pregnant(x)

- **interlingua**
  - woman(x) & baby(y) & inside(y, x)

- **TARGET**
  - words
  - syntax
  - semantics
Generation is half the job!

So, what can we be doing?
I’m Ready for My Input, Mr. DeMille.
“I’m Ready for My Input, Mr. DeMille”

• This doesn’t take us anywhere

• Plus, we’re not ready
  – Imagine the input arrives…
    • Structured representation
    • Large inventory of concepts and relations
  – How can we be ready by 2020?
    • Ontobank?
    • Ontobank for MT?
    • Ontobank for NLG?
  – Even just for syntax…
    • Parallel data
    • Human parsed (both sides)
    • Human word-aligned

Very important to do work in the context of parallel texts.
What does English generalized possession map onto in Chinese?
Plus, we’re getting input already!
People are tunneling through the MT pyramid
MT Pyramid

interlingua

semantics

syntax

words

SOURCE

Brown et al 93, Och 99, Koehn et al 03

TARGET

words
What hunger have I,
Hungry I am so,
I am so hungry,
Have I that hunger …

Que hambre tengo yo

I am so hungry
MT Pyramid

Interlingua

Syntax

Semantics

Source

Target

Galley et al 04
Venugopal & Zollmann 06
Brown et al 93, Och 99, Koehn et al 03
MT Pyramid

words

interlingua

semantics

syntax

SOURCE

TARGET

Quirk et al 05, Huang et al 06, Liu et al 06

Galley et al 04
Venugopal & Zollmann 06

Brown et al 93,
Och 99,
Koehn et al 03
MT Pyramid

People who tunnel get the accuracy they deserve!?
By the way, can MT get away without semantics?

- “People understand the meaning of the source sentence, then generate the target language.”

- How to translate “transformational grammar” into Spanish?
  - Hmm, what kind of grammar is transformational grammar?
  - At the very least, I need to know what is being transformed…?
  - Aw, forget it: “grammática transformacionál”
So, what is this talk going to be about, anyway?

I actually have three topics:

1. Transformation formalisms
2. Language models
3. Software tools

as relate to NLG and MT
1st topic: Transformation Systems

- NLG is a mapping problem (or a series of them)

- Capture possibly-infinite set of <input, output> pairs
  - e.g., sentence s89383 is (or is not) a realization of meaning m5634

- Can extend this to capturing sets of <input, output, probability> triples
  - e.g., sentence s89383 is the most likely realization...

- Formal devices exist to concisely encode such infinite sets of tuples
  - Finite-state transducers
  - Tree transducers
  - Synchronous grammars
  - Rewrite systems
Top-Down Tree Transducer

Original input:

```
S
  NP
    PRO
      he
    VP
      VBG
        listening
      P
        to
      NP
        music
  VP
    PRO
      he
    VBG
      enjoys
    SBAR
    NP
    VP
      NP
        listening
      P
        to
      NP
        music
```
Top-Down Tree Transducer

Original input:

Transformation:
Top-Down Tree Transducer

Original input:
S
 NP  VP
 PRO  VBG  NP
 he  listening  SBAR
 enjoys
 VP  P  NP
 listening  to  music

Transformation:
S
 NP  VP
 PRO  VBG  NP
 he  listening  SBAR
 enjoys
 VP  P  NP
 listening  to  music
Top-Down Tree Transducer

Original input:  Transformation:

S               NP
   NP       VP
      PRO  VBZ  NP
     he  enjoys  SBAR
       VBG  VP
      listening  P  NP
        to  music

kare, wa, SBAR, ga, enjoys

NP  VP
   VBZ
      VBG  NP
     listening  P  to
        to  music
Top-Down Tree Transducer

Original input:  Final output:

he enjoys listening to music

kare, wa, ongaku, o, kiku, no, ga, daisuki, desu
Top-Down Tree Transducer

Original input:

Transformation:
Top-Down Tree Transducer

Original input:

```
S
  NP  VP
    PRO VBZ NP
      he   enjoys  SBAR
        VBG  VP
          listening  P  NP
            to  music
```

Transformation:

```
q S
  x0:NP  VP
       x1:VBZ  x2:NP
0.2
  \rightarrow  s x0, wa, r x2, ga, q x1
```

```
q S
  NP  VP
    PRO VBZ NP
      he  enjoys  SBAR
        VBG  VP
          listening  P  NP
            to  music
```
**Top-Down Tree Transducer**

**Original input:**

```
S
  NP
    PRO
      he
    VP
      VBG
        listening
      P
        to
    SBAR
      NP
        enjoys
      VP
```

**Transformation:**

```
S
  NP
    s
      PRO
      he
    VP
      VBG
        listening
      P
        to
      NP
        music
    SBAR
      r
        wa
      SBAR
        ga
      VP
        enjoys
```

```
NP
  VP
    q
      VBZ
        enjoys
  NP
    VP
      VBG
        listening
      P
        to
      NP
        music
```
Top-Down Tree Transducer

Original input:

Transformation:
Top-Down Tree Transducer

Original input:

Transformation:
Top-Down Tree Transducer

Original input:  
Final output:

To get total probability, multiply probabilities of the individual steps.
Tree Transducers are Expressive

Phrasal Translation

Non-constituent Phrases

Non-contiguous Phrases

Context-Sensitive Word Insertion

Multilevel Re-Ordering

Lexicalized Re-Ordering

also QA, compression, paraphrasing, etc
most probabilistic tree-based models proposed 2000-2005 can be so cast
Limitations of the Top-Down Transducer Model

*Who* does John think Mary believes I saw?  
John thinks Mary believes I saw *who*?
Limitations of the Top-Down Transducer Model

Who does John think Mary believes I saw? \( \rightarrow \) John thinks Mary believes I saw who?
Limitations of the Top-Down Transducer Model

Who does John think Mary believes I saw?  →  John thinks Mary believes I saw who?
Limitations of the Top-Down Transducer Model

*Whose blue dog* does John think Mary believes I saw?  →  John thinks Mary believes I saw *whose blue dog?*

```
Whose blue dog
  /
  /
/    /
DT  ADJ  N

q S
  /
  /
/    /
WhNP  VP
  /
  /
/    /
ADJ  N

S
  /
  /
/    /
V

NP
  /
  /
/    /
John

V

S
  /
  /
/    /
think

NP
  /
  /
/    /
Mary

V

S
  /
  /
/    /
believes

NP
  /
  /
/    /
I

V

V

S

VP

Can't do this

saw

WhNP
  /
  /
/    /
ADJ  N

whose blue dog
```
Advantages of Using Transformation Systems

• Efficient algorithms off the shelf
  – e.g., enumerate the 5,000 most probable realizations of meaning m3425
  – e.g., train the <probability> part on
    <input, output> training pairs
# String & Tree Automata

<table>
<thead>
<tr>
<th>String Automata</th>
<th>Tree Automata</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-best …</td>
<td>… paths through a lattice (Viterbi 67; Eppstein 98)</td>
</tr>
<tr>
<td><strong>EM training</strong></td>
<td>Forward-backward EM (Baum &amp; Welch 71)</td>
</tr>
<tr>
<td>Determinization …</td>
<td>… of weighted string acceptors (Mohri 97)</td>
</tr>
<tr>
<td>Intersection</td>
<td>WFSA intersection</td>
</tr>
<tr>
<td>Applying transducers</td>
<td>string → WFST → WFSA</td>
</tr>
<tr>
<td>Transducer composition</td>
<td>WFST composition (Pereira &amp; Riley 96)</td>
</tr>
</tbody>
</table>

- **String & Tree Automata**
  - **String Automata**
    - N-best paths through a lattice (Viterbi 67; Eppstein 98)
    - EM training: Forward-backward EM (Baum & Welch 71)
    - Determinization: … of weighted string acceptors (Mohri 97)
  - **Tree Automata**
    - … trees in a forest (Huang & Chiang 05)
    - Tree transducer EM training (Graehl & Knight 04)
    - … of weighted tree acceptors (Borchart & Vogler 03)
    - Tree acceptor intersection (despite CFG not closed)
    - Tree → TT → weighted tree acceptor
Another Advantage: Reversibility

- E.g., enumerate all meanings that can be realized as sentence s54673
- Statistical generators like Nitrogen & Halogen implicitly encode a set of <input, output, probability> triples, but can only give answers in one direction
- If the generation system were reversible, we could build a paraphrasing system:
  
  sentence $\rightarrow$ top meaning $\rightarrow$ top-100 sentences

Why build a paraphrasing system?

Speech translation
- is current user utterance a paraphrase of something in my (large) phrasebook, which I already know how to translate accurately?
Reversibility

- If the generation system were reversible, we might be able to build a better generator:
  - Convert meaning to sentence
  - Ask what that sentence means (reverse)
  - Does *what I’m saying* mean *what I think it means*?
Transformation Systems

- MT systems are routinely built on reversible formalisms
  - Phrase substitution/transformation systems
  - Synchronous grammars
  - Tree transducers
  ...

Transformation Systems

• Reversible: same set of 50m transducer rules
  – Chinese strings $\rightarrow$ English trees (Galley et al 04)
  – English trees $\rightarrow$ Chinese strings (Huang, Knight, Joshi 06)

• Noisy channel
  – $P(e | f) \sim P(e) \times P(f | e)$

• Phrase-based MT systems chained together for paraphrasing (Bannard & Callison-Burch 05)
  – $P(e_2 | e_1) \sim P(e_2) \times \sum_f P(f | e_1) \times P(e_2 | f)$

• No one has yet tried to improve one-way translation by translating back & checking…
Transformation Systems

• (Wong & Mooney 06)
  – Does semantic analysis (text to meaning)
  – Database query and other domains
  – Uses synchronous grammars
    • used for something like what (Shieber 90) proposed!
  – Automatically trains on input/output pairs

• (Wong & Mooney 07)
  – Tackles reverse problem of generation

• These systems work in multiple languages
• Possibilities for interlingua-based MT
2nd topic: Language Models

Spanish/English Bilingual Text → Statistical Analysis

Spanish → TRANSLATION MODEL → Broken English

English Text → Statistical Analysis

What hunger have I,
Hungry I am so,
I am so hungry,
Have I that hunger …

Que hambre tengo yo → I am so hungry
Language Models

- Statistical MT overgenerates
- A target language model contributes a score to help grade candidate translations
  - typically $P(\text{string})$
  - typically n-gram model
    \[ P(\text{string}) = P(w_1) \times P(w_2 \mid w_1) \times P(w_3 \mid w_2) \times \ldots \]
- Lexical selection and word ordering are helped greatly
the banking trustco is said to expand its purchase part of its royal international plan operations

royal trustco said the purchase is part of its plan to expand its international banking operations

these n-grams have semantics

for its part, stressed the longstanding relationship with its own, chinese boeing

boeing, for its part, stressed its own longstanding relationship with the chinese

[Soricut & Marcu 05]
Language Models

• $P(\text{string})$
  – Used in every statistical MT and ASR system

• $P(\text{tree})$
Language Models

• P(tree)
  – Parsing researchers have been building P(tree) models for 10 years, trained on the Penn Treebank

  – PARSING: use P(tree) to grade different parses of the same sentence
    • [Collins 97, Charniak 01, …]

  – TRANSLATION: use P(tree) to grade different translations of the same sentence
    • [Charniak, Knight, Yamada 03] -- not yet in the mainstream of MT.

  – GENERATION: use P(tree) to grade different realizations of the same meaning
    • [Bangalore & Rambow 00, …]
Language Models

• $P(\text{meaning})$

• $P(\text{interlingua})$
  – “I saw the Grand Canyon flying to New York”
MT Ziggurat
MT Ziggurat
3rd topic: Software Tools

- What tools do MT researchers use?
  1. Giza++, Moses, … (translation modeling)
  2. SRI-LM (or other) language modeling toolkit
  3. Collins (or other) English parser
  4. BBN Serif (or other) named entity finder
  5. AT&T FSM (or other) finite-state toolkit

- No specific generation tools
Pipelines Considered Harmful

• Ideal MT system doesn’t pipeline
• It brings all knowledge sources to bear during incremental construction of the translation
What hunger have I,
Hungry I am so,
I am so hungry,
Have I that hunger …

Que hambre tengo yo → I am so hungry
Word-Based Statistical MT

Spanish/English Bilingual Text

Statistical Analysis

TRANSLATION MODEL

Language Text

Statistical Analysis

LANGUAGE MODEL

Que hambre tengo yo

I am so hungry
Pipelines Considered Harmful

- Ideal MT system doesn’t pipeline
- It brings all knowledge sources to bear during incremental construction of the translation

<table>
<thead>
<tr>
<th>Approach</th>
<th>LM used</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>No LM</td>
<td>--</td>
<td>x</td>
</tr>
<tr>
<td>TM and LM in pipeline (n-best)</td>
<td>3-gram</td>
<td>y &gt; x</td>
</tr>
<tr>
<td>TM and LM integrated</td>
<td>3-gram</td>
<td>z &gt; y</td>
</tr>
</tbody>
</table>
Pipelines Considered Harmful

• But there is some pipelining in MT...
  1. Pre-processing
     • Source tokenization, normalization
     • Source lower-casing
  2. Decoding
  3. Post-processing
     • Target de-tokenization (suck up punctuation)
     • Target re-capitalization

Actual separable generation tool! Not just a game.
Generation Tools

- **Morphology**
  - When we translate into morpho-complex languages, we need to reliably take words apart and put them back together

- **Normalization**
  - Tools for reducing language variation, to improve training
  - Need to put Humpty Dumpty together again

- **Lexical choice**

- **Linearization**
Generation Tools

• How to “box up” something like linearization?
• Even if a generation tool isn’t bolted onto an MT system, its knowledge can be integrated into incremental translation

• Bag generation
  – What program can assign the best ordering to an unordered bag of words? (Brown et al 90)
  – Several models considered in (Soricut & Marcu 05)
  – Can be set up as a community challenge
  – Breakthroughs will be extremely significant for MT
Challenges & Evaluations

<table>
<thead>
<tr>
<th>ANALYSIS</th>
<th>GENERATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphological analysis</td>
<td>Morphological generation (in many languages)</td>
</tr>
<tr>
<td>Word sense disambiguation</td>
<td>Lexical selection</td>
</tr>
<tr>
<td>Co-reference resolution</td>
<td>Reference generation</td>
</tr>
<tr>
<td>Semantic role labeling</td>
<td>Meaning-to-text</td>
</tr>
<tr>
<td>Parsing</td>
<td>Linearization</td>
</tr>
</tbody>
</table>

“Evaluations drive progress”
“It’s not just a game”
“It has to be kind of fun”
“Somebody from outside has to be interested”
NLG and the MT Triangle

• Y. Wilks: “Natural language understanding is like counting from 1 to infinity. Natural language generation is like counting from infinity down to 1.”

• Infinity is a little worrisome, despite providing some job security…
• Y. Wilks: “Natural language understanding is like counting from 1 to infinity. Natural language generation is like counting from infinity down to 1.”
• Infinity is a little worrisome, despite providing some job security…
• How about let’s approach it this way:
  – 1
  – 2, 1
  – 3, 2, 1
  – 4, 3, 2, 1
  etc
thanks!