Towards developing text-to-text generation systems

Daniel Marcu

Information Sciences Institute and Department of Computer Science

University of Southern California

4676 Admiralty Way, Suite 1001

Marina del Rey, CA 90292

marcu@isi.edu

http://www.isi.edu/~marcu/

What goes on outside NLG?

- The most robust NL modules/components/ systems are those that search through millions of potential solutions, using parameters estimated by inspecting large amounts of training data.
  - Parsing [Lexicalized PCFGs].
  - POS tagging and named-entity recognition [Finite State Transducers].
  - Machine translation [Stochastic translation models].

Keys to Success

- Ability to
  - frame scientific problems so that they can be solved through extensive search over large spaces of solutions.
  - collect large amounts of training material that is compatible with the manner in which a problem is framed.
  - define parameters that explain how training data is generated.
  - estimate parameters from the data.
  - identify metrics that reflect the goodness of the output.

Example: Machine Translation

- Framing of the problem:
  - A noisy-channel translation model can be used to enumerate all the ways in which a sentence can be translated.
- Training:
  - Parallel corpus.
- Evaluation:
  - ?? (In general, very difficult).
  - But when training, it is straightforward: ability to generate the translations in a parallel corpus.

How can we apply a similar approach to building NLG applications?

- Pose generation problems and sub-problems so that they can be solved using search techniques that probe spaces defined by parameters that are automatically learned.
- Worry about
  - the framing of the problems and the definition of parameters;
  - the acquisition of training material that can be used to estimate the parameters;
  - the definition of scoring metrics to assess the goodness of the output.

We need corpora!!!

- We have a significant body of research on NLG from logical-forms.

- In spite of this, I believe that building corpora of <logical-form, text> tuples is not a good idea.
- Rather, I believe it is better to build corpora of <text_source, text_target> tuples.

Examples

- <badly-written texts, well-written texts>
- <long sentences, short sentences>
- <paragraphs, one-sentence summaries of them>
- <complex sentences, simple sentences>
- <sentences, same-meaning sentences>
- <texts with specialized vocabularies and terminology, texts written for laypeople>

NLG from KB vs. NLG from text

	KB-based	Text-based
Input	Heterogeneous Unreasonably rich	Homogeneous Uncontroversial
Corpus creation	Expensive Difficult	Inexpensive Less difficult
Technology transfer	Very difficult	Very easy
Evaluation	Very difficult	Training: easy Test: difficult (but see IBM-Bleu, MT)
System comparison	Very difficult	Easy
Resource and corpora sharing	Very difficult	Easy

Examples (Knight and Marcu, AAAI'00)

- The documentation is typical of Epson quality: excellent.
- Documentation is excellent.

- All of our design goals were achieved and the delivered performance matches the speed of the underlying device.
- All design goals were achieved.

- Reach's E-mail product, MailMan, is a message-management system designed initially for VINES LANs that will eventually be operating system-independent.
- MailMan will eventually be operating system-independent.

- Although the modules themselves may be physically and/or electrically incompatible, the cable-specific jacks on them provide industry-standard connections.
- Cable-specific jacks provide industry-standard connections.

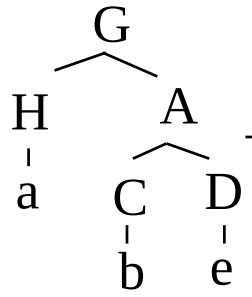
- Ingres/Start prices start at $2,000
- Ingres/Start prices start at $2,000.

Statistical sentence compression (syntax-based model)

Lots of parse trees of short sentences

P(s)

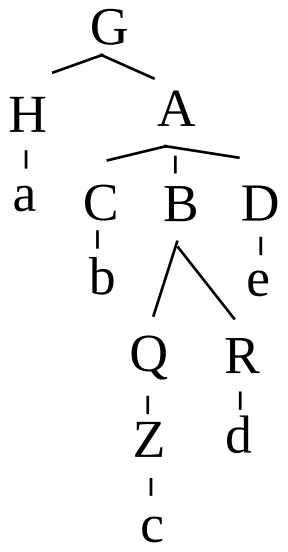
editor



Lots of tuples of parse trees of short and long sentences

P(l | s)

reporter



Source model:

P(s) → PCFG: $P(G \rightarrow H\ A \mid G) \times$
$P(H \rightarrow a \mid H) \times \ldots$

Bigrams: $P(a \mid BOS) \times$
$P(b \mid a) \times \ldots$

Channel model:

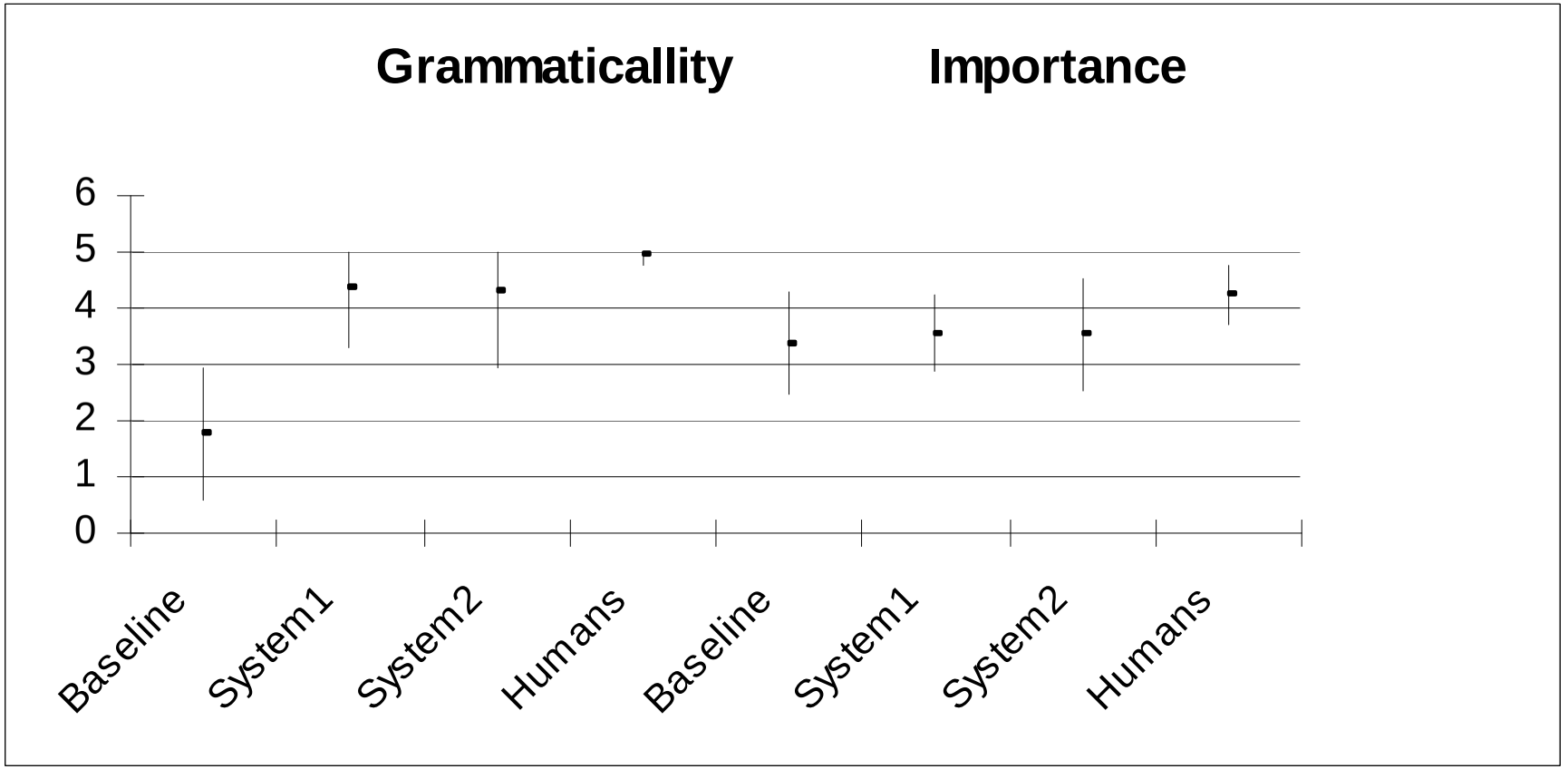
P(l | s) → Insertion: $P(A \rightarrow C\ B\ D \mid A \rightarrow C\ D)$

Expansion: $P(B \rightarrow Q\ R \mid B) \times$
$P(Q \rightarrow Z \mid Q) \times \ldots$

Compression Examples

```
Beyond that basic level, the operations of the three products vary widely(1514588)
Beyond that level, the operations of the three products vary widely (1430374)
Beyond that basic level, the operations of the three products vary (1333437)
Beyond that level, the operations of the three products vary (1249223)
Beyond that basic level, the operations of the products vary (1181377)
The operations of the three products vary widely (939912)
The operations of the products vary widely (872066)
The operations of the products vary (748761)
The operations of products vary (690915)
Operations of products vary (809158)
The operations vary (522402)
Operations vary (662642)
```

Performance

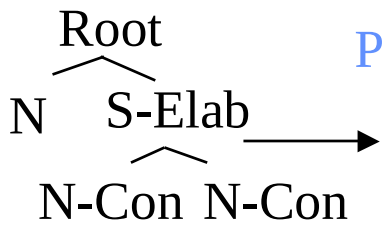


Hal Daume: Scaling up sentence compression to text compression (discourse and syntax-based model)

Lots of discourse trees of short texts (extracts)

P(extract)

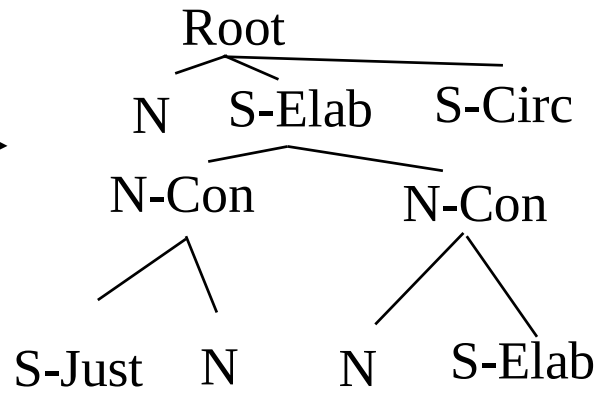
editor



Lots of tuples of discourse trees of texts and their extracts

P(text | extract)

reporter



Source model:

P(extract) Discourse PCFG:

$P(\text{Root} \rightarrow \text{N S-elab} \mid \text{Root}) \times$
$P(\text{S-elab} \rightarrow \text{N-Con N-Con} \mid \text{S-Elab})$

Channel model:

P(text | extract)

Insertion: $P(\text{R} \rightarrow \text{N S-Elab S-Circ} \mid \text{R} \rightarrow \text{N S-Elab})$

Expansion: $P(\text{N-Con} \rightarrow \text{S-Just N} \mid \text{N-Con})$

Examples:

Telxon Corp. said its vice president for manufacturing resigned and its Houston work force has been trimmed by 40 people, or about 15%.
The maker of hand-held computers and computer systems said the personnel changes were needed to improve the efficiency of its manufacturing operation. The company said it hasn't named a successor to Ronald Bufton, the vice president who resigned.
Its Houston work force now totals 230.

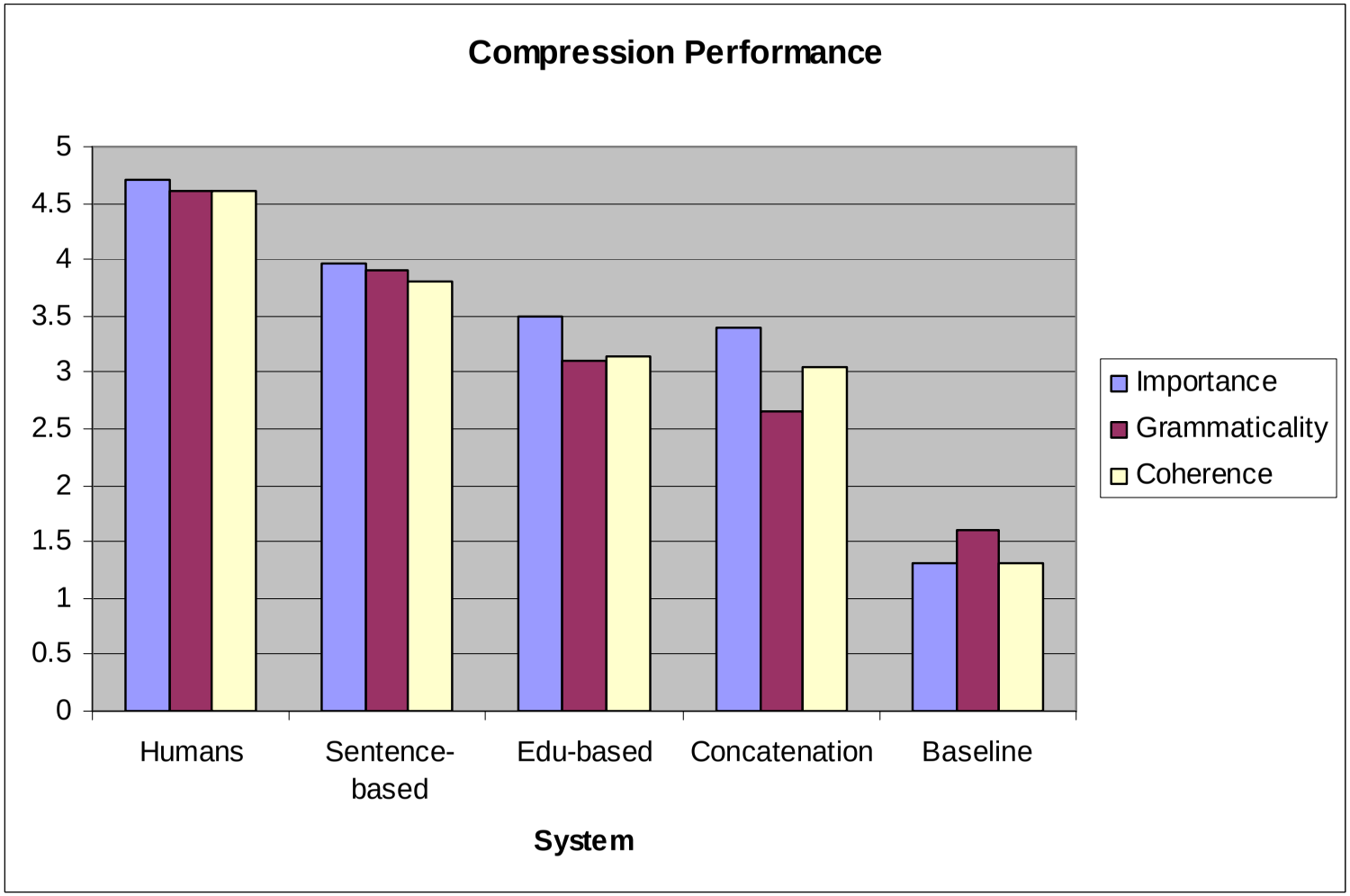
Summary 1:

Telxon Corp. said The company said it has not named a successor to Ronald Bufton , the vice president Its work force now totals 230 .

Summary 2:

Telxon Corp. said Its Houston work force now totals 230 .

Evaluation


Compression Performance
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Humans
Sentence-based
Edu-based
Concatenation
Baseline
System
Importance
Grammaticality
Coherence

More sophisticated models are needed (examples of two-to-one sentence compressions)

He holds a bachelor's degree in chemistry. <1>
"Maintaining an organization like ISFUG is like building a castle in the sand; it just requires constant work to keep it in trim," said Berkman, an economist with the Commerce Department's Bureau of Economic Analysis.

Berkman, who holds a bachelor's degree in chemistry, is an economist with the US Department of Commerce.

Nonetheless, policies on its use vary. <1>
Agencies such as the Internal Revenue Service and the Farm Credit Administration promote shareware use, but one NASA center shuns it.

Federal agencies generally use shareware, but policies on its use vary.

Summary

- Build text-to-text corpora and focus on text-to-text generation.
  - During training – evaluation is straightforward.
  - For testing – use IBM-Bleu evaluation schema.

- Note: this does not solve all our problems, but it can help us build systems that are more robust than those we can build now.

Thank you!