Introduction to Corpus-based Computational Semantics: *Paraphrases*

Alexander Koller
Columbia University

Sabine Schulte im Walde
University of Stuttgart

Introductory Course
ESSLLI 2007
Trinity College, Dublin, Ireland
August 6-10, 2007

Now what?

- Semantic relations between words ← yesterday
  - » Can we learn Wordnet relations automatically?
  - » Word relations vs. sense relations
- Similarity between phrases ← today
  - » Paraphrase: Two semantically similar word sequences
  - » Extended distributional hypothesis

What are paraphrases?

- A paraphrase is a pair of two expressions which can be exchanged without changing the meaning of a sentence (at least in some contexts).
- Approximation of phrase-level synonymy.
- In the literature, varied definitions:
  - » "approximate conceptual equivalence" (Barzilay & McKeown 01)
  - » "a sentence containing p' might contain an answer to the question from which p was extracted" (Lin & Pantel 01)

What are paraphrases?

- Example:
  - » 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.

Why do we care about paraphrases?

- Information retrieval: Find not just documents that match your query, but also documents that match paraphrases.
- Information extraction: Identify information not just by the pattern you defined, but also by its paraphrases.
- Question answering: Enable your system to recognise a paraphrase as the answer.
- Semantic world knowledge?

Application to Textual Entailment

- Remember last night's RTE example?
  - » Prime Minister Mahmoud Abbas offered the hand of peace to Israel after his landslide victory in Sunday's presidential election.
  - » Mahmoud Abbas has claimed victory in the presidential elections.
- Problem for knowledge-based system because it didn't realize that "X's landslide victory" implies "X claimed victory".
- Paraphrase?

Types of paraphrases

- Lexical paraphrase: both expressions are single words.
  - » China has regarded Taiwan as a rebel province since a civil war split them in 1949.
  - » China has regarded Taiwan as a rebel province since a civil war separated them in 1949.
- Lexical paraphrases are typically synonyms or hyponyms.

Types of paraphrases

- Structural paraphrases: expressions are more complex phrases.
  - » Detroit building reduced to rubble.
  - » Detroit building blasted to the ground.
- Sometimes there is a "grey area":
  - » After the latest Fed rate cut, stocks rose across the board.
  - » Winners strongly outpaced losers after Greenspan cut interest rates again.

Paraphrase representations

- As pairs of strings: (comfort, console)
- As templates with argument slots:
  (X comforts Y, X consoles Y)
  (X buys Y from Z, Z sells Y to X)
  (X wrote Y, X is the author of Y)

Paraphrases in this lecture

- We will discuss three main approaches:
  - » learning from unannotated corpora using the distributional hypothesis (Lin & Pantel 2001)
  - » learning from multiparallel corpora (Barzilay & McKeown 01)
  - » learning from parallel corpora using methods from machine translation (Bannard and Callison-Burch 2005)
- There are many more variants in the literature.

The Distributional Hypothesis

Grandma baked an apple pie for my birthday.

- *„You shall know a word by the company it keeps.“*
  Firth (1957), *A Synopsis of Linguistic Theory*

- *„Each language can be described in terms of a distributional structure, i.e., in terms of the occurrence of parts relative to other parts. The distribution of an element is understood as the sum of all its environments”.*
  Harris (1968), *Distributional Structure*

Lin's implementation of the DH

Dekang Lin (1998): „Automatic retrieval and clustering of similar words". In: *Proceedings of the 17th International Conference on Computational Linguistics.*

- Model word similarity by functional context similarity
- Basis: distributional patterns of word *w* and relations *r:* *f(w,r,w')* for all patterns *(w,*,*),* using dependency triples

  f(cell, subj-of, adapt) = 1
  f(cell, obj-of, attack) = 6 ...

- Mutual information *I(w,r,w'):*

$$I(w,r,w') = \log_2 \frac{p(w,r,w')}{p(*,r,*) \times p(w,r,*) \times p(*,r,w')} = \log_2 \frac{f(w,r,w') \times f(*,r,*)}{f(w,r,*) \times f(*,r,w')}$$

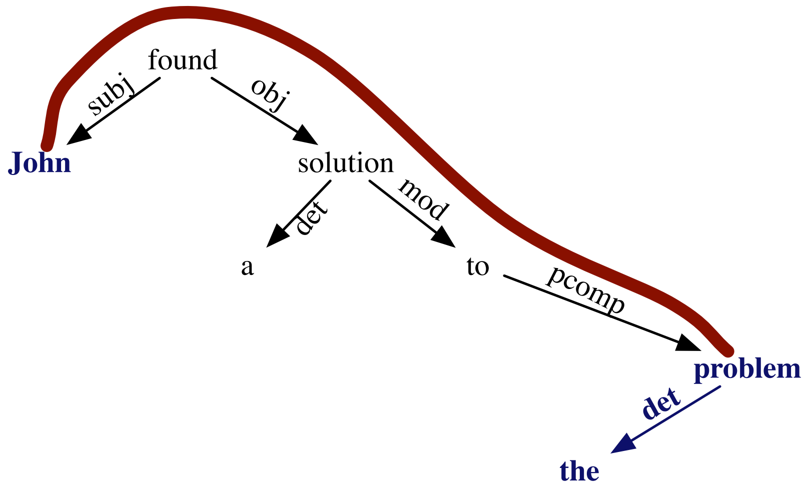
The Distributional Hypothesis

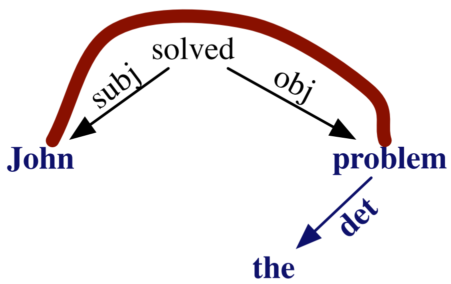
- Words that appear frequently in similar contexts have similar meanings.
- Extended Distributional Hypothesis: Contexts that frequently connect similar words have similar meanings. → paraphrases

"*X finds a solution to Y*"		"*X solves Y*"	
SlotX	SlotY	SlotX	SlotY
commission	strike	committee	problem
committee	civil war	clout	crisis
committee	crisis	government	problem
government	crisis	he	mystery
government	problem	she	problem
he	problem	petition	woe
I	situation	researcher	mystery
legislator	budget deficit	resistance	crime
sheriff	dispute	sheriff	murder

Lin & Pantel 2001: The DIRT system

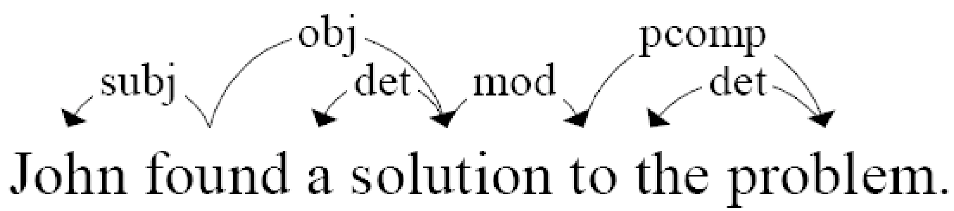
- Generalisation of Lin's noun similiarity system to multi-word template paraphrases.
- Here paraphrases are paths in syntactic dependency trees that can be exchanged:





Parsing

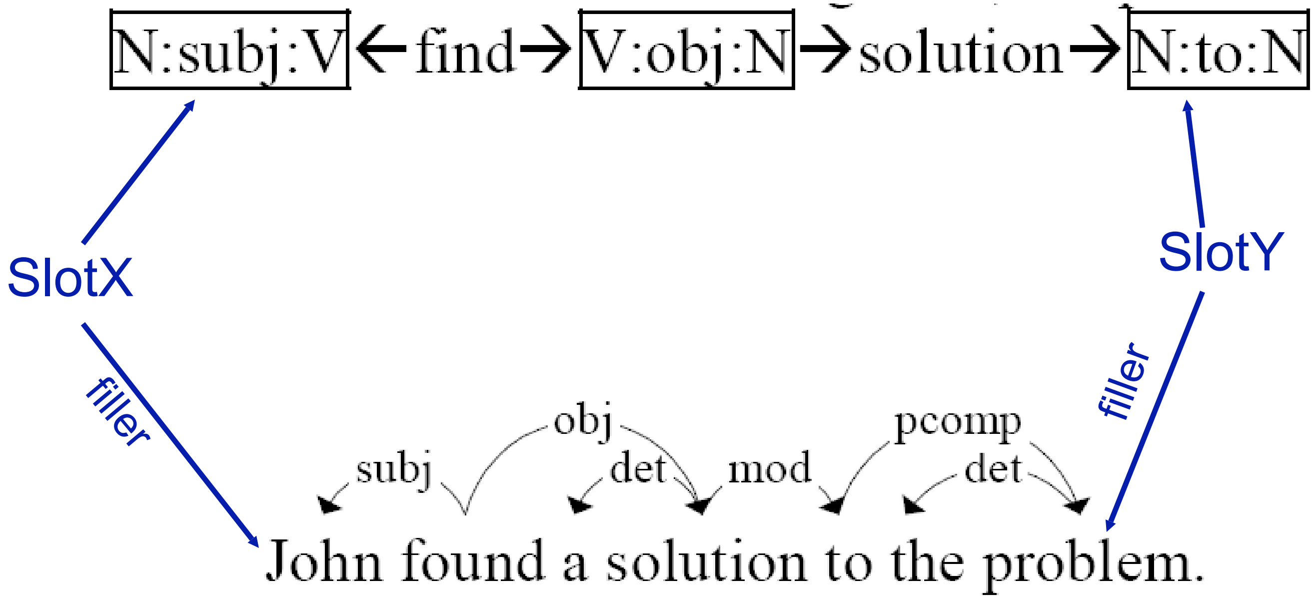
- We start with an unannotated corpus (1 GB of English newspaper text).
- Parsing using broad-coverage Minipar parser gives us a dependency structure for each sentence:



- Parse tree is (automatically) simplified a bit.

Paths

- An (undirected) path in a dependency tree is specified by the words and relations along the path.
- Example:



Triples

- For each path p in the corpus, we collect the words that can fill its two slots in triples:

  (p, SlotX, w_1)

  (p, SlotY, w_2)

- The more common triples two paths have, the more similar we take them to be.

Mutual Information

- We need MI for three events:

$$mi(x,y,z) = \log \frac{P(x,y,z)}{P(x)P(y)P(z)}$$

- Apply this to our problem, assuming conditional independence of path and filler given a slot:

$$mi(p,Slot,w) = log \frac{P(p,Slot,w)}{P(Slot)P(p \mid Slot)P(w \mid Slot)}$$

- Estimate these probabilities by counting in corpus.

Similarity between two paths

- Define the similarity of two pairs $slot_1 = (p_1,s)$ and $slot_2 = (p_2,s)$ of paths and a slot:

$$sim(slot_1, slot_2) = \frac{\sum_{w \in T(p_1,s) \cap T(p_2,s)} mi(p_1,s,w) + mi(p_2,s,w)}{\sum_{w \in T(p_1,s)} mi(p_1,s,w) + \sum_{w \in T(p_2,s)} mi(p_2,s,w)}$$

- $T(p,s)$ are fillers with positive mutual information.
- Define the similarity of two paths as the geometric average of the similarities of their slots:

$$S(p_1, p_2) = \sqrt{sim(SlotX_1, SlotX_2) \times sim(SlotY_1, SlotY_2)}$$

Finding similar paths in the corpus

- Extract all paths that satisfy certain constraints. In the newspaper corpus, this gives 200.000 distinct paths.
- Compute similarity for each pair of paths.
- Because this is a computational nightmare, apply some clever heuristics to make it feasible.

Top paraphrases for "X solves Y"

1.	Y is solved by X	26.	X clears up Y
2.	X resolves Y	27.	*X creates Y
3.	X finds a solution to Y	28.	*Y leads to X
4.	X tries to solve Y	29.	Y is eased between X
5.	X deals with Y	30.	X gets down to Y
6.	Y is resolved by X	31.	X worsens Y
7.	X addresses Y	32.	X ends Y
8.	X seeks a solution to Y	33.	*X blames something for Y
9.	X do something about Y	34.	X bridges Y
10.	X solution to Y	35.	X averts Y
11.	Y is resolved in X	36.	*X talks about Y
12.	Y is solved through X	37.	X grapples with Y
13.	X rectifies Y	38.	*X leads to Y
14.	X copes with Y	39.	X avoids Y
15.	X overcomes Y	40.	X solves Y problem
16.	X eases Y	41.	X combats Y
17.	X tackles Y	42.	X handles Y
18.	X alleviates Y	43.	X faces Y
19.	X corrects Y	44.	X eliminates Y
20.	X is a solution to Y	45.	Y is settled by X
21.	X makes Y worse	46.	*X thinks about Y
22.	X irons out Y	47.	X comes up with a solution to Y
23.	*Y is blamed for X	48.	X offers a solution to Y
24.	X wrestles with Y	49.	X helps somebody solve Y
25.	X comes to grip with Y	50.	*Y is put behind X

Evaluation

- Basis for evaluation: Questions from TREC-8 question answering competition.
- Compute all paraphrases that DIRT finds for each question.
- Use human judge to check how many of these are correct.
- Evaluate overlap of DIRT paraphrases and human-produced manual paraphrases.

Evaluation: Results

QUESTION	PATHS	MANUAL	DIRT (CORRECT)	INTERSECTION	ACCURACY
Q_1	X is author of Y	7	21	2	52.5%
Q_2	X is monetary value of Y	6	0	0	0%
Q_3	X manufactures Y	13	37	4	92.5%
Q_4	X spend Y	7	16	2	40.0%
	spend X on Y	8	15	3	37.5%
Q_5	X is managing director of Y	5	14	1	35.0%
Q_6	X asks Y	2	23	0	57.5%
	asks X for Y	2	14	0	35.0%
	X asks for Y	3	21	3	52.5%
Q_7	X leave Y	4	0	0	0%
Q_8	X is disease with Y	5	0	0	0%
Q_9	∅	N/A	N/A	N/A	N/A
Q_{10}	X is designer of Y	5	7	2	17.5%
Q_{11}	∅	N/A	N/A	N/A	N/A
Q_{12}	∅	N/A	N/A	N/A	N/A
Q_{13}	rent X for Y	14	16	1	40.0%
Q_{14}	X is producer of Y	10	31	3	77.5%
Q_{15}	∅	N/A	N/A	N/A	N/A

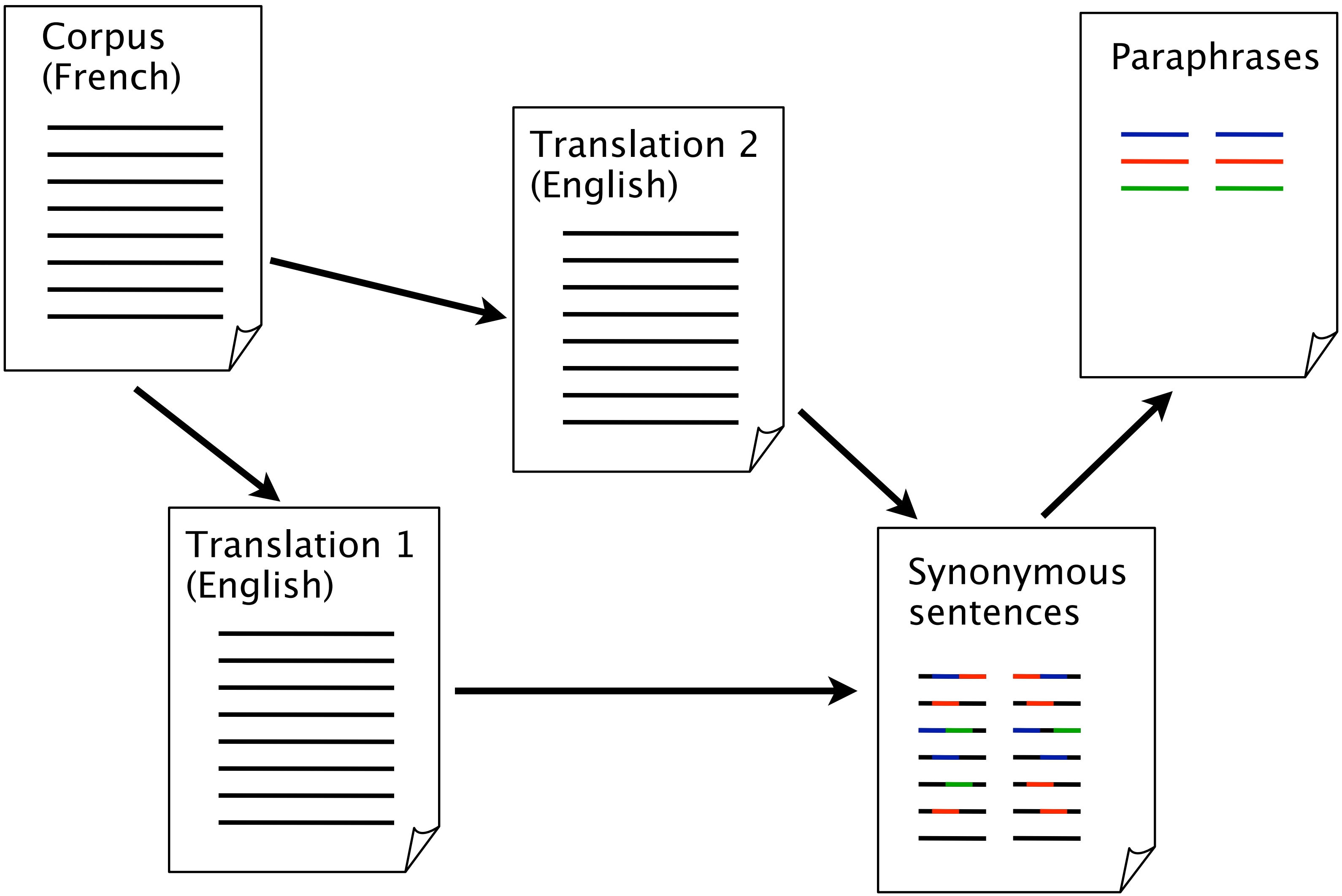
Lin & Pantel: Conclusions

- Run a broad-coverage dependency parser.
- Extract paths and compute pairwise similarity.
- Advantage: Get syntactic paraphrase templates.
- Problems:
  - » Comparing each pair of paths is computationally problematic
  - » Is the Extended Distributional Hypothesis correct in general?

Fixing the EDH

- The EDH may not be correct in general.
- But it would probably be more true if the two contexts are in synonymous sentences.
- Now, if only we had access to pairs of synonymous sentences ...

Barzilay & McKeown: The big picture



Parallel translations

- Idea: Different translations of the same foreign-language sentence are synonymous.
  - » use multiple English translations of five French novels
- Compute a sentence-level alignment of each document (methods from statistical MT).
  - » sentences aligned to same F sentence are synonymous
- Simultaneously learn ("co-training")
  - » pairs of synonymous pharases (i.e. paraphrases) EDH
  - » pairs of contexts that surround synonymous phrases DH

Learning paraphrases and contexts

Emma burst into tears, and he tried to comfort her, ...

Emma cried, and he tried to console her, ...

Peter managed to solve it, ...

Peter managed to fix it, ...

left context:

tried	to	her	,
VB_1	TO_2	$PRP\$_3$	,

right context:

tried	to	her	,
VB_1	TO_2	$PRP\$_3$	,

Algorithm

1. Initialization: Equal words in parallel sentences are paraphrases, all other word pairs aren't.
2. Update contexts:
   - context strength: strength+(x) = count(x+)/count(x)
   - pick 10 strongest contexts, only with strength > 95%
3. Update paraphrases:
   - find matches of contexts in corpus and take fillers as new paraphrases
   - paraphrases may be multi-word
4. Go back to 2 with new paraphrases.

Examples: Paraphrases

lexical paraphrases

(countless, lots of)	(repulsion, aversion)
(undertone, low voice)	(shrubs, bushes)
(refuse, say no)	(dull tone, gloom)
(sudden appearance, apparition)	

morpho-syntactic paraphrase patterns

$(NN_0\ POS\ NN_1)$	$\leftrightarrow(NN_1\ IN\ DT\ NN_0)$
King's son	son of the king
$(IN\ NN^0)$	$\leftrightarrow(VB^0)$
in bottles	bottled
$(VB_0\ to\ VB^1)$	$\leftrightarrow(VB_0\ VB^1)$
start to talk	start talking
$(VB_0\ RB_1)$	$\leftrightarrow(RB_1\ VB_0)$
suddenly came	came suddenly
$(VB\ NN^0)$	$\leftrightarrow(VB^0)$
make appearance	appear

Evaluation

- Precision:
  - » two human judges, 500 random paraphrases
  - » without context: precision = 87.8% / 85.2% (K=0.68)
  - » with original context: precision = 91.8% / 91.4% (K=0.97)
- Recall:
  - » hard to evaluate: no resources to compare against
  - » One judge extracted 70 paraphrases from 50 sentences by hand. Of these, the system found 69%.

Comparison to Wordnet

- What Wordnet relation do single-word paraphrases stand in?
  - » check this for 112 word pairs that occurred at least 20 times
  - » Synonymy: 35% (rise/stand up, hot/warm)
  - » Hyponymy: 32% (landlady/hostess, reply/say)
  - » Sisters: 18% (city/town, pine/fir)
  - » Unrelated: 10% (sick/tired, next/then)

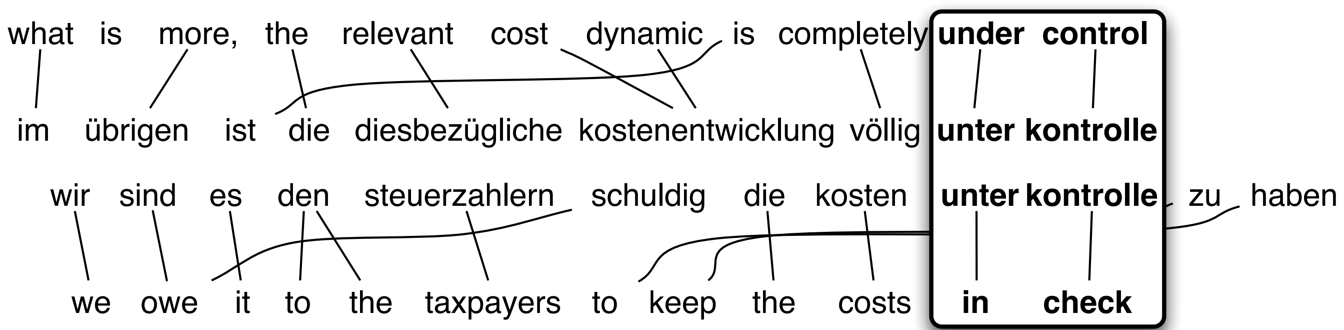
Barzilay & McKeown: Summary

- Obtain synonymous sentences from parallel translations.
- Use co-training to learn good paraphrases and paraphrase contexts in tandem.
- Problems:
  - » multi-parallel translations are rare
  - » and unannotated
  - » statistical parsers trained on different text genres

Paraphrases from multilingual corpora

(Bannard & Callison-Burch 05)

- Idea: Use methods from statistical MT to align phrases in multilingual corpora.
- Alternative translations of same phrase may be paraphrases.



(Europarl corpus: Proceedings of European Parliament)

Statistical MT

- Goal: Given an input sentence in one language, find best translation in another language.
- Use parallel corpus to train a statistical model P(e|f): Probability of English sentence e as translation of foreign language sentence f.
- Then, given input sentence f, compute English sentence e such that P(e|f) is maximized.
- Phrase-based and word-based models; both based on (manual or automatic) word alignments.
- This is an extremely active research field right now.

Statistical MT for paraphrasing

- Define paraphrase probability $P(e_2|e_1)$ for e_2 being a paraphrase of e_1.
- Using parallel corpus, express $P(e_2|e_1)$ in terms of translation probabilities:

$$p(e_2|e_1) = \sum_f p(f|e_1)p(e_2|f)$$

- Best paraphrase for e_1: maximizes $P(e_2|e_1)$.

Some example paraphrases

under control	checked, curb, curbed, *in check*, limit, slow down
sooner or later	*at some point*, eventually
military force	armed forces, defence, *force*, forces, military forces, peace-keeping personnel
long ago	a little time ago, a long time, *a long time ago*, a lot of time, a while ago, a while back, far, for a long time, for some time, for such a long time, long, long period of time, long term, long time, long while, overdue, some time, some time ago
green light	approval, call, *go-ahead*, indication, message, sign, signal, signals, formal go-ahead
great care	a careful approach, greater emphasis, *particular attention*, special attention, specific attention, very careful
first half	*first six months*
crystal clear	absolutely clear, all clarity, clear, clearly, in great detail, no mistake, no uncertain, obvious, obviously, particularly clear, perfectly clear, quite clear, quite clearly, quite explicitly, quite openly, very clear, *very clear and comprehensive*, very clearly, very sure, very unclear, very well
carbon dioxide	*co2*
at work	at the workplace, employment, held, holding, in the work sphere, operate, organised, taken place, took place, *working*

Italicized phrases: Considered best paraphrases by the model.

Evaluation: Method

- Manual evaluation of highest-ranked paraphrases:
  - » Insert highest-ranked paraphrase into original context:
    This situation is in check in terms of security.
    This situation is checked in terms of security.
    This situation is slow down in terms of security.
  - » Ask human to judge whether paraphrased sentence is both grammatical and preserves original meaning.

Evaluation: Results

- With manual word alignment: 85% correct meaning, 75% also grammatically correct.
- With automatic word alignment: 65% correct meaning, about 50% also grammatically correct.
- Improve performance with automatic alignment to 70% / 62% by
  - » using a language model
  - » using multiple foreign languages
  - » controlling for word senses

Conclusion

- Paraphrases capture meaning equivalence of word sequences.
- Various methods:
  - » single corpus, distributional hypothesis
  - » multiple translations, co-training with distributional hypothesis
  - » parallel corpus, methods from machine translation
- Best methods get results that seem pretty good.
- Evaluations hard to compare.