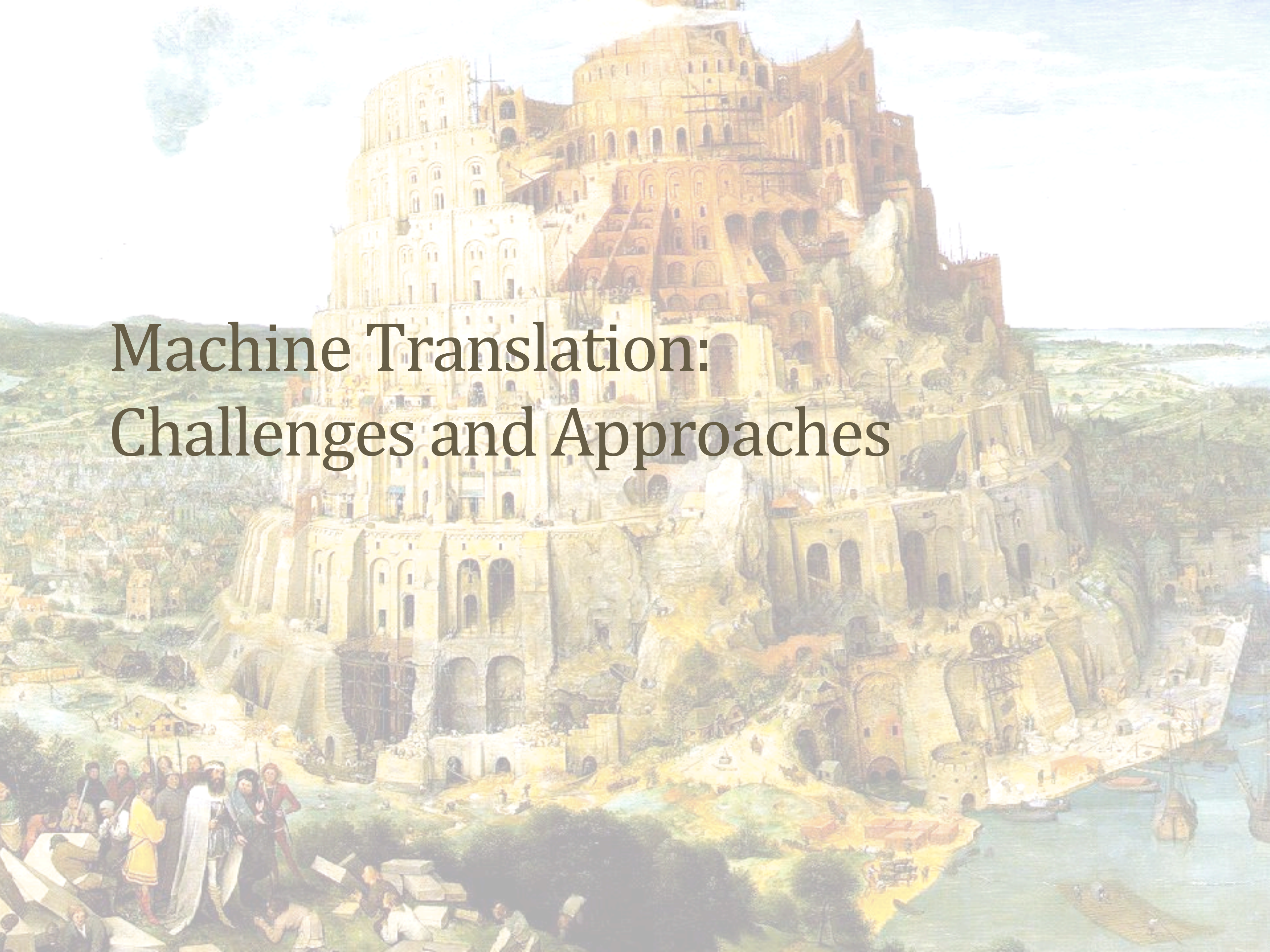


Machine Translation: Challenges and Approaches

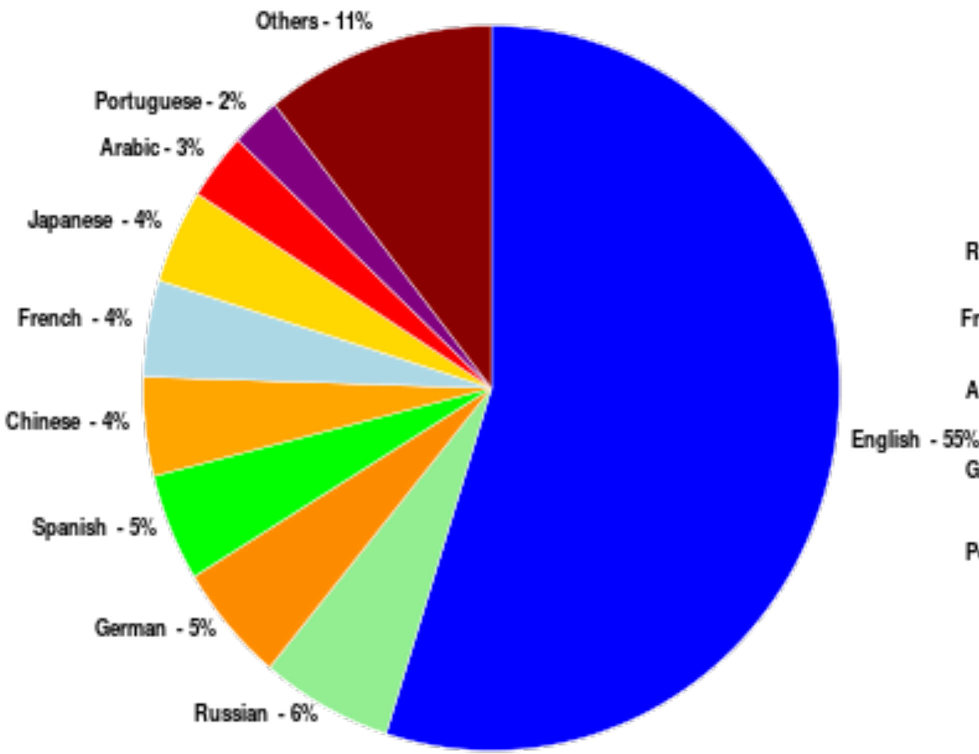
Announcements

- Final exam, Dec. 21st, 1;10-4PM
- Dan Jurafsky, Stanford Univ., "Does This Vehicle Belong to You?" Processing the Language of Policing for Improving Police-Community Relations, Dec. 5th, 5pm Davis Auditorium
- Rupal Patel, Northwestern Univ., Speech recordings – a life altering form of biological donation, Dec. 4th, 11:30AM, Davis Auditorium

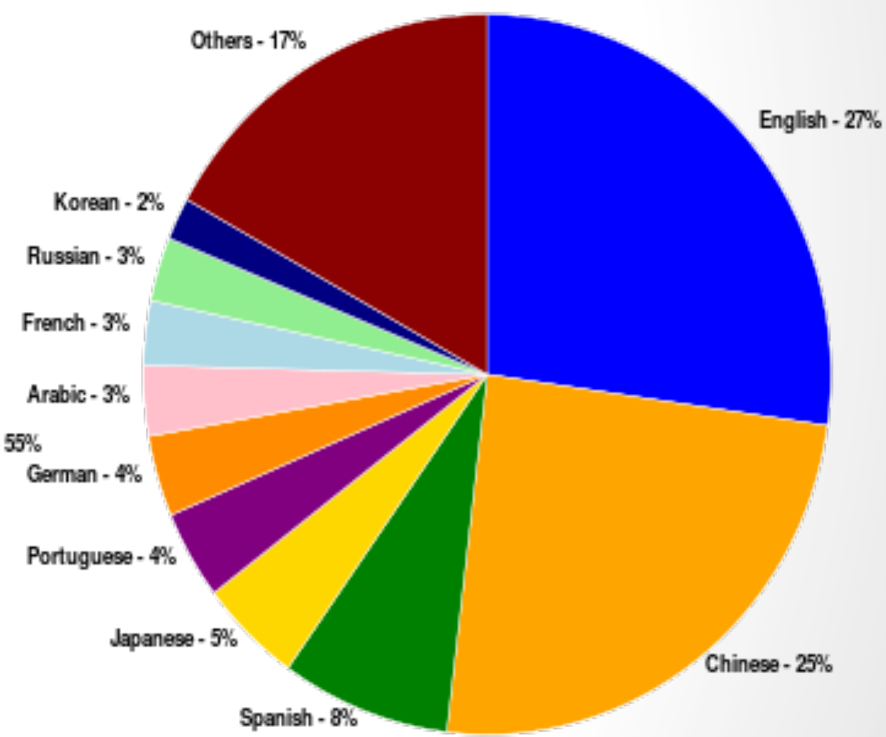
Slide from Radev

Multilingual Users

- Content languages for websites



Percentage of Internet users by language



April 2013

http://en.wikipedia.org/wiki/Global_Internet_usage

Google Translate

Yiddish
Yoruba
Zulu

Afrikaans	Bulgarian	Greek	German	Igno	Kurdish	Malayalm	Polish	sindhi	Tamil
Albanian	Catalan	English	Gujarati	Indonesian	Kyrgyz	Maltese	Portuguese	Sinhala	Telugu
Amharic	Cebuano	Esperanto	HaitianCreole	Irish	Lao	Maori	Punjabi	Slovak	Thai
Arabic	Chichewa	Estonian	Hausa	Italian	Latin	Marathi	Romanian	Sloveian	Turkish
Armenian	Chinese	Filipino	Hawaiian	Japanese	Latvian	Mongolian	Russian	Somali	Ukranian
Azerbaijani	Corsican	Finnish	Hebrew	Javanese	Lithuanian	Myanmar	Samoan	Spanish	Urdu
Basque	Croatian	French	Hindi	Kannada	Luxembourgish	Nepala	Scots Gaelic	Sundanese	Uzbek
Belarusian	Czech	Frisian	Hmong	Kazakh	Macedonian	Norwegian	Serbian	Swahili	Veitnamese
Bengali	Danish	Galician	Hungarian	Khmer	Malagasy	Pashto	Sesotho	Swedish	welsh
Bosnian	Dutch	Georgian	Icelandic	Korean	Malay	Persian	Shona	Tajik	Xhosa

Thank you for your attention!

Questions?

- Romance languages handled well
- Similar language pairs handled well (e.g., Spanish, Portuguese)
- Formal genres handled better

→ Still many problems!

Today

- Multilingual Challenges for MT

- MT Approaches
  - Statistical
  - Neural net (Thursday)

- MT Evaluation

Today

- Multilingual Challenges for MT
- MT Approaches
  - Statistical
  - Neural net
- MT Evaluation

Multilingual Challenges

- Orthographic Variations
  - Ambiguous spelling
    - كتب الاولاد اشعارا   كَتَبَ الأَوْلادُ اشعَاراً
  - Ambiguous word boundaries
    - 美单方削减中国纺织品出口配额
- Lexical Ambiguity
  - **Bank ➔** بنك (financial) vs. ضفة(river)
  - **Eat ➔** essen (human) vs. fressen (animal)

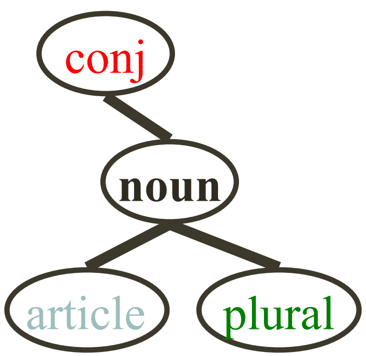
Slide from Nizar Habash

Multilingual Challenges
Morphological Variations

- Affixation vs. Root+Pattern

write	→	**written**	كتب	→	مكتوب
kill	→	**killed**	قتل	→	مقتول
do	→	**done**	فعل	→	مفعول

- Tokenization

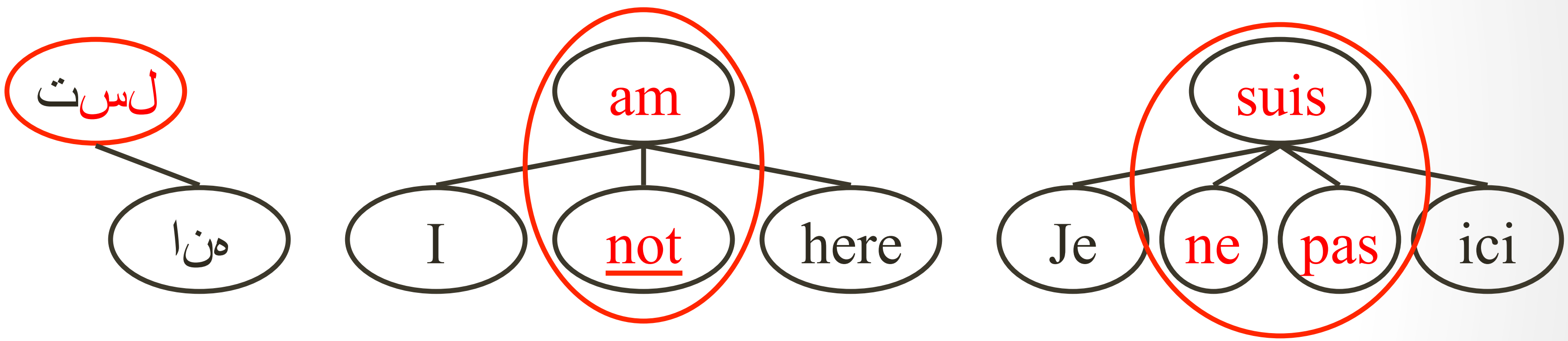


And the cars	→	*and the cars*
والسيارات	→	w Al SyArAt
Et les voitures	→	et le voitures

Slide from Nizar Habash

Translation Divergences

conflation



لست انه
I-am-not here

I am not here

Je ne suis pas ici
I not am not here

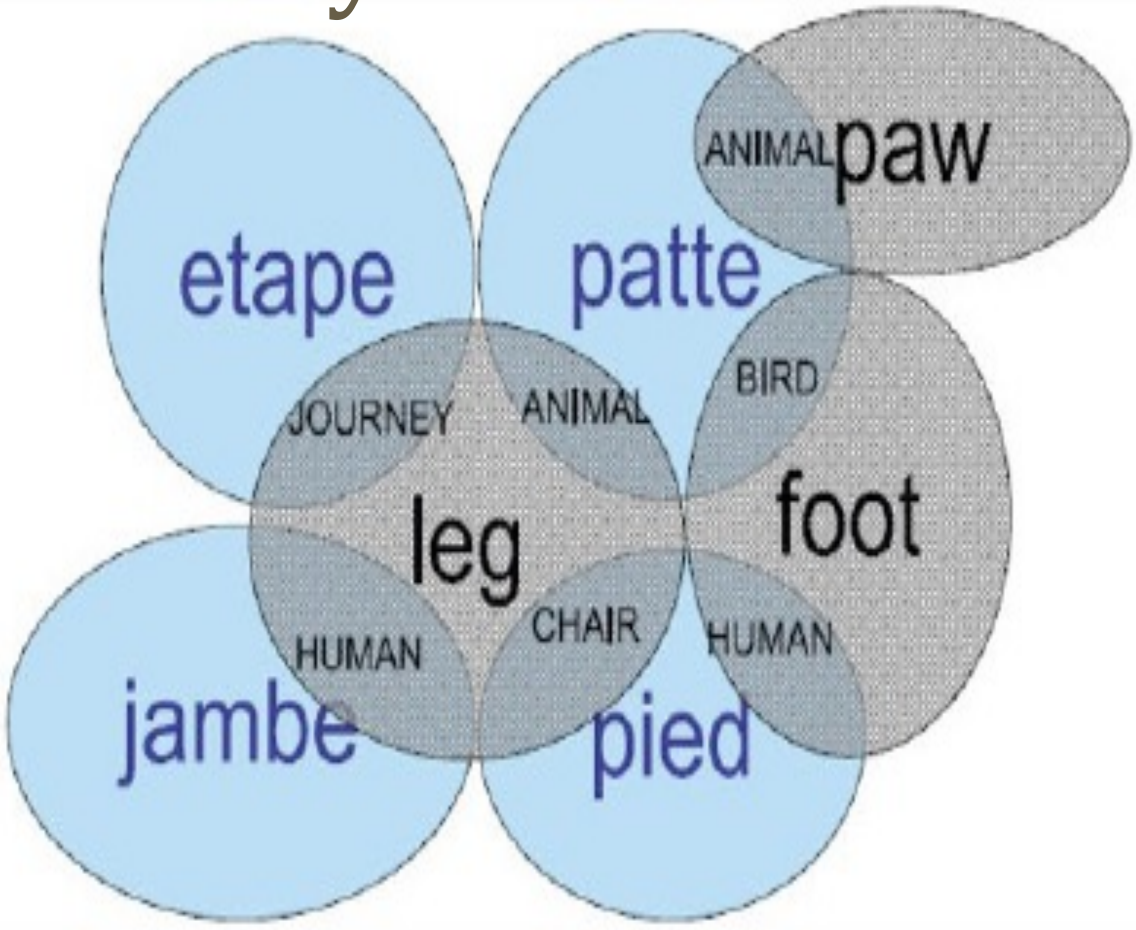
Slide from Nizar Habash

Translation Divergences

English	John swam across the river quickly
Spanish	Juan cruzó rapidamente el río nadando *Gloss: John crossed fast the river swimming*
Arabic	اسرع جون عبور النهر سباحة *Gloss: sped john crossing the-river swimming*
Chinese	约翰 快速 地 游 过 这 条 河 *Gloss: John quickly (DE) swam cross the (Quantifier) river*
Russian	Джон быстро переплыл реку *Gloss: John quickly cross-swam river*

Slide from Nizar Habash

Language Differences - vocabulary



[Example from Jurafsky and Martin]

Language Differences - Syntax

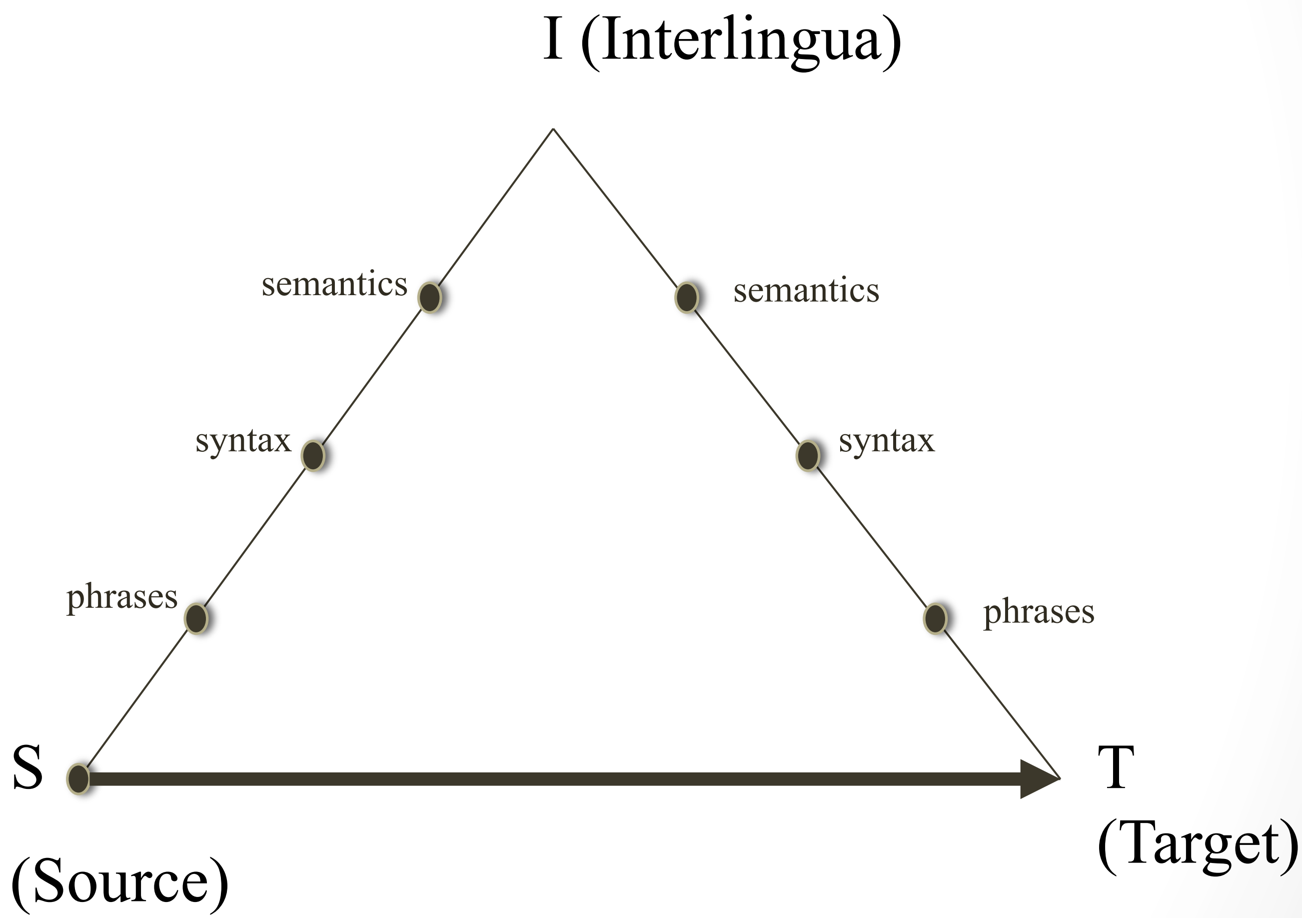
- Word order
  - SVO: English, Mandarin
  - VSO: Irish, Classical Arabic
  - SOV: Hindi, Japanese
- Word order in phrases (Fr.)
  - la maison bleue, the blue house
- Word order in sentences (Jap.)
  - I like to drink coffee
  - watashi wa kohii o nomu no ga suki desu
  - I-subj coffee-obj drink-dat-rheme like
- Prepositions (Jap.)
  - to Mariko, Mariko-ni

Slide adapted from Radevc

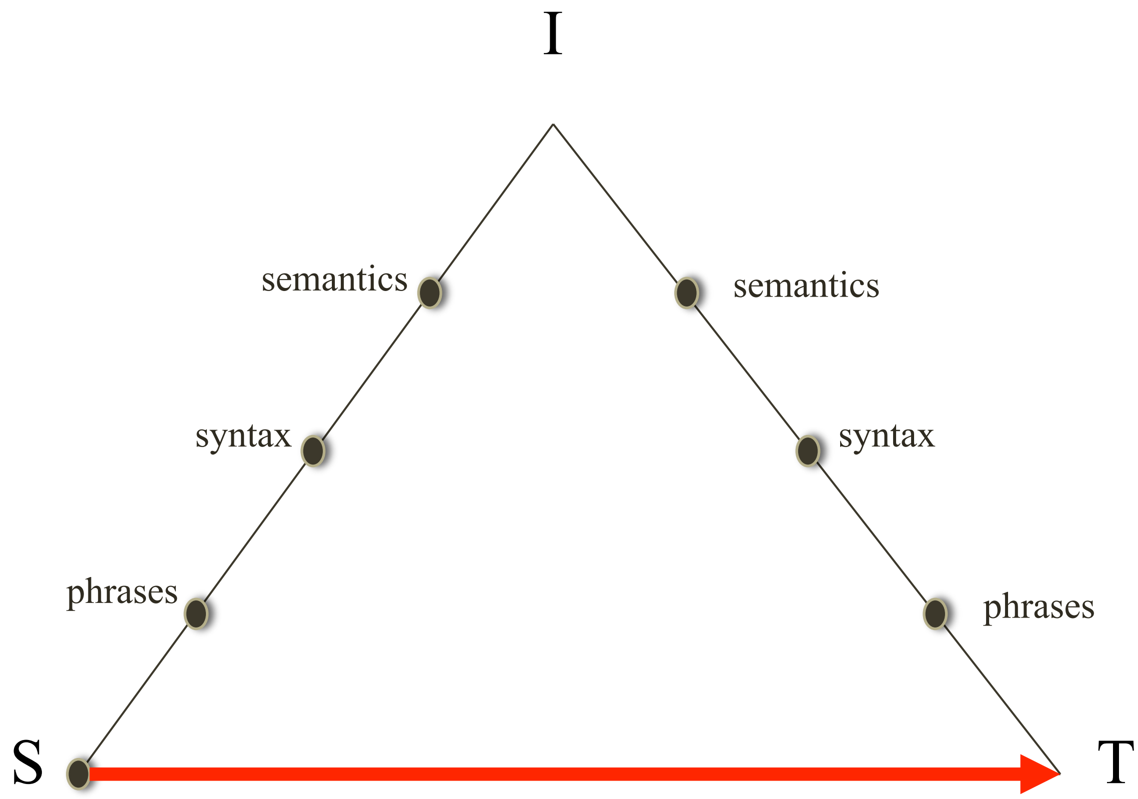
Today

- Multilingual Challenges for MT
- MT Approaches
  - Statistical
  - Neural net
- MT Evaluation

MT Approaches
MT Pyramid



String-to-String Translation



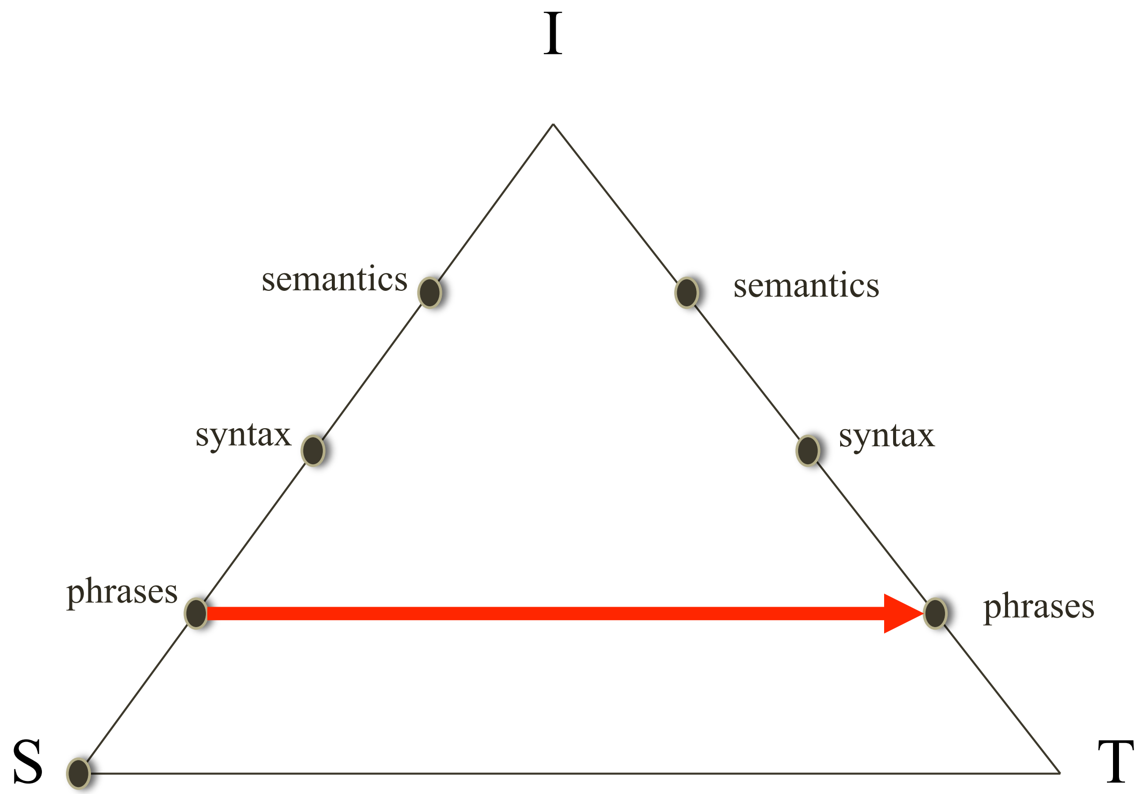
Slide from Radev

MT Approaches
Gisting Example

Sobre la base de dichas experiencias se estableció en 1988 una metodología.

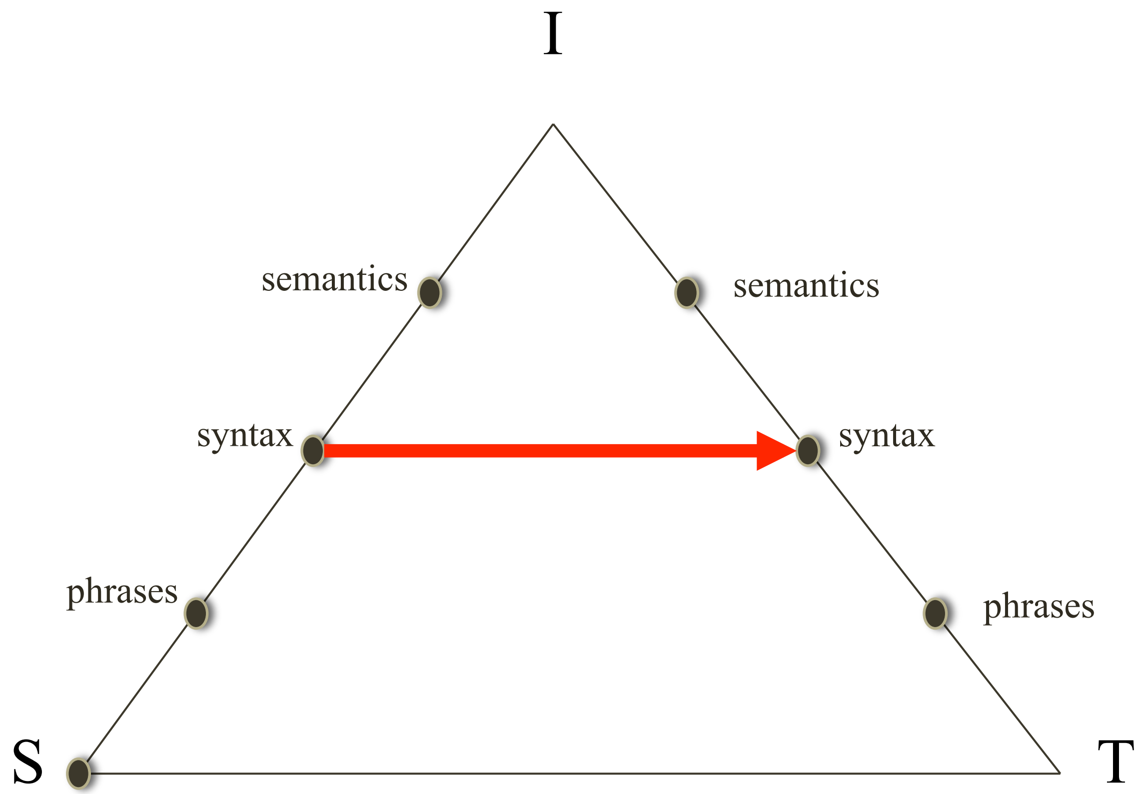
On the basis of these experiences, a methodology was arrived at in 1988.

Phrase-Based Translation



Slide from Radev

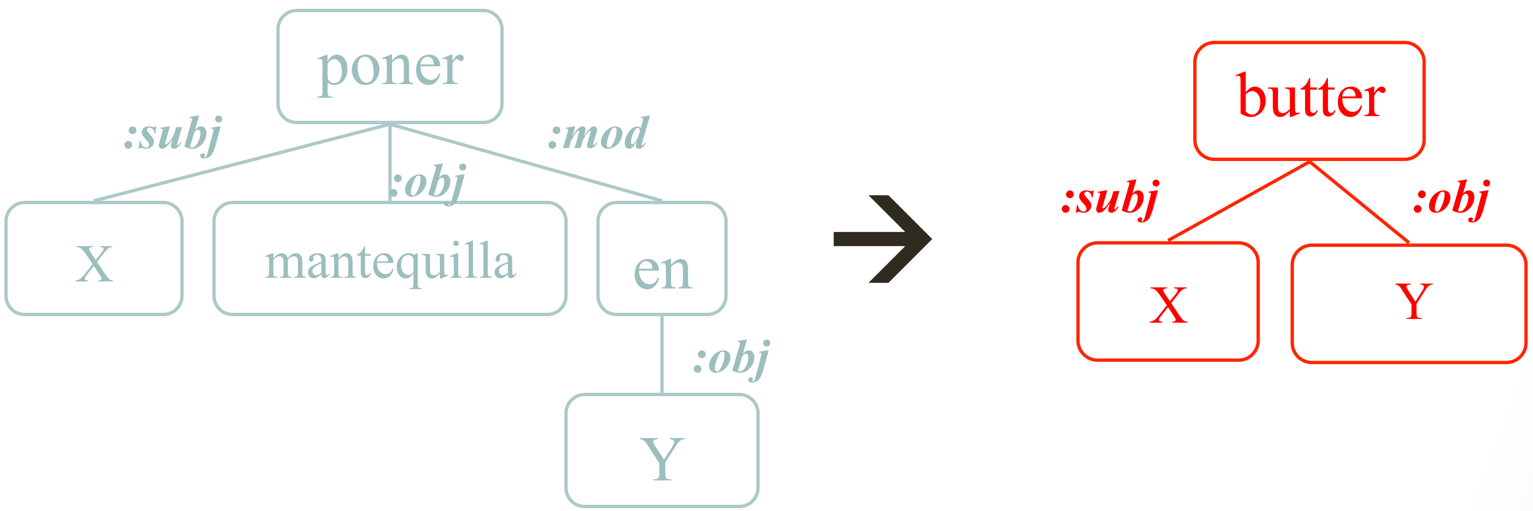
Tree-to-Tree Translation



Slide from Radev

MT Approaches
Transfer Example

- Transfer Lexicon
  - Map SL structure to TL structure

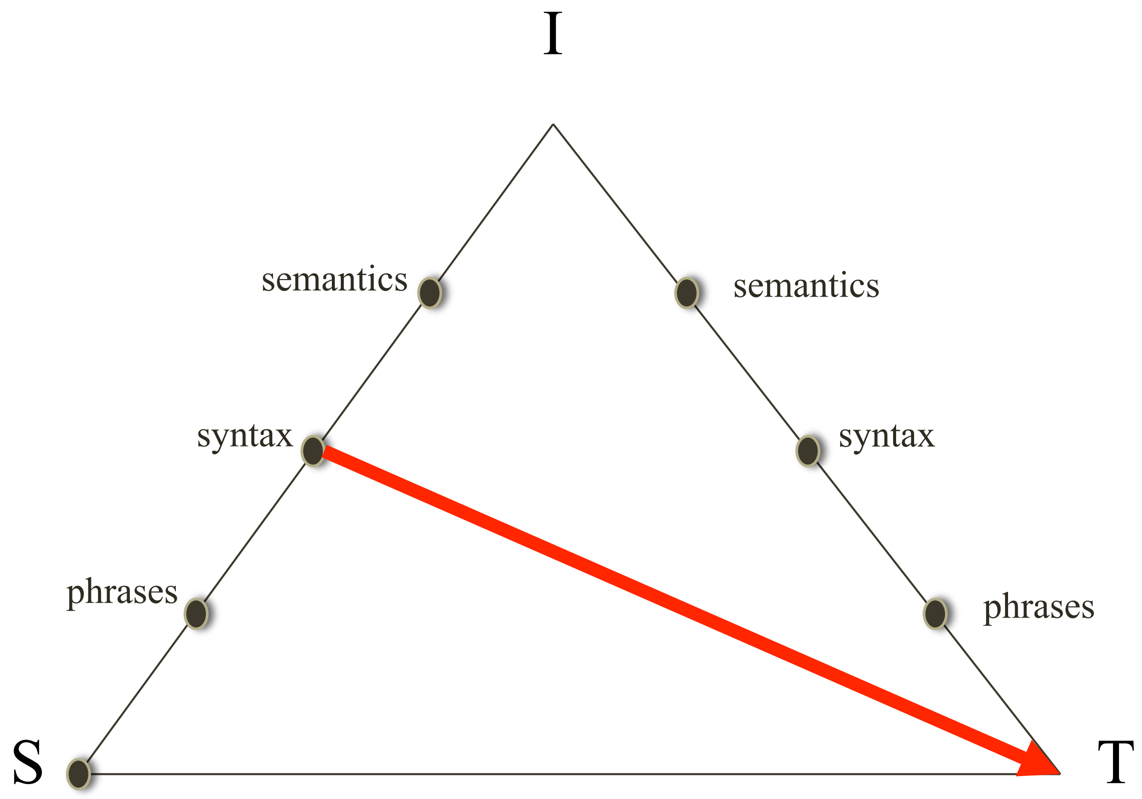


X puso mantequilla en Y

X buttered Y

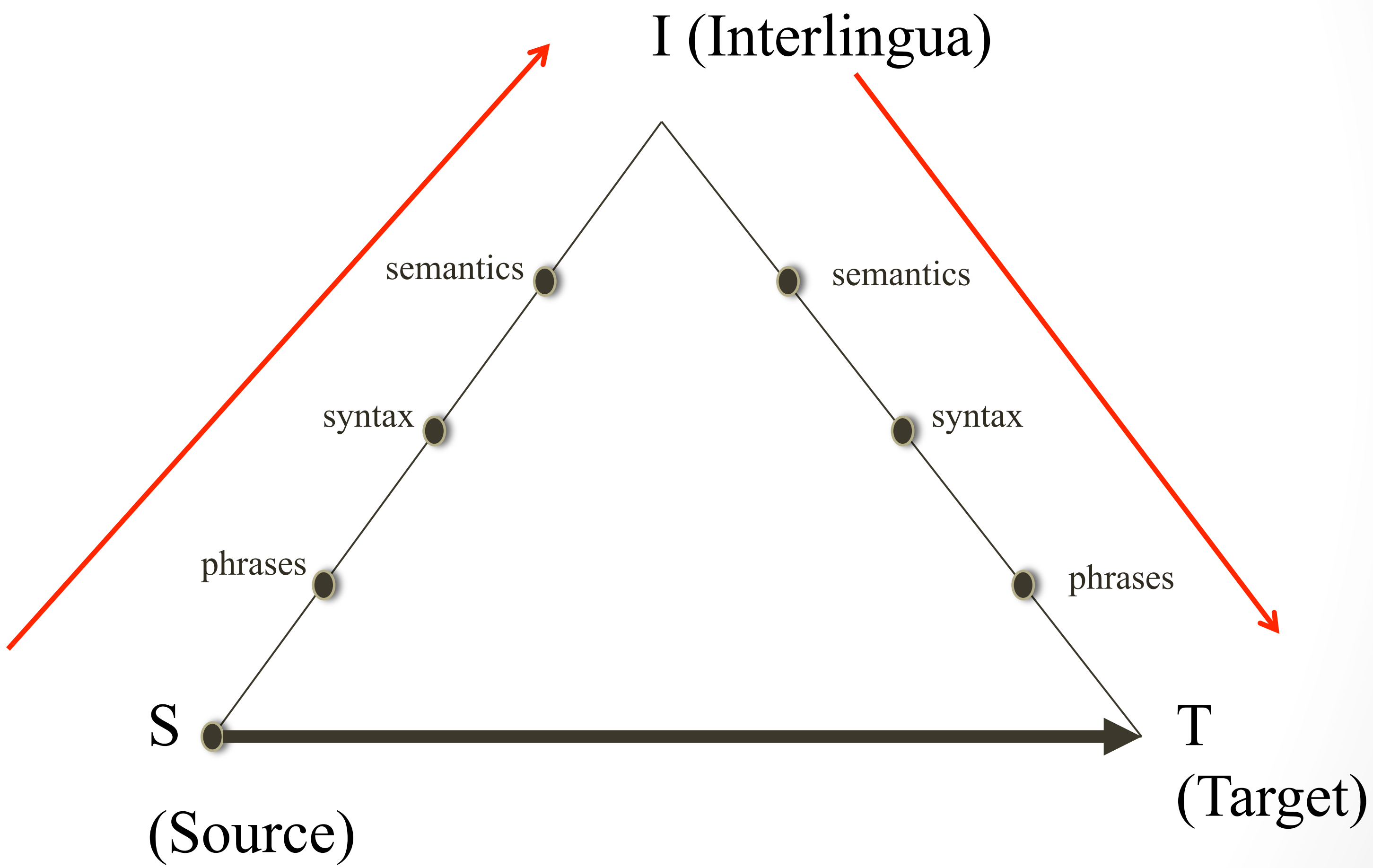
Slide from Nizar Habash

Tree-to-String Translation



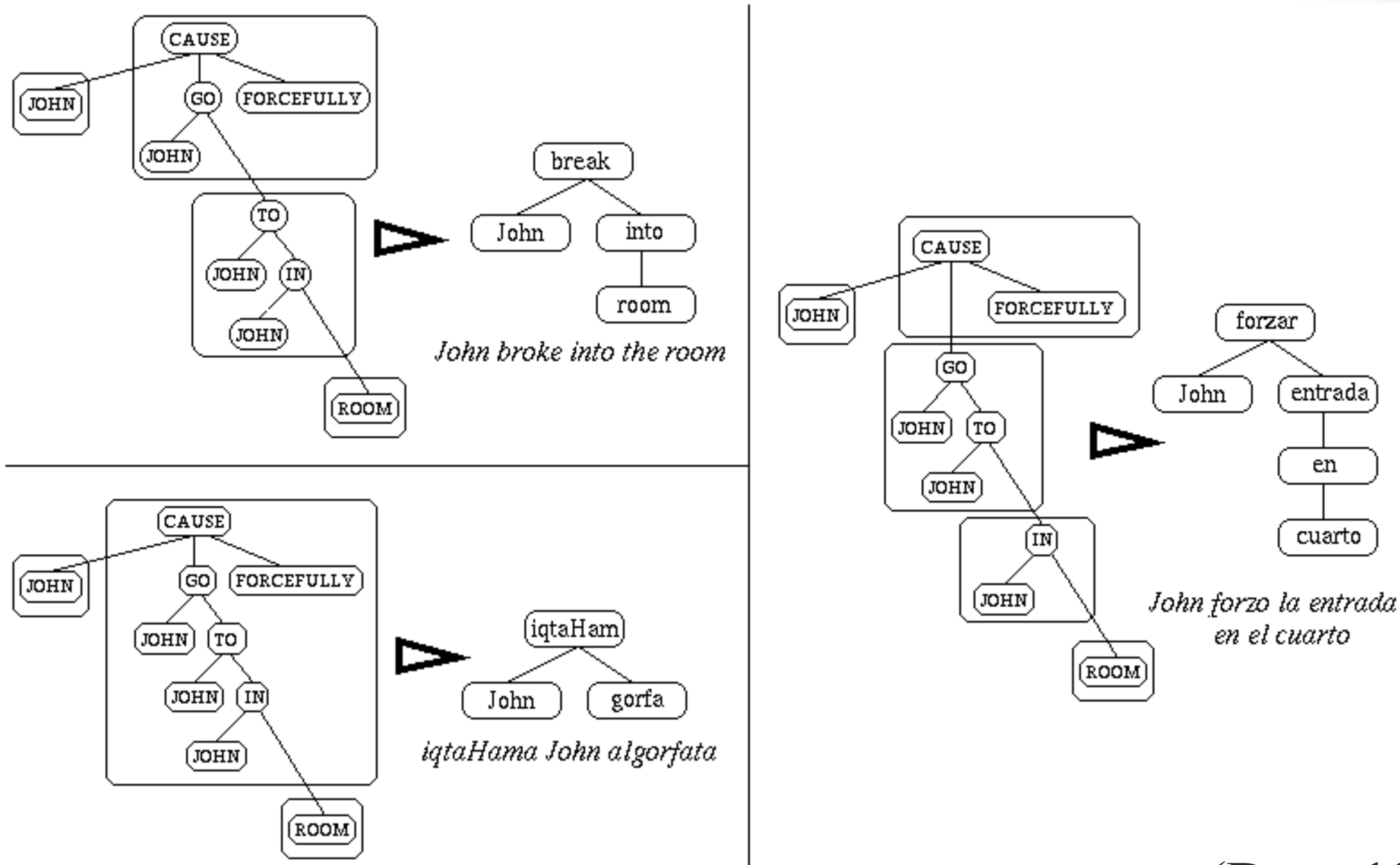
Slide from Radev

MT Approaches
MT Pyramid



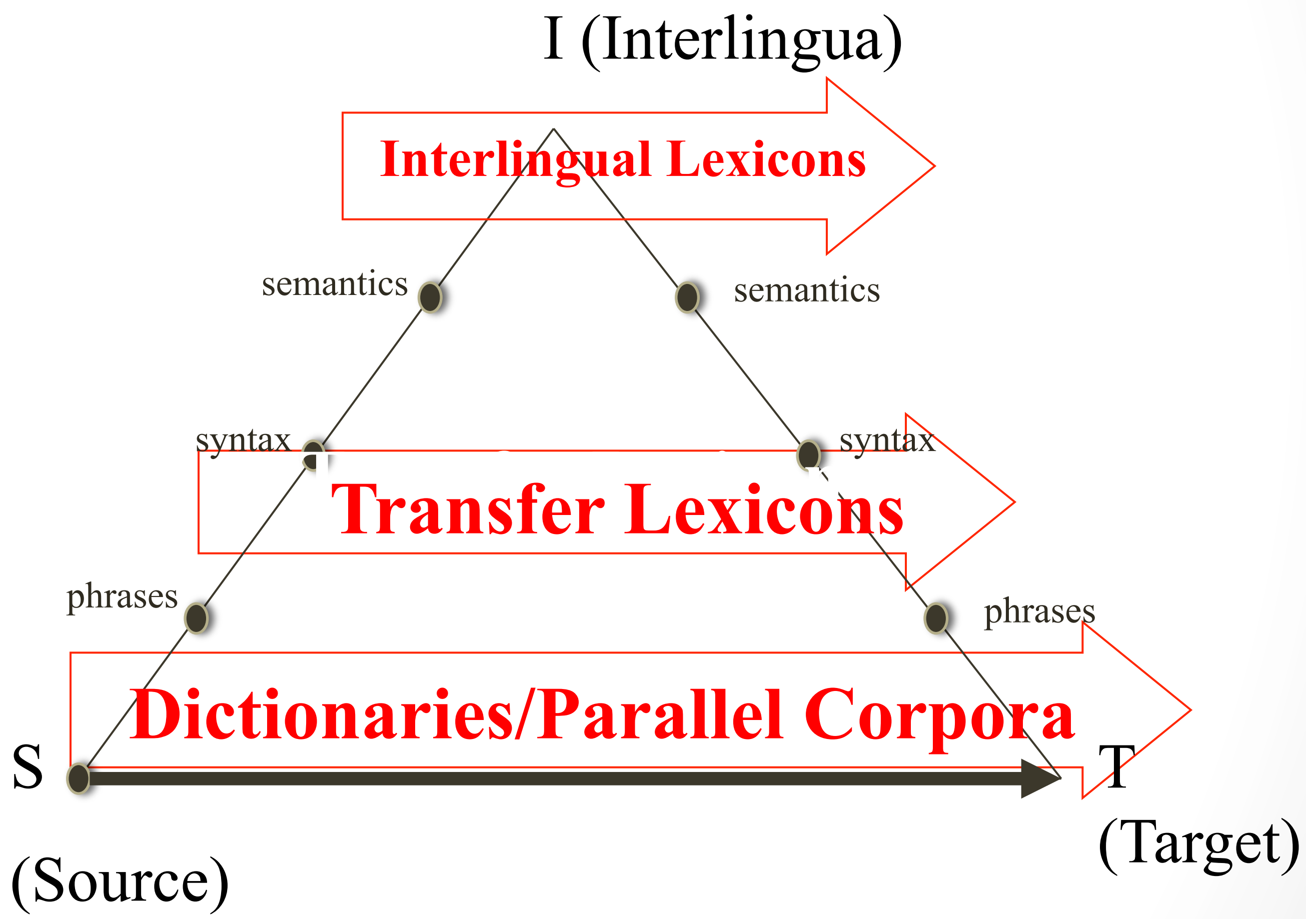
MT Approaches

Interlingua Example: Lexical Conceptual Structure



(Dorr, 1993)

MT Approaches
MT Pyramid



Today

- Multilingual Challenges for MT
- MT Approaches
  - Statistical
  - Neural net
- MT Evaluation

Translation as Decoding

- "One naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.' "
  - Warren Weaver, "Translation (1955)"

Slide from Radev

The first parallel corpus: The Rosetta Stone

Carved in 196 BC in Egypt
Deciphered by Champollion in 1822
Mixture of Egyptian (hieroglyphs and Demotic) and Greek

http://www.ancientegypt.co.uk/writing/rosetta.html

Slide from Radev

Europarl: A Parallel Corpus for Statistical Machine Translation

- Proceedings of the European Parliament
- 21 European languages
  - Romanic (French, Italian, Spanish, Portuguese, Romanian), Germanic (English, Dutch, German, Danish, Swedish), Slavik (Bulgarian, Czech, Polish, Slovak, Slovene), Finni-Ugric (Finnish, Hungarian, Estonian), Baltic (Latvian, Lithuanian), and Greek
- 60 million words/language
- Must be aligned first

Koehn, MT Summit, 2005

http://homepages.inf.ed.ac.uk/pkoehn/publications/europarl-mtsummit05.pdf

Danish: det er næsten en personlig rekord for mig dette efterår .
German: das ist für mich fast persönlicher rekord in diesem herbst .
Greek: πρόκειται για το προσωπικό μου ρεκόρ αυτό το φθινόπωρο .
English that is almost a personal record for me this autumn !
Spanish: es la mejor marca que he alcanzado este otoño .
Finnish: se on melkein minun ennätykseni tänä syksynä !
French: c ' est pratiquement un record personnel pour moi , cet automne !
Italian: e ' quasi il mio record personale dell ' autunno .
Dutch: dit is haast een persoonlijk record deze herfst .
Portuguese: é quase o meu recorde pessoal deste semestre !
Swedish: det är nästan personligt rekord för mig denna höst !

Figure 2: One sentence aligned across 11 languages

Koehn, MT Summit, 2005
http://homepages.inf.ed.ac.uk/pkoehn/publications/europarl-mtsummit05.pdf

What other parallel corpora can you think of?

Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

Statistical MT
Noisy Channel Model

Portions from http://www.clsp.jhu.edu/ws03/preworkshop/lecture_yamada.pdf

Statistical MT

Translate from French: “une fleur rouge”?

	p(e)	**p(f|e)**	**p(e)*p(f|e)**
1. a flower red			
2. red flower a			
3. flower red a			
4. a red dog			
5. dog cat mouse			
6. a red flower			

Slide from Radev

Which phrases have high p(e)?

2

3

4

5

Statistical MT

Translate from French: “une fleur rouge”?

	p(e)	**p(f|e)**	**p(e)*p(f|e)**
1. a flower red	Low		
2. red flower a	Low		
3. flower red a	Low		
4. a red dog	High		
5. dog cat mouse	Low		
6. a red flower	High		

Slide from Radev

…ich phrases have high probability under p(f|…

phrases 1-3, 6

phrase 4

phrase 5

Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

Statistical MT

Translate from French: “une fleur rouge”?

	p(e)	p(f|e)	p(e)*p(f|e)
1. a flower red	Low	High	
2. red flower a	Low	High	
3. flower red a	Low	High	
4. a red dog	High	Low	
5. dog cat mouse	Low	Low	
6. a red flower	High	High	

Slide from Radev

Statistical MT

Translate from French: “une fleur rouge”?

	p(e)	p(f|e)	p(e)*p(f|e)
1. a flower red	Low	High	Low
2. red flower a	Low	High	Low
3. flower red a	Low	High	Low
4. a red dog	High	Low	Low
5. dog cat mouse	Low	Low	Low
6. a red flower	High	High	High

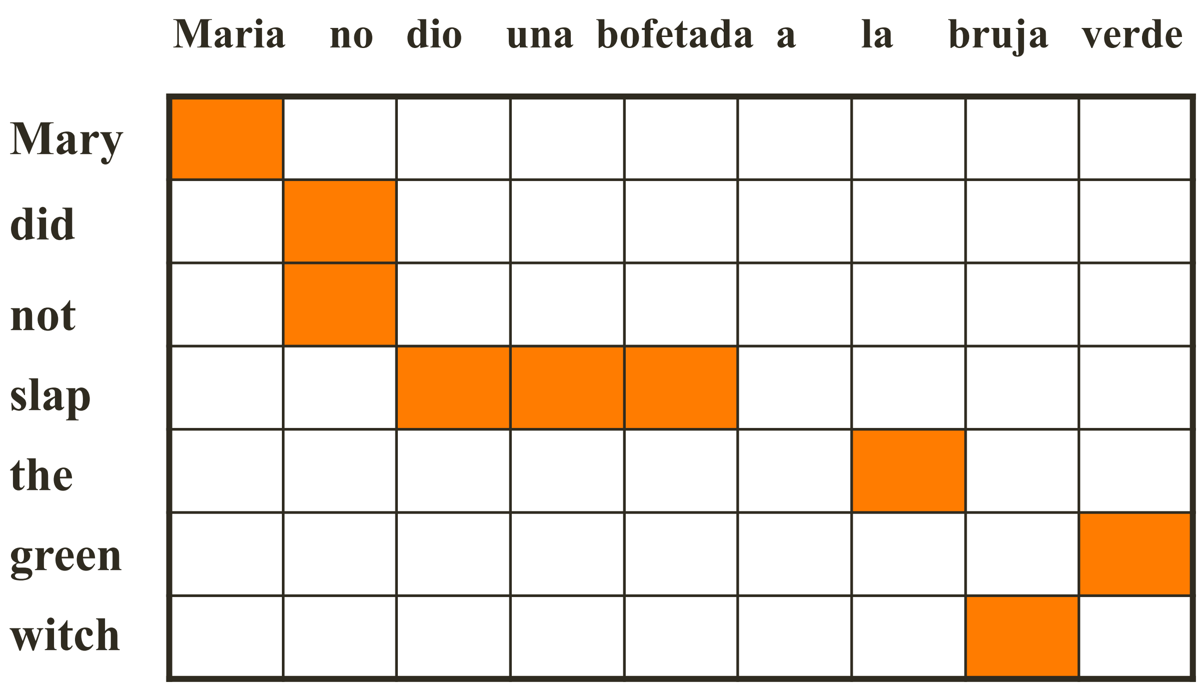
Slide from Radev

Slide based on Kevin Knight's http://www.sims.berkeley.edu/courses/is290-2/f04/lectures/mt-lecture.ppt

Statistical MT

Automatic Word Alignment

- GIZA++
  - A statistical machine translation toolkit used to train word alignments.
  - Uses Expectation-Maximization with various constraints to bootstrap alignments



Slide from Nizar Habash

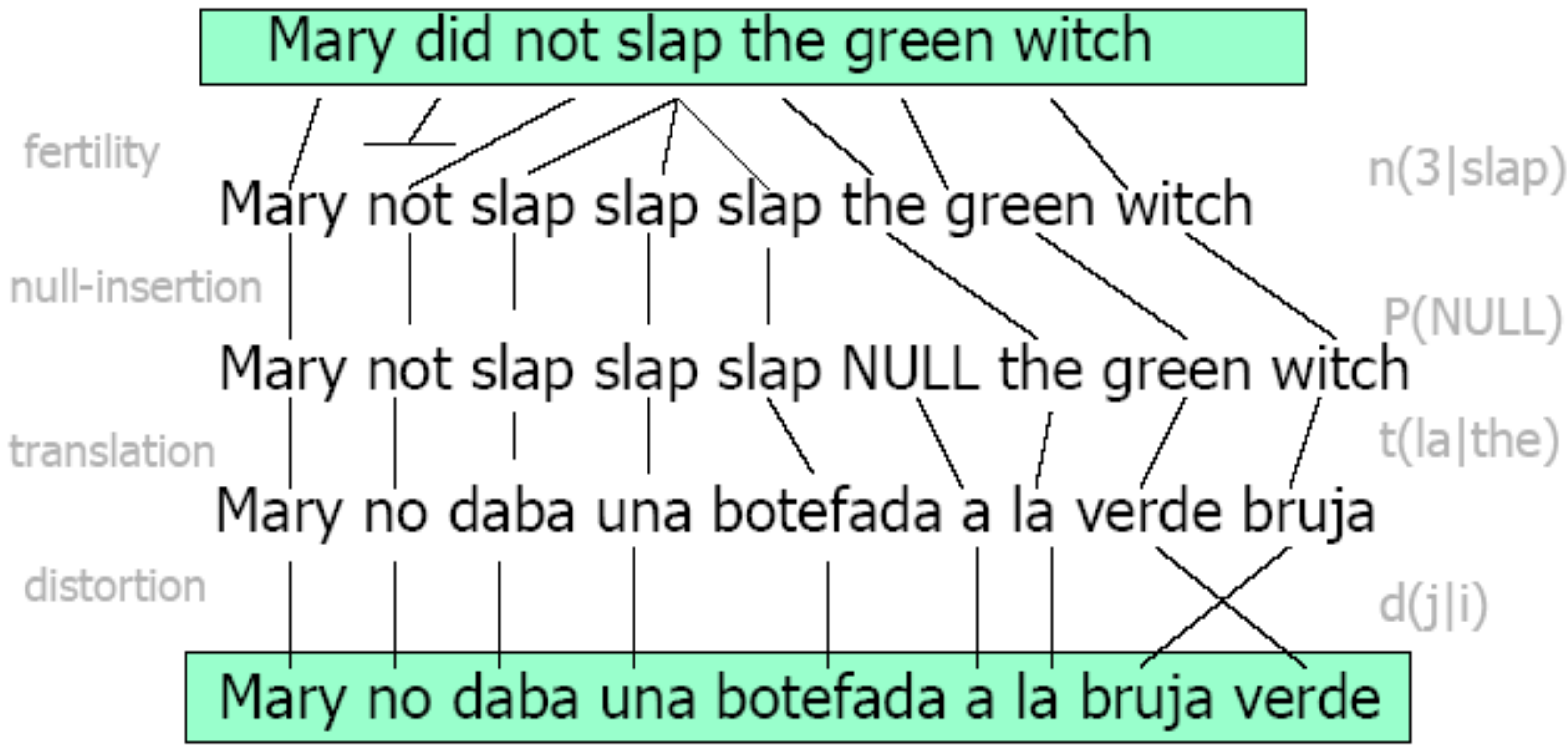
nstraints might be used to bootstrap word ali

Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

Statistical MT

IBM Model (Word-based Model)



IBM's EM trained models (1-5)

- Word translation
- Local alignment
- Fertilities
- Class-based alignment
- Re-ordering

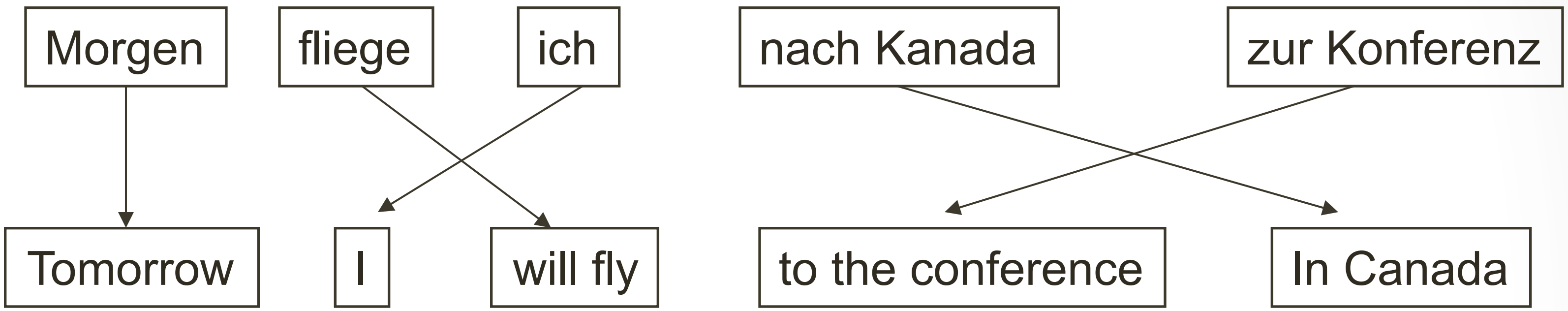
All are separate models to train!

Model 1:

$$p(f,a \mid e) = p(a \mid e) * p(f \mid a,e) = \frac{c}{(n+1)^m} \prod_{j=1}^{m} p(f_j \mid e_{a_j})$$

Slide courtesy of Kevin Knight http://www.sims.berkeley.edu/courses/is290-2/f04/lectures/mt-lecture.ppt

Phrase-Based Statistical MT



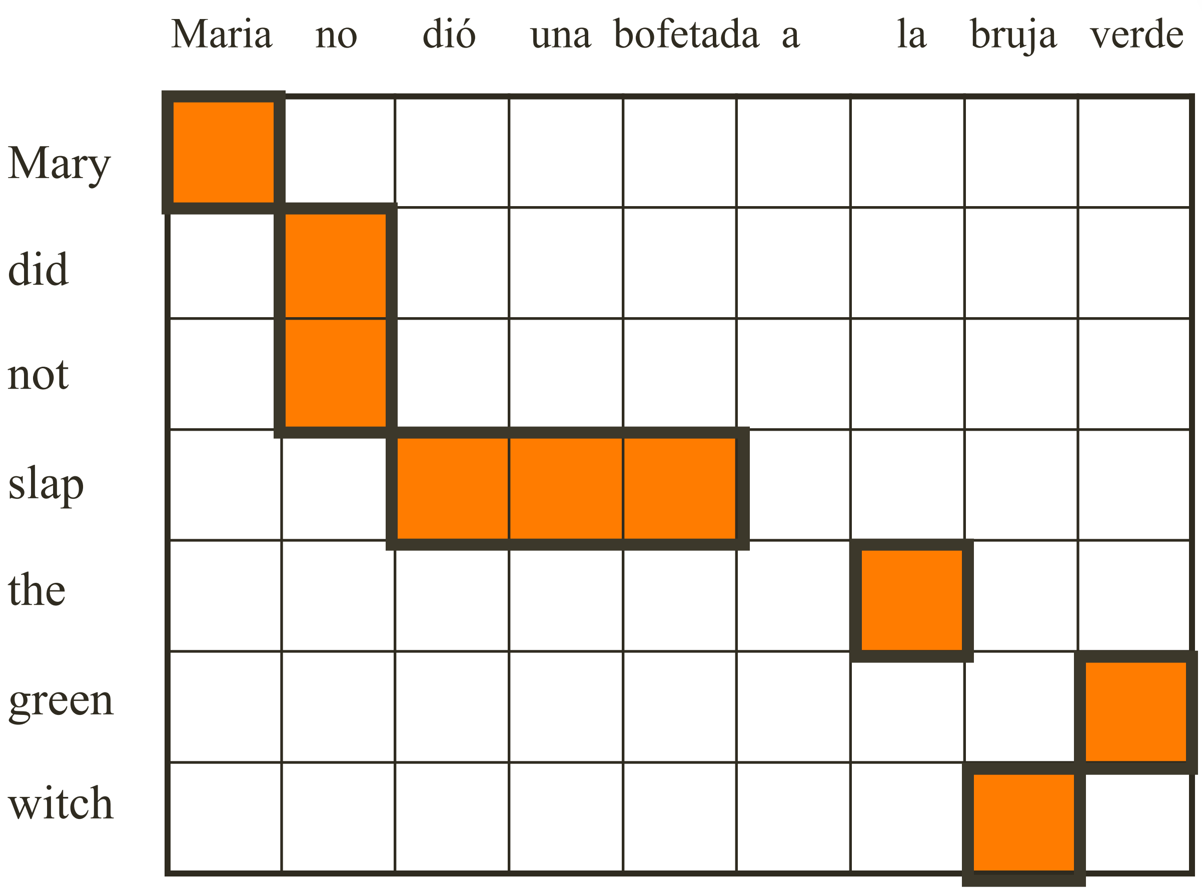
- Foreign input segmented in to phrases
  - "phrase" is any sequence of words
- Each phrase is probabilistically translated into English
  - P(to the conference | zur Konferenz)
  - P(into the meeting | zur Konferenz)
- Phrases are probabilistically re-ordered

See [Koehn et al, 2003] for an intro.

This was state-of-the-art before neural MT

Slide courtesy of Kevin Knight http://www.sims.berkeley.edu/courses/is290-2/f04/lectures/mt-lecture.ppt

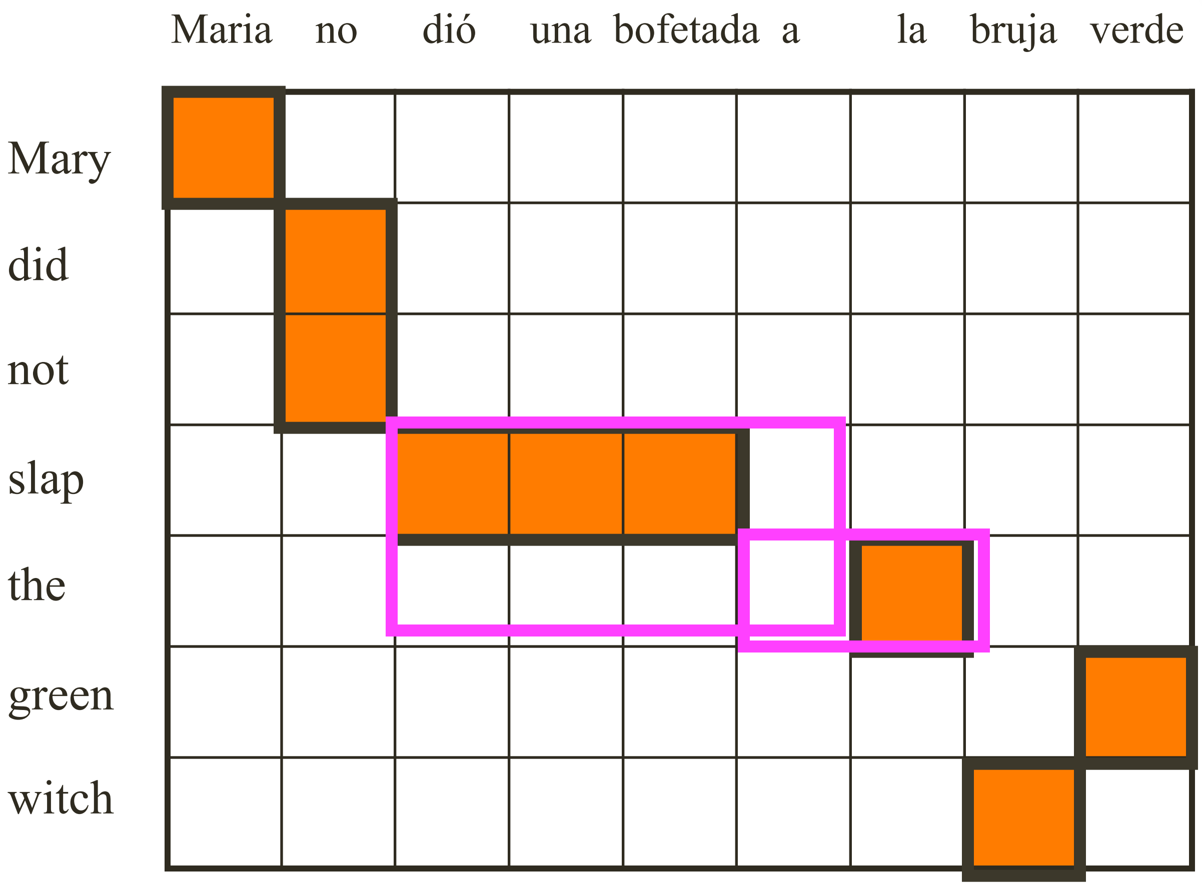
Word Alignment Induced Phrases



(Maria, Mary) (no, did not) (slap, dió una bofetada) (la, the) (bruja, witch) (verde, green)

Slide courtesy of Kevin Knight http://www.sims.berkeley.edu/courses/is290-2/f04/lectures/mt-lecture.ppt

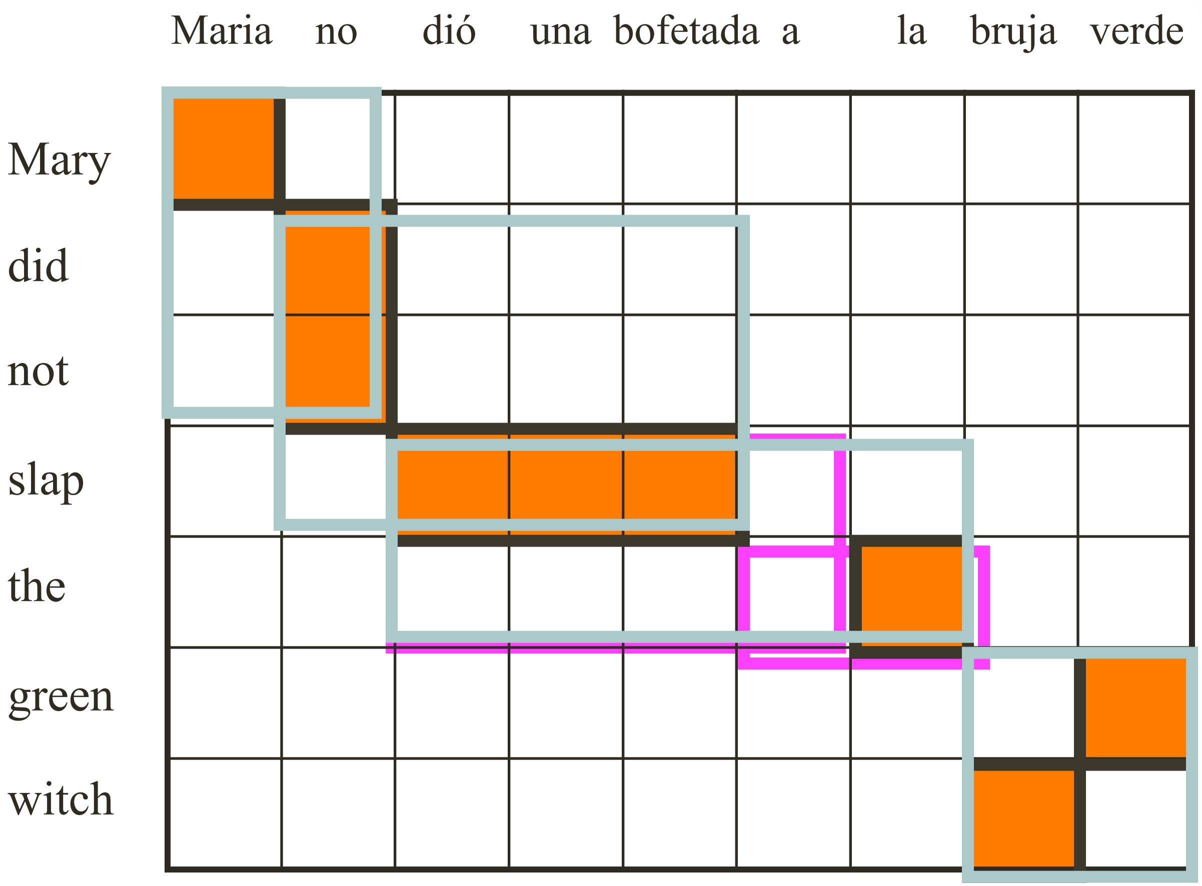
Word Alignment Induced Phrases



(Maria, Mary) (no, did not) (slap, dió una bofetada) (la, the) (bruja, witch) (verde, green)
(a la, the) (dió una bofetada a, slap the)

Slide courtesy of Kevin Knight http://www.sims.berkeley.edu/courses/is290-2/f04/lectures/mt-lecture.ppt

Word Alignment Induced Phrases



(Maria, Mary) (no, did not) (slap, dió una bofetada) (la, the) (bruja, witch) (verde, green)

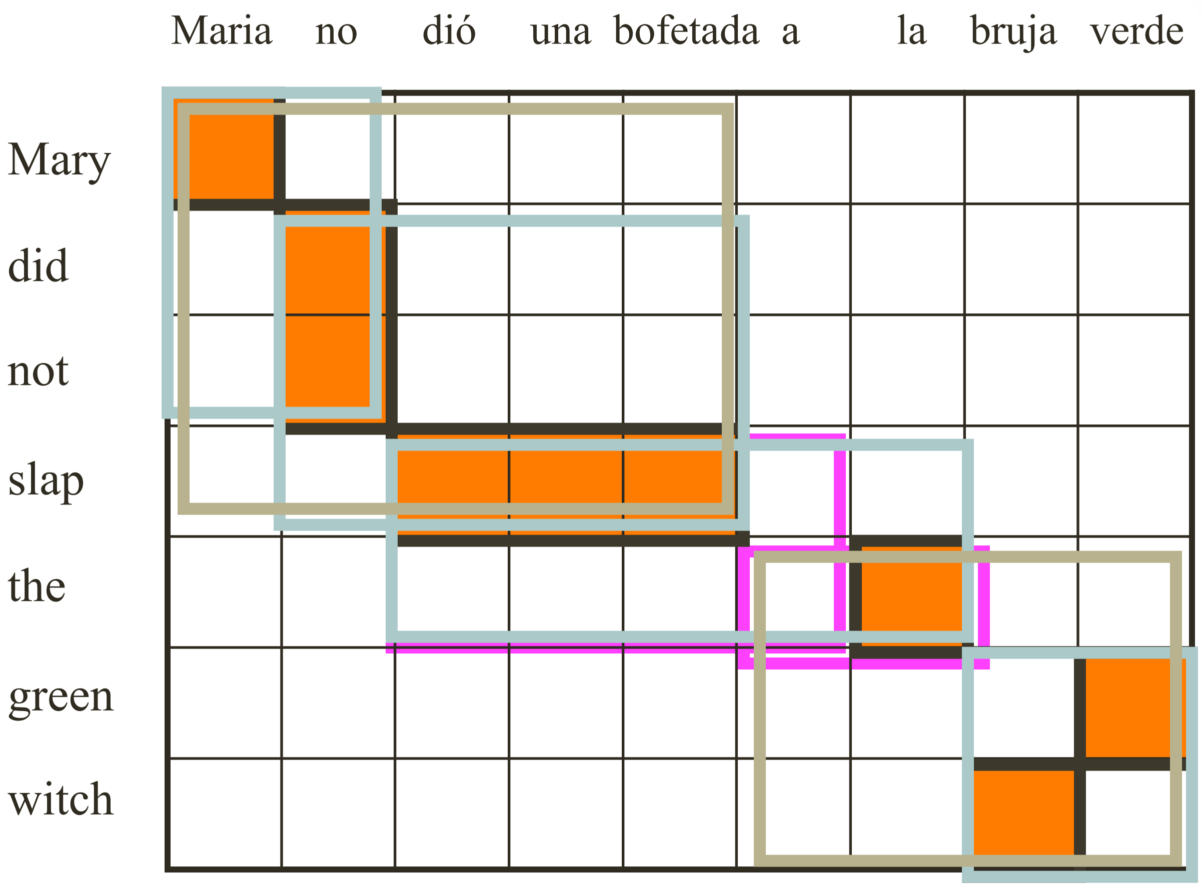
(a la, the) (dió una bofetada a, slap the)

(Maria no, Mary did not) (no dió una bofetada, did not slap), (dió una bofetada a la, slap the)

(bruja verde, green witch)

Slide courtesy of Kevin Knight http://www.sims.berkeley.edu/courses/is290-2/f04/lectures/mt-lecture.ppt

Word Alignment Induced Phrases



(Maria, Mary) (no, did not) (slap, dió una bofetada) (la, the) (bruja, witch) (verde, green)

(a la, the) (dió una bofetada a, slap the)

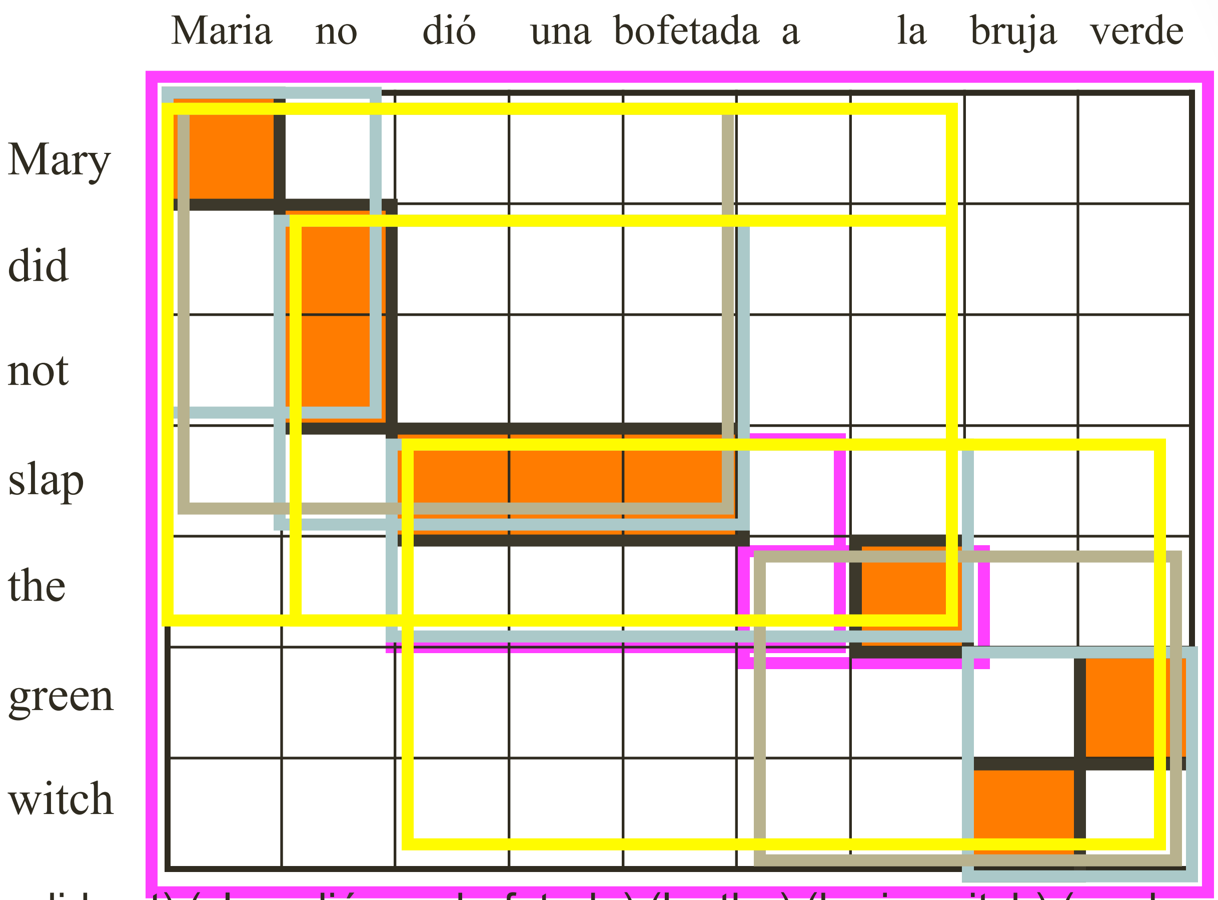
(Maria no, Mary did not) (no dió una bofetada, did not slap), (dió una bofetada a la, slap the)

(bruja verde, green witch) (Maria no dió una bofetada, Mary did not slap)

(a la bruja verde, the green witch) …

Slide courtesy of Kevin Knight http://www.sims.berkeley.edu/courses/is290-2/f04/lectures/mt-lecture.ppt

Word Alignment Induced Phrases



(Maria, Mary) (no, did not) (slap, dió una bofetada) (la, the) (bruja, witch) (verde, green)

(a la, the) (dió una bofetada a, slap the)

(Maria no, Mary did not) (no dió una bofetada, did not slap), (dió una bofetada a la, slap the)

(bruja verde, green witch) (Maria no dió una bofetada, Mary did not slap)

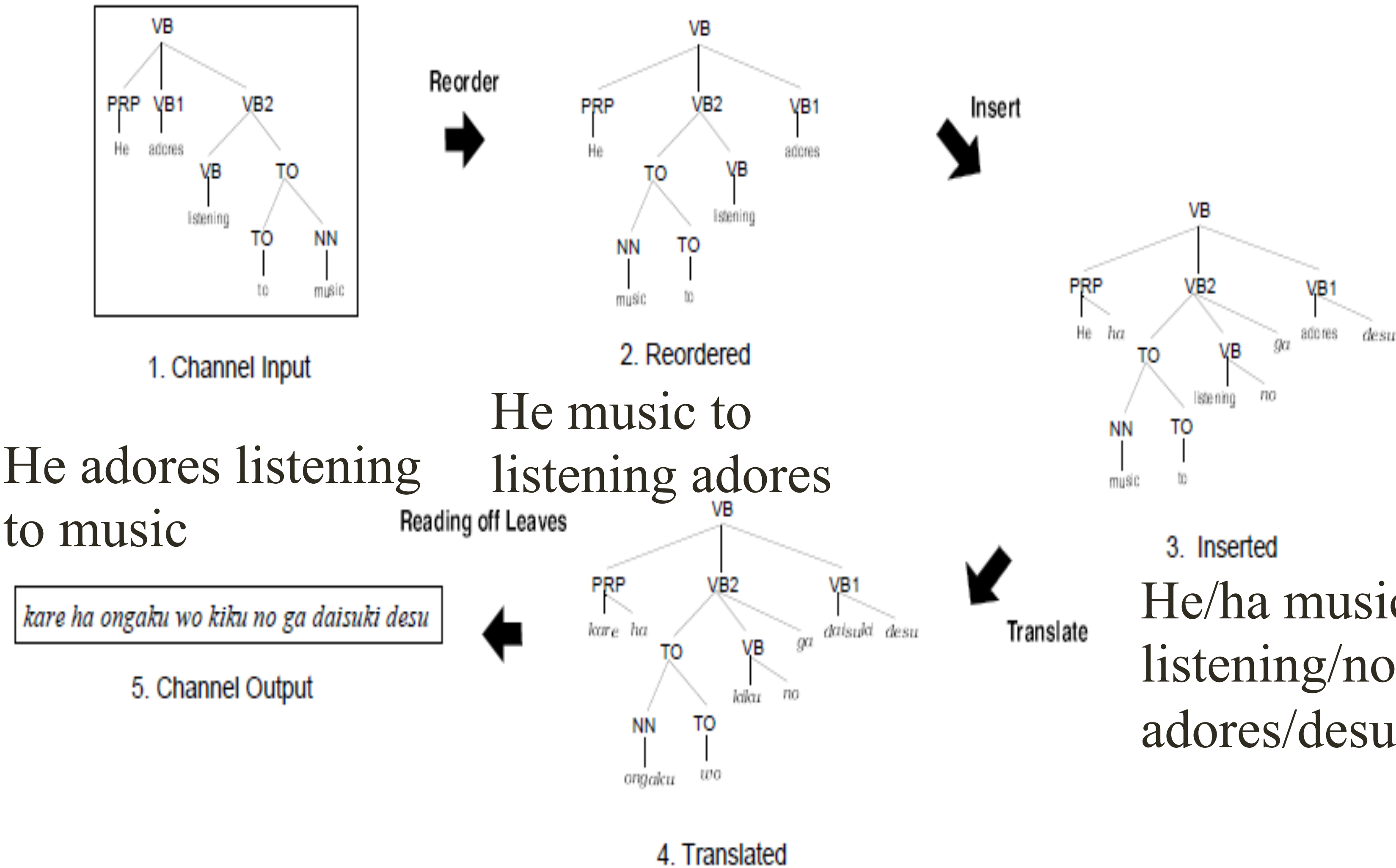
(a la bruja verde, the green witch) …

(Maria no dió una bofetada a la bruja verde, Mary did not slap the green witch)

Advantages of Phrase-Based SMT

- Many-to-many mappings can handle non-compositional phrases
- Local context is very useful for disambiguating
  - "Interest rate" → …
  - "Interest in" → …
- The more data, the longer the learned phrases
  - Sometimes whole sentences

String to Tree Translation



He adores listening to music

He music to listening adores

He/ha music to listening/no ga adores/desu

Figure 1: Channel Operations: Reorder, Insert, and Translate

(Yamada and Knight 2001)

Clause restructuring (Collins et al.)

- Ich werde Ihnen den Report aushaendigen … damit Sie den eventuell uebernehmment koennen.
- I will pass_on to_you the report, so_that you can adopt that perhaps
- verb initial: that perhaps adopt can -> adopt that perhaps can
- verb second: so that you adopt…can -> so that you can adopt
- move subject: so that can you adopt -> so that you can adopt
- particles: we accept the presidency *Particle* -> we accept the presidency

  (in German, split-prefix phrasal verbs are very common,
  e.g., “anrufen” -> “rufen sie bitte noch einmal an” – call right back please)

Slide from Radev

Synchronous Grammars

- Generate parse trees in parallel in two languages using different rules
- E.g.,
  - NP -> ADJ N (in English)
  - NP -> N ADJ (in Spanish)
- ITG (Inversion Transduction Grammar) [Wu 1995]
  - Don’t allow all permutations in derivations
  - Only <> and [ ] are allowed

Slide from Radev

MT Approaches
Practical Considerations

- Resources Availability
  - Parsers and Generators
    - Input/Output compatability
  - Translation Lexicons
    - Word-based vs. Transfer/Interlingua
  - Parallel Corpora
    - Domain of interest
    - Bigger is better
- Time Availability
  - Statistical training, resource building

Slide from Nizar Habash

Today

- Multilingual Challenges for MT
- MT Approaches
  - Statistical
  - Neural net (Thursday)
- MT Evaluation

MT Evaluation

- More art than science
- Wide range of Metrics/Techniques
  - interface, …, scalability, …, faithfulness, ... space/time complexity, … etc.
- Automatic vs. Human-based
  - *Dumb Machines vs. Slow Humans*

Slide from Nizar Habash

Human-based Evaluation Example
Accuracy Criteria

5	contents of original sentence conveyed (might need minor corrections)
4	contents of original sentence conveyed BUT errors in word order
3	contents of original sentence generally conveyed BUT errors in relationship between phrases, tense, singular/plural, etc.
2	contents of original sentence not adequately conveyed, portions of original sentence incorrectly translated, missing modifiers
1	contents of original sentence not conveyed, missing verbs, subjects, objects, phrases or clauses

Slide from Nizar Habash

Human-based Evaluation Example
Fluency Criteria

5	clear meaning, good grammar, terminology and sentence structure
4	clear meaning BUT bad grammar, bad terminology or bad sentence structure
3	meaning graspable BUT ambiguities due to bad grammar, bad terminology or bad sentence structure
2	meaning unclear BUT inferable
1	meaning absolutely unclear

Slide from Nizar Habash

Today: Crowdsourcing

- Amazon Mechanical Turk or CrowdFlower
- Create a HIT for each sentence
- Get multiple workers to rate
- Pay .01 to .10 per hit
- Complete an evaluation in hours (vs days/ weeks)
- *Ethics?*

Automatic Evaluation Example

Bleu Metric

(Papineni et al 2001)

- Bleu
  - *BiLingual Evaluation Understudy*
  - Modified n-gram precision with length penalty
  - Quick, inexpensive and language independent
  - Correlates highly with human evaluation
  - Bias against synonyms and inflectional variations

Slide from Nizar Habash

Automatic Evaluation Example
Bleu Metric

Test Sentence

colorless green ideas sleep furiously

Gold Standard References

all dull jade ideas sleep irately
drab emerald concepts sleep furiously
colorless immature thoughts nap angrily

Slide from Nizar Habash

Automatic Evaluation Example
Bleu Metric

Test Sentence

colorless green ideas sleep furiously

Gold Standard References

all dull jade ideas sleep irately
drab emerald concepts sleep furiously
colorless immature thoughts nap angrily

Unigram precision = 4/5

Slide from Nizar Habash

Automatic Evaluation Example
Bleu Metric

Test Sentence	Gold Standard References
colorless green	**all dull jade <u>ideas</u> <u>sleep</u> irately**
green ideas	**drab emerald concepts <u>sleep</u> <u>furiously</u>**
<u>ideas sleep</u>	**colorless immature thoughts nap angrily**
<u>sleep furiously</u>	

Unigram precision = 4 / 5 = 0.8

Bigram precision = 2 / 4 = 0.5

$$\text{Bleu Score} = (a_1\, a_2 \ldots a_n)^{1/n}$$
$$= (0.8 \times 0.5)^{½} = 0.6325 \rightarrow 63.25$$

Slide from Nizar Habash

BLEU scores for 110 translation systems trained on Europarl

Source Language	Target Language										
	da	de	el	en	es	fr	fi	it	nl	pt	sv
da	-	18.4	21.1	28.5	26.4	28.7	14.2	22.2	21.4	24.3	28.3
de	22.3	-	20.7	25.3	25.4	27.7	11.8	21.3	23.4	23.2	20.5
el	22.7	17.4	-	27.2	**31.2**	**32.1**	11.4	26.8	20.0	27.6	21.2
en	25.2	17.6	23.2	-	**30.1**	**31.1**	13.0	25.3	21.0	27.1	24.8
es	24.1	18.2	28.3	**30.5**	-	**40.2**	12.5	**32.3**	21.4	**35.9**	23.9
fr	23.7	18.5	26.1	**30.0**	**38.4**	-	12.6	**32.4**	21.1	**35.3**	22.6
fi	20.0	14.5	18.2	21.8	21.1	22.4	-	18.3	17.0	19.1	18.8
it	21.4	16.9	24.8	27.8	**34.0**	**36.0**	11.0	-	20.0	**31.2**	20.2
nl	20.5	18.3	17.4	23.0	22.9	24.6	10.3	20.0	-	20.7	19.0
pt	23.2	18.2	26.4	**30.1**	**37.9**	**39.0**	11.9	**32.0**	20.2	-	21.9
sv	**30.3**	18.9	22.8	**30.2**	28.6	29.7	15.3	23.9	21.9	25.9	-

Table 2: BLEU scores for the 110 translation systems trained on the Europarl corpus

Koehn, MT Summit, 2005
http://homepages.inf.ed.ac.uk/pkoehn/publications/europarl-mtsummit05.pdf

Language	From	Into	Diff
Danish (da)	23.4	23.3	0.0
German (de)	**22.2**	**17.7**	**-4.5**
Greek (el)	23.8	22.9	-0.9
English (en)	**23.8**	**27.4**	**+3.6**
Spanish (es)	26.7	29.6	+2.9
French (fr)	26.1	31.1	+5.1
Finnish (fi)	19.1	12.4	-6.7
Italian (it)	24.3	25.4	+1.1
Dutch (nl)	19.7	20.7	+1.1
Portuguese (pt)	26.1	27.0	+0.9
Swedish (sv)	24.8	22.1	-2.6

Table 3: Average translation scores for systems when translating *from* and *into* a language. Note that German (de) and English (en) are similarly difficult to translate *from*, but English is much easier to translate *into*.

Automatic Evaluation Example
METEOR
(Lavie and Agrawal 2007)

- Metric for Evaluation of Translation with Explicit word Ordering
- Extended Matching between translation and reference
  - Porter stems, wordNet synsets
- Unigram Precision, Recall, parameterized F-measure
- Reordering Penalty
- Parameters can be tuned to optimize correlation with human judgments
- Not biased against "non-statistical" MT systems

Slide from Nizar Habash

Metrics MATR Workshop

- Workshop in AMTA conference 2008
  - Association for Machine Translation in the Americas
- Evaluating evaluation metrics
- Compared 39 metrics
  - 7 baselines and 32 new metrics
  - Various measures of correlation with human judgment
  - Different conditions: text genre, source language, number of references, etc.

Slide from Nizar Habash

Automatic Evaluation Examp

SEPIA

(Habash and ElKholy 2008)

- A syntactically-aware evaluation metric
  - (Liu and Gildea, 2005; Owczarzak et al., 2007; Giménez and Màrquez, 2007)
- Uses dependency representation
  - MICA parser (Nasr & Rambow 2006)
  - 77% of all structural bigrams are surface n-grams of size 2,3,4
- Includes dependency surface span as a factor in score
  - long-distance dependencies should receive a greater weight than short distance dependencies
    - Higher degree of grammaticality?