Semantics & Machine Translation

Marine Carpuat
University of Maryland
marine@cs.umd.edu

sem•an•tics (sĭ-măn'tĭks)

n. *Linguistics* The study or science of meaning in language.

n. *Linguistics* The study of relationships between signs and symbols and what they represent.

n. The meaning or the interpretation of a word, sentence, or other language form: *We're basically agreed; let's not quibble over semantics.*

from The American Heritage® Dictionary of the English Language, 4th Edition

وسيتولى قيادة الفرقة الموسيقية الأستاذ أحمد عاشور قائد الأوركسترا

...

Reference:

mr. ahmad ashour , conductor of the tunisian symphony orchestra , will lead the orchestra , which will be infused with a number of levantine instruments […]

MT:

mr. ahmad ashour , commander of the tunisian symphony orchestra and will be vaccinated by a group of eastern instruments

[…]

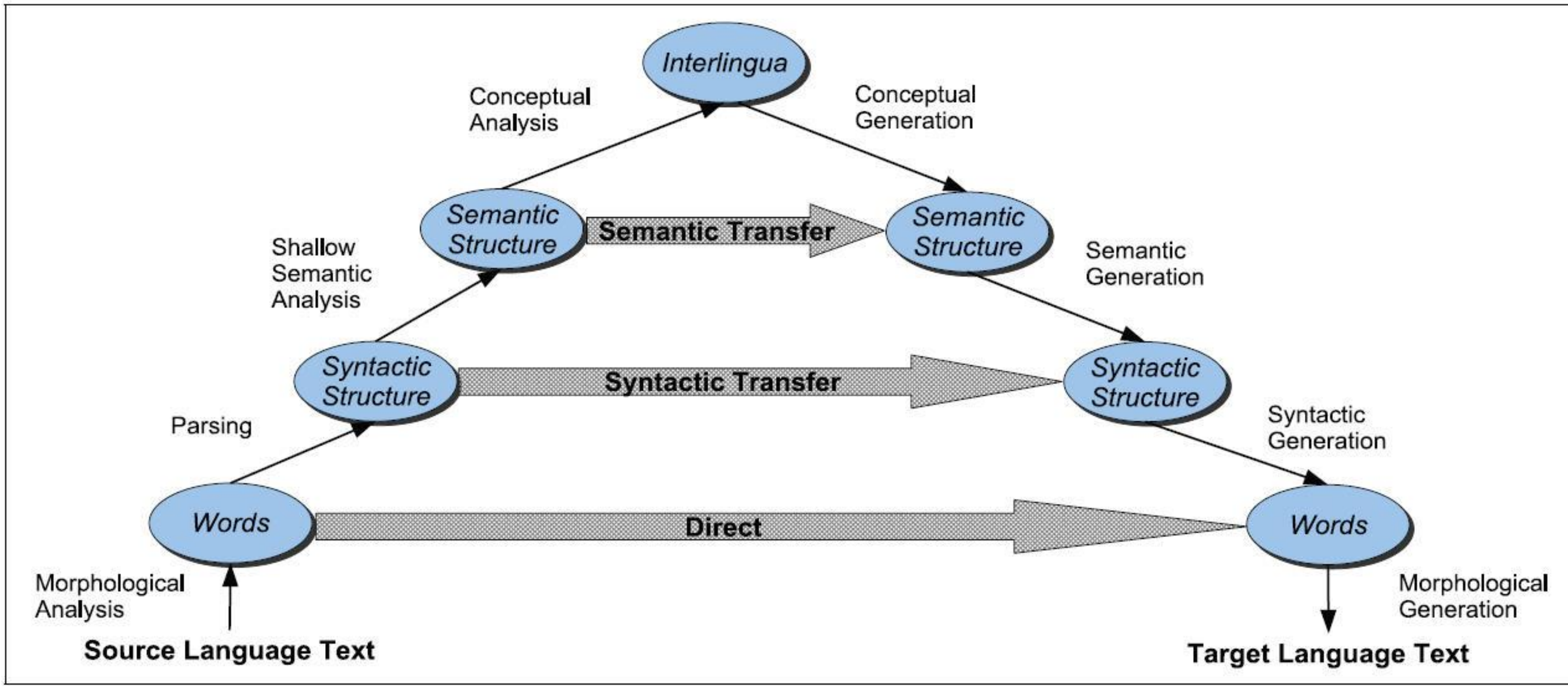
Semantics & Machine Translation

- Machine Translation goal: produce fluent translations that **adequately capture meaning of the source**.

- In NLP, this is the focus of semantic analysis, a long-standing research area

- Yet, most MT systems do not have dedicated semantic analysis components

The Vauquois Triangle



Semantics & Machine Translation

This lecture

- Key problems/tasks in semantic analysis of text and how they relate to MT
- How can semantic analysis be used in MT?
- How can MT (data and models) help improve semantic analysis?

Semantic Analysis Challenges for NLP/MT

`bank: wall street or river?`

Ambiguity

- words can have more than one meaning/sense
- NLP task: word sense disambiguation

Semantic Analysis Challenges for NLP/MT

`shore - coast`

Variability

- distinct word forms might have the same meaning
- NLP task: word semantic similarity, paraphrase detection

Semantic Analysis
Challenges for NLP/MT

```
John broke the window
The window was broken by John
John destroyed the window
```

Sentence-level variability

- many different sentences can express the same meaning
- NLP task: paraphrase detection, semantic textual similarity, semantic composition

Semantic Analysis
Challenges for NLP/MT

```
He misses Paris
Paris lui manque
```

Structural divergence across languages

- same meaning different structure across languages
- NLP task:
  - semantic parsing
  - “who did what to whom”

Predicate argument structure
"who did what to whom"

(1) John broke the window

(2) The window broke

(3) The window was broken by John

(4) John busted the window

(5) The window was destroyed by John

(6) John tore down the window

Semantic Roles:

AGENT – an initiator/doer in the event [Who?]

PATIENT - an affected entity [to Whom / to What?]

Example from Ivan Titov

Semantics & Machine Translation

Case studies: MT with models of

- **word senses**
- word similarity
- predicate argument structure
- paraphrase

Word Sense Disambiguation (WSD)

- Goal: identify correct meaning of a word ("sense") in context:

- Long standing area of research in NLP
  - Perhaps first formulation by Warren Weaver
  - Used in early MT [Brown et. al 1991, Berger et al. 1996]

Warren Weaver's intuition, 1955

If one examines the words in a book, one at a time as through an opaque mask with a hole in it one word wide, then it is obviously impossible to determine, one at a time, the meaning of the words…

But if one lengthens the slit in the opaque mask, until one can see not only the central word in question but also say N words on either side, then if N is large enough one can unambiguously decide the meaning of the central word…

The practical question is : ``What minimum value of N will, at least in a tolerable fraction of cases, lead to the correct choice of meaning for the central word?"

From Word Sense Disambiguation…

- Predicts **P(sense|s,context)**
- Meanings of s represented by hand defined **sense inventory**
  - Small, exhaustive
  - Mutually exclusive senses
  - Learning requires annotation by hand

```
Senses of "ring (n)" in WordNet

1.ring (a characteristic sound)
2.ring, halo, annulus, doughnut (a toroidal shape)
3.gang, pack, ring, mob (an association of criminals)
      …
8. ring, band (jewelry consisting of a circlet of precious metal worn on the finger)
```

From Word Sense Disambiguation…

- Motivation for WSD in MT: disambiguating sense of word should help eliminate incorrect translations

- But using senses to constrain translation lexicon hurts MT quality
  - On Chinese-English translation
  - Using HowNet sense representations
- Why?
  - Representation mismatch
  - Training data mismatch

[Carpuat & Wu, 2005]

… to Context-Dependent Translation Lexicons

- Learn **P(t|s,context)**
- Represents meanings of s using **translations**
  - Very large, very sparse
  - Only partially observed
  - More than 1 correct translation possible
  - Learn from automatically constructed annotation

```
The twelve stars of the
European flag are depicted
on the outer ring.
  es: anillo

The terrors which Mr Cash
expresses about our future
in the community have a
familiar ring about them.
   es: sonar
```

… to Context-Dependent Translation Lexicons

Context-dependent translation lexicons

- MT-driven sense representation
- Modest improvements in translation quality

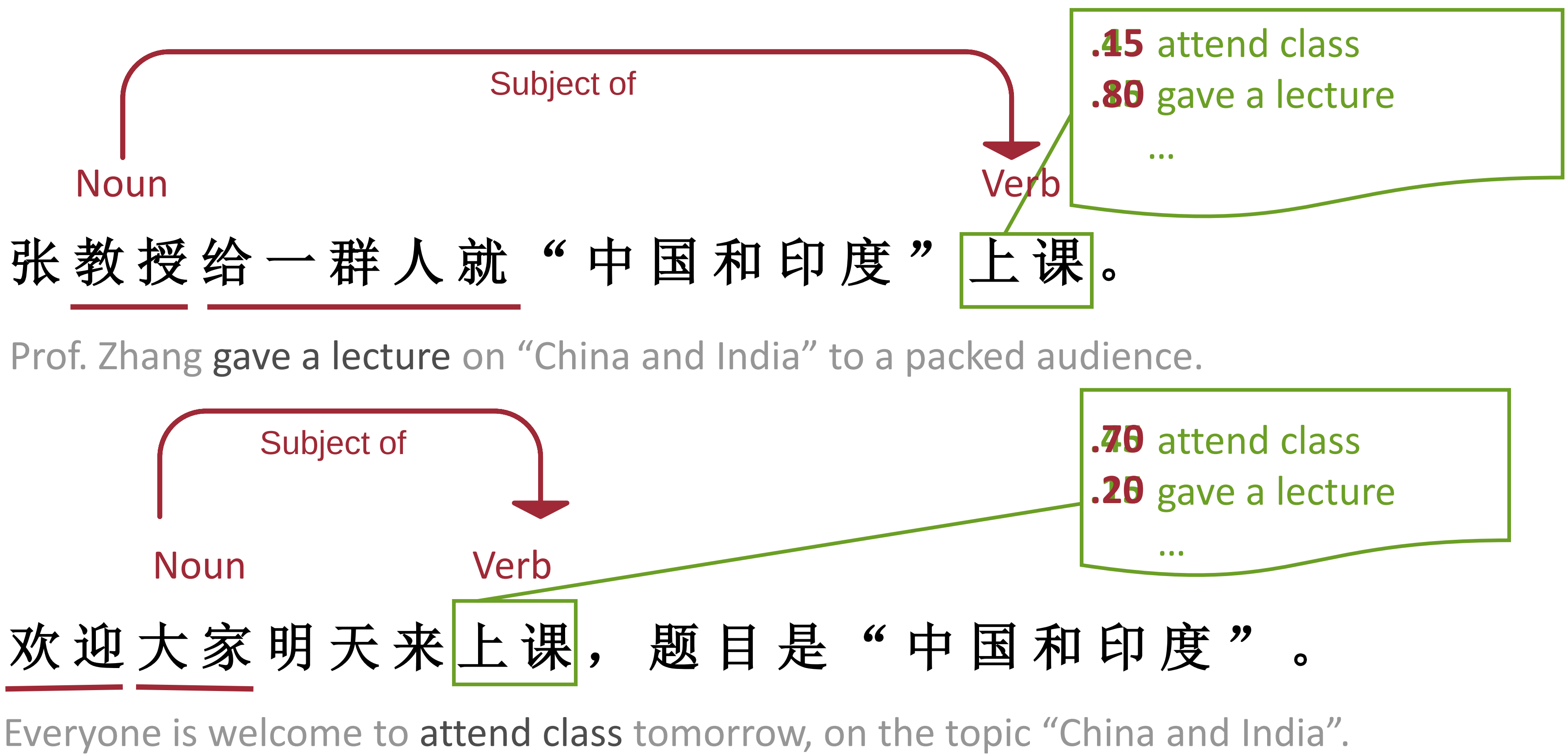
[Carpuat & Wu, 2007]

See also:

[Vickrey et al. 2005, Cabezas & Resnik 2005, Chan et al. 2007, Gimenez & Marquez 2007, Stroppa et al. 2008]

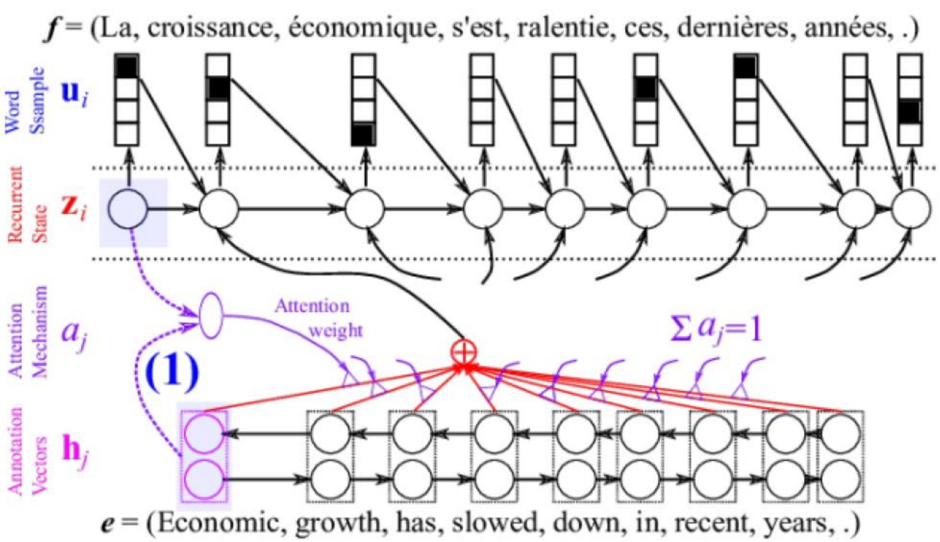
[Apidianaki 2009, Xiong and Zhang 2014]

Context-dependent translation lexicons in PBMT



Word Senses & Neural MT

(Devlin et al. 2014)



(Bahdanau et al. 2015)

- Neural MT already conditions on broader source context
- But does this help address word sense translation errors?

Semantics & Machine Translation

Case studies: MT with models of

- word senses
- **word similarity**
- predicate argument structure
- paraphrase

Modeling Word Semantic Similarity

- Task: score how similar are “coast” and “shore”?

- **Approach: Distributional hypothesis**
  - Intuition: if 2 words appear in the same context, then they must be similar
  - “You shall know a word by the company it keeps!” [Firth, 1957]
  - “Differences of meaning correlates with differences of distribution” [Harris, 1970]

Representing words as features vectors

- **Word co-occurrence within a window**

	arts	boil	data	function	large	sugar	summarized	water
apricot	0	1	0	0	1	1	0	1
pineapple	0	1	0	0	1	1	0	1
digital	0	0	1	1	1	0	1	0
information	0	0	1	1	1	0	1	0

$$Similarity(apricot, pineapple) = \cos(\overrightarrow{apricot}, \overrightarrow{pineapple})$$

				...			...				...					...		*nmod*, body	*nmod*, bone marrow
cell	1	1	1		16	30		3	8	1		6	11	3	2		3	2	2

How can we use word semantic similarity in MT?

Improve coverage

- by replacing source OOVs or creating new rules with known paraphrases
- based on monolingual distributional information

[e.g., Marton et al. 2009]

`Src: the` **`conductor`** `of the Tunisian orchestra`

`If “conductor” is OOV,`

`replace it with similar term`

`New src: the` **`music director`** `of the Tunisian orchestra`

Word semantic similarity for MT evaluation

```
MT1  the music director of the Tunisian orchestra
MT2  the commander of the Tunisian orchestra
REF  the conductor of the Tunisian orchestra
```

Approach

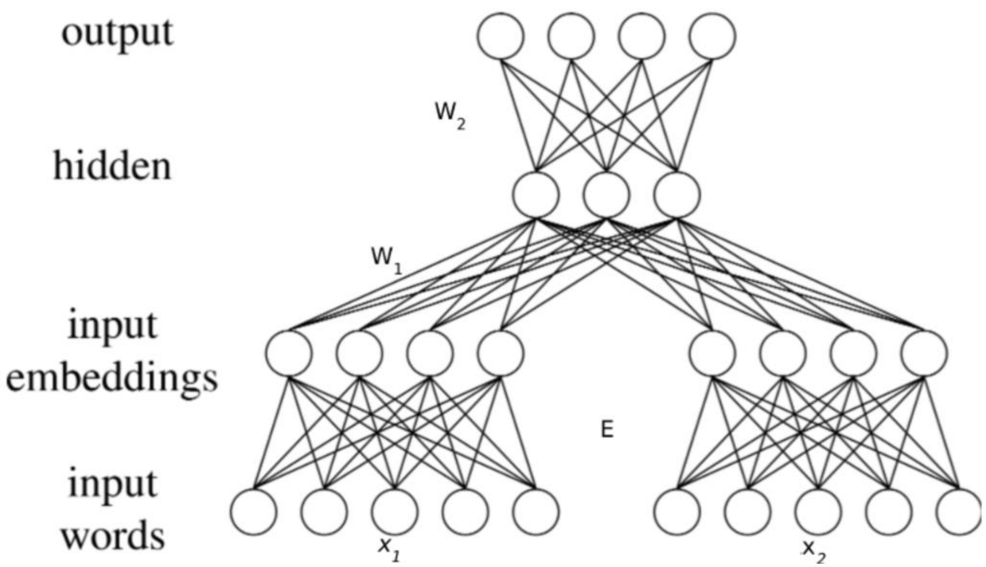
- Replace exact match with semantic similarity score
- Or allow matches with words that share a sense (e.g., WordNet synset)

  [e.g., Liu et al. 2010, Dahlmeier et al. 2011, Denkowski and Lavie 2011]

Neural language models induce continuous word representations

Word representations learned in monolingual neural LM (and variants) improve NLP tasks

[Turian et al. 2010, and many many others since]



	1	2		d
cat	2.059	-1.134	...	2.004
dog	2.011	-1.005	...	0.135
...	...	...	...	...
January	-3.193	0.145	...	0.001
February	-3.016	0.196	...	0.025
...	...	...	...	...

Neural Translation Models Capture Monolingual Semantic Similarity

		Monolingual models			Biling. models		NMT models	
		Skipgram	**Glove**	**CW**	**FD**	**BiCVM**	**RNNenc**	**RNNsearch**
WordSim-353	ρ	0.52	0.55	0.51	**0.69**	0.50	0.57	0.58
MEN	ρ	0.44	0.71	0.60	**0.78**	0.45	0.63	0.62
SimLex-999	ρ	0.29	0.32	0.28	0.39	0.36	**0.52**	0.49
SimLex-333	ρ	018	0.18	0.07	0.24	0.34	**0.49**	0.45
TOEFL	%	0.75	0.78	0.64	0.84	0.87	**0.93**	**0.93**
Syn/antonym	%	0.69	0.72	0.75	0.76	0.70	**0.79**	0.74

English word representations learned by Neural MT models correlate better than monolingual representations with human judgements of

- similarity (e.g., coast-shore)
- relatedness (e.g., clothes-closet)

[Hill et al. ICLR 2015]

Neural Translation Models Capture Monolingual Semantic Similarity

	Skipgram	**Glove**	**CW**	**FD**	**BiCVM**	**RNNenc**	**RNNsearch**
teacher	*vocational*	*student*	*student*	*elementary*	*faculty*	*professor*	*instructor*
	in-service	*pupil*	*tutor*	*school*	*professors*	*instructor*	*professor*
	college	*university*	*mentor*	*classroom*	*teach*	*trainer*	*educator*
eaten	*spoiled*	*cooked*	*baked*	*ate*	*eating*	*ate*	*ate*
	squeezed	*eat*	*peeled*	*meal*	*eat*	*consumed*	*consumed*
	cooked	*eating*	*cooked*	*salads*	*baking*	*tasted*	*eat*
Britain	*Northern*	*Ireland*	*Luxembourg*	*UK*	*UK*	*UK*	*England*
	Great	*Kingdom*	*Belgium*	*British*	*British*	*British*	*UK*
	Ireland	*Great*	*Madrid*	*London*	*England*	*America*	*Syria*

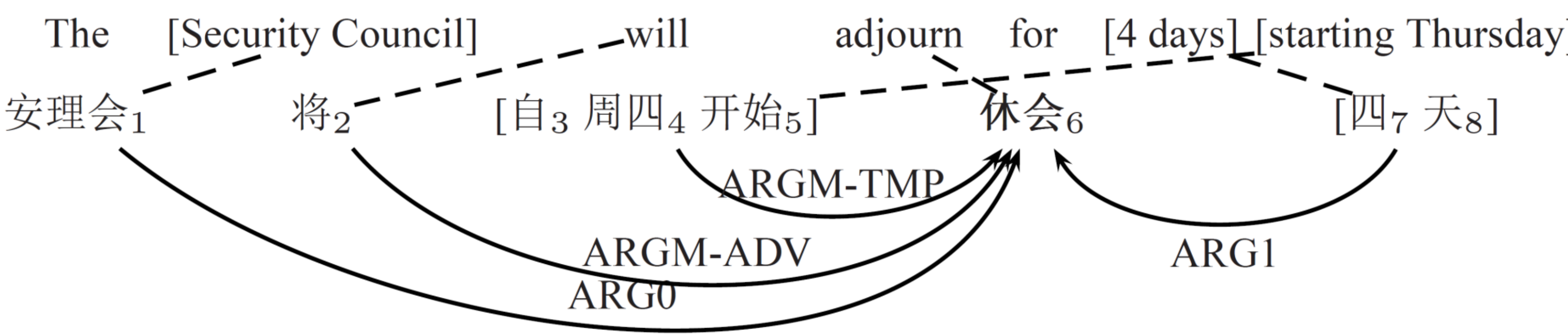
3 nearest neighbors for “teacher”, “eaten” and “Britain”, based on different word embeddings
From Hill et al. ICLR 2015

Semantics & Machine Translation

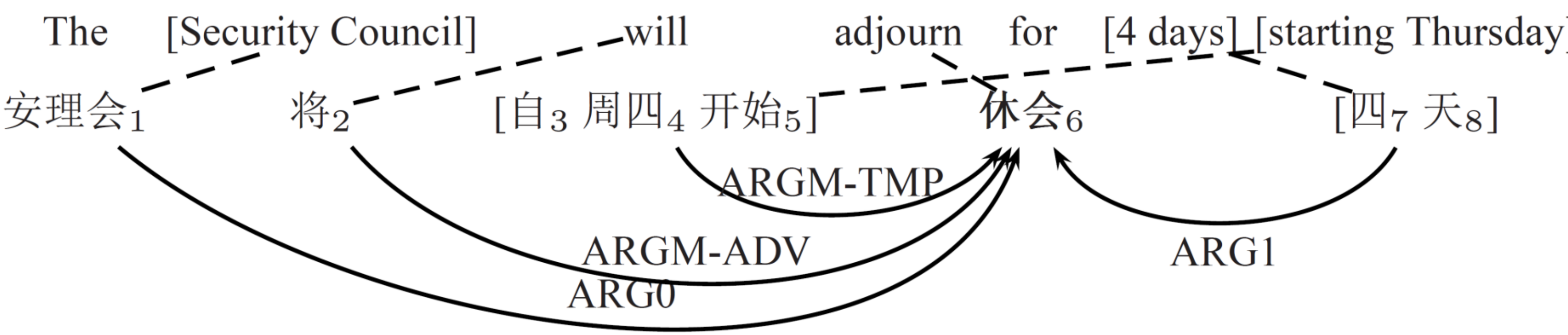
Case studies: MT with models of

- word senses
- word similarity
- **semantic structure**
- paraphrase

Predicate argument structure for MT



Predicate argument structure for MT

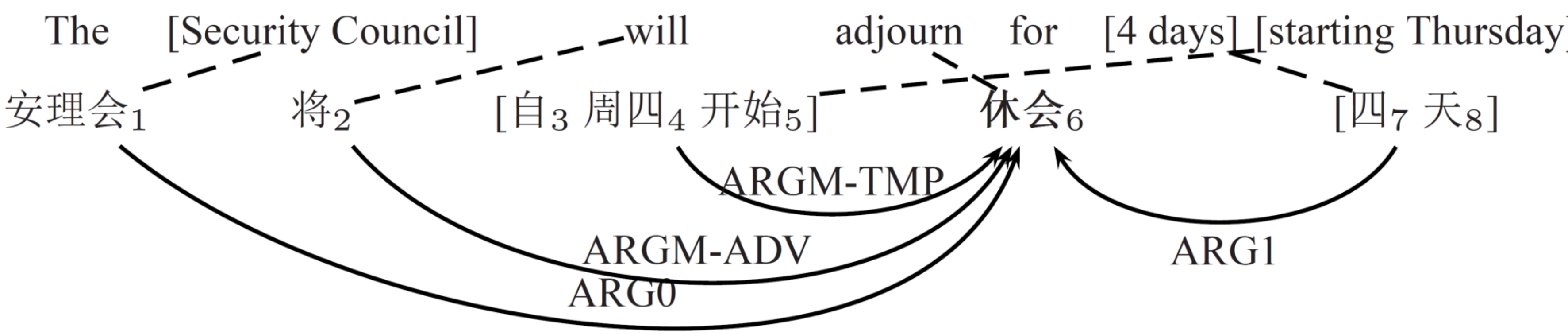


Chinese to English MT, with Chinese PAS

- Predicate translation model:
  - context-dependent lexicon P(English|Chinese predicate, context) where context includes PAS features

[Xiong et al. 2012]

Predicate argument structure for MT



Chinese to English MT, with Chinese PAS

- Argument reordering model:
  - predict relative position of English argument wrt English predicate
  - integrate in hierarchical MT system

[Xiong et al. 2012]

Other ways to use predicate argument structure

For reordering

- Detect PAS on MT output, and reorder in postprocessing step [Wu & Fung 2009]
- Detect POS on source and automatically reorder before MT [Wu et al. 2011]
- Soft reordering constraints based on source side semantic roles in a tree-to-string translation model [Liu & Gildea 2010]
- Soft semantic constraints not as useful if syntactic constraints already included in system [Li et al. 2014]

For MT evaluation

- Constrain matches between MT and reference using semantic roles [e.g., MEANT Lo et al.]

Semantics & Machine Translation

Case studies: MT with models of

- word senses
- word similarity
- predicate argument structure
- **paraphrases**

Detecting Paraphrases

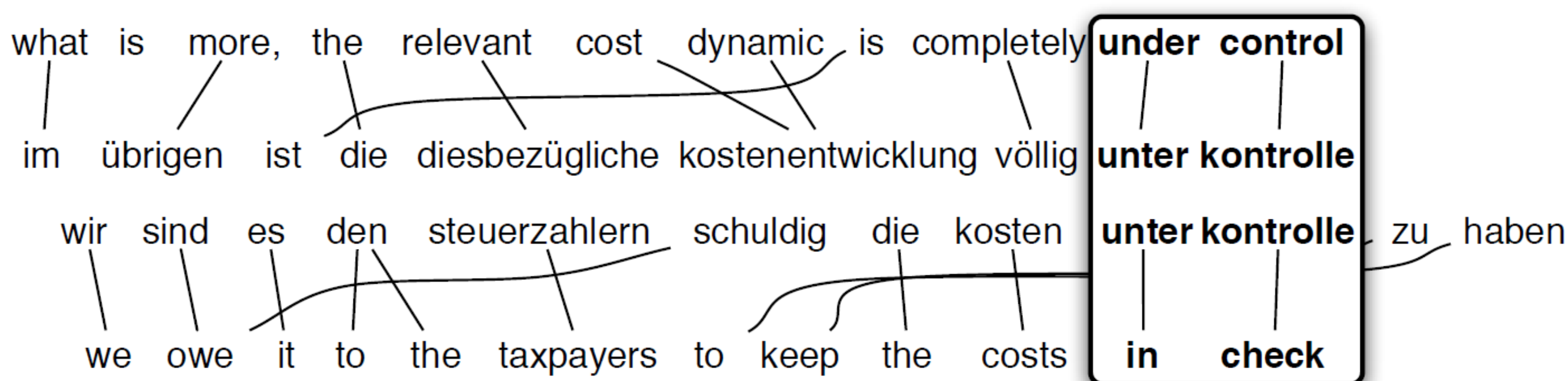
- Task: predict whether "`John broke the window`" and "`the window was broken by John`" are paraphrases

- Advantages of this formulation
  - A representation-agnostic form of semantic analysis
  - Can use similar models for word, phrase, sentence

Paraphrasing source language to improve MT

[Callison-Burch et al. 2006, among others]

– Paraphrase OOVs
– Paraphrases are learned by bilingual projection



Paraphrasing source language to improve MT

- Expand training data by paraphrasing noun compounds and prepositional phrases in source [Nakov 2008]
  - Hand-written paraphrasing rules

I welcome the Commissioner 's statement about the progressive and rapid beef import ban lifting .

I welcome the progressive and rapid beef import ban lifting Commissioner 's statement .

I welcome the Commissioner 's statement about the beef import ban 's progressive and rapid lifting .

I welcome the beef import ban 's progressive and rapid lifting Commissioner 's statement .

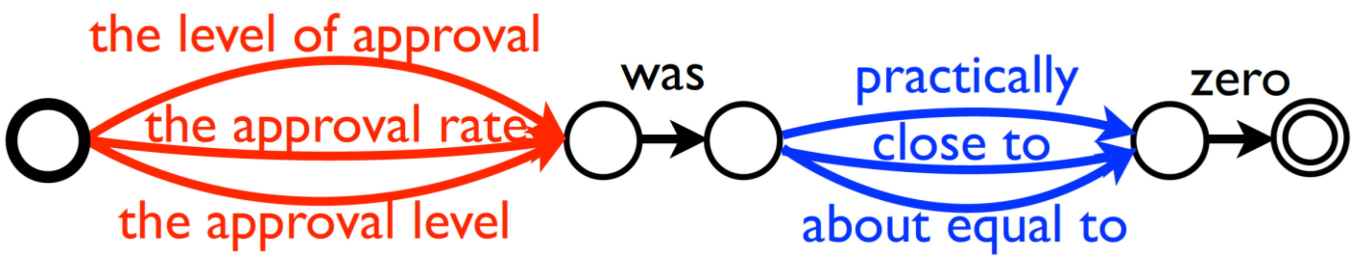
I welcome the Commissioner 's statement about the progressive and rapid lifting of the *ban on beef imports* .

Paraphrasing target language to improve MT

```
MT   the music director of the Tunisian orchestra
REF  the conductor of the Tunisian orchestra
REF’ the music director of the Tunisian orchestra
```

- [Madnani et al. 2007]
  - Goal
    - create additional reference translations for tuning
  - How?
    - view paraphrasing as monolingual MT
    - monolingual MT rules learned by pivoting

(Manually) paraphrasing references for MT evaluation



- HyTER networks [Dreyer and Marcu 2012]
  - requires significant annotation effort
  - encodes exponential number of reference translations for a given source sentence
  - improves MT evaluation (HyTER = TER + network)
  - but even networks that encode billions of translations do not contain all independently produced human translations!

Semantics for Machine Translation

Case studies: MT with models of

- word senses
- word similarity
- semantic structure
- paraphrases

Semantics for Machine Translation

- Many approaches, no clear single-best solution

- But common themes emerge

  – Application-driven semantic analysis
    - Word translations rather than WordNet senses
    - Paraphrases rather than PAS (?)

  – Models are often learned on same parallel corpus as MT

  – Semantic analysis often easier to integrate in evaluation than MT

(Machine) Translation for Semantics

Idea: view 2nd language as supervision/annotation for making predictions in the 1st language

- Cross-lingual projection
  [Yarowsky et al. 2001; Pado & Lapata 2009; Banea et al. 2008…]
- Learning semantics
  - paraphrases [Callison-Burch, Ganitkevitch et al. 2013…]
  - word senses [Diab 2004, Apidianaki 2008…]
  - semantic role induction [Titov & Klementiev 2012]
  - …

(Machine) Translation for Semantics

- Retrofitting sense-specific word vectors using parallel text
  [Ettinger, Resnik & Carpuat, NAACL 2016 (Short)]

- Sparse Bilingual Word Representations for Cross-lingual lexical entailment
  [Vyas & Carpuat, NAACL 2016]

- Learning Easy to Compose Word Representations using Bilingual Supervision
  [Elgohary & Carpuat, ACL 2016 (Short)]

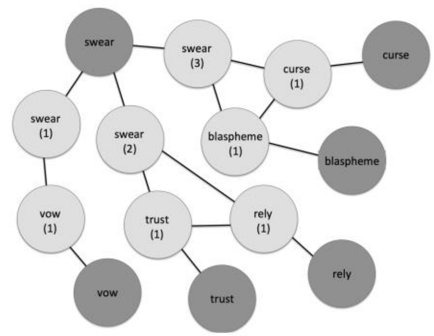


Figure 1: Illustration of WordNet-based sense graph.

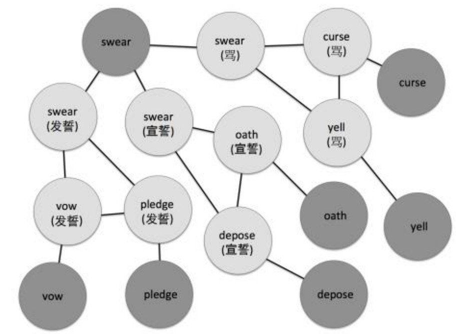


Figure 2: Illustration of parallel-text-based sense graph.

English-English	**English-French**
affection $\rightarrow$ feeling	affection $\rightarrow$ sentiment
aspirin $\rightarrow$ drug	aspirin $\not\rightarrow$ drogue
water $\rightarrow$ wet	water $\rightarrow$ humide
feeling $\not\rightarrow$ nostalgia	feeling $\not\rightarrow$ nostalgie

Semantics & Machine Translation

- Intersection of semantics & MT is a fertile area!
- Still many long-standing open questions
- Neural models offer an opportunity to revisit them with new tools

Semantics & Machine Translation

Marine Carpuat
University of Maryland
marine@cs.umd.edu