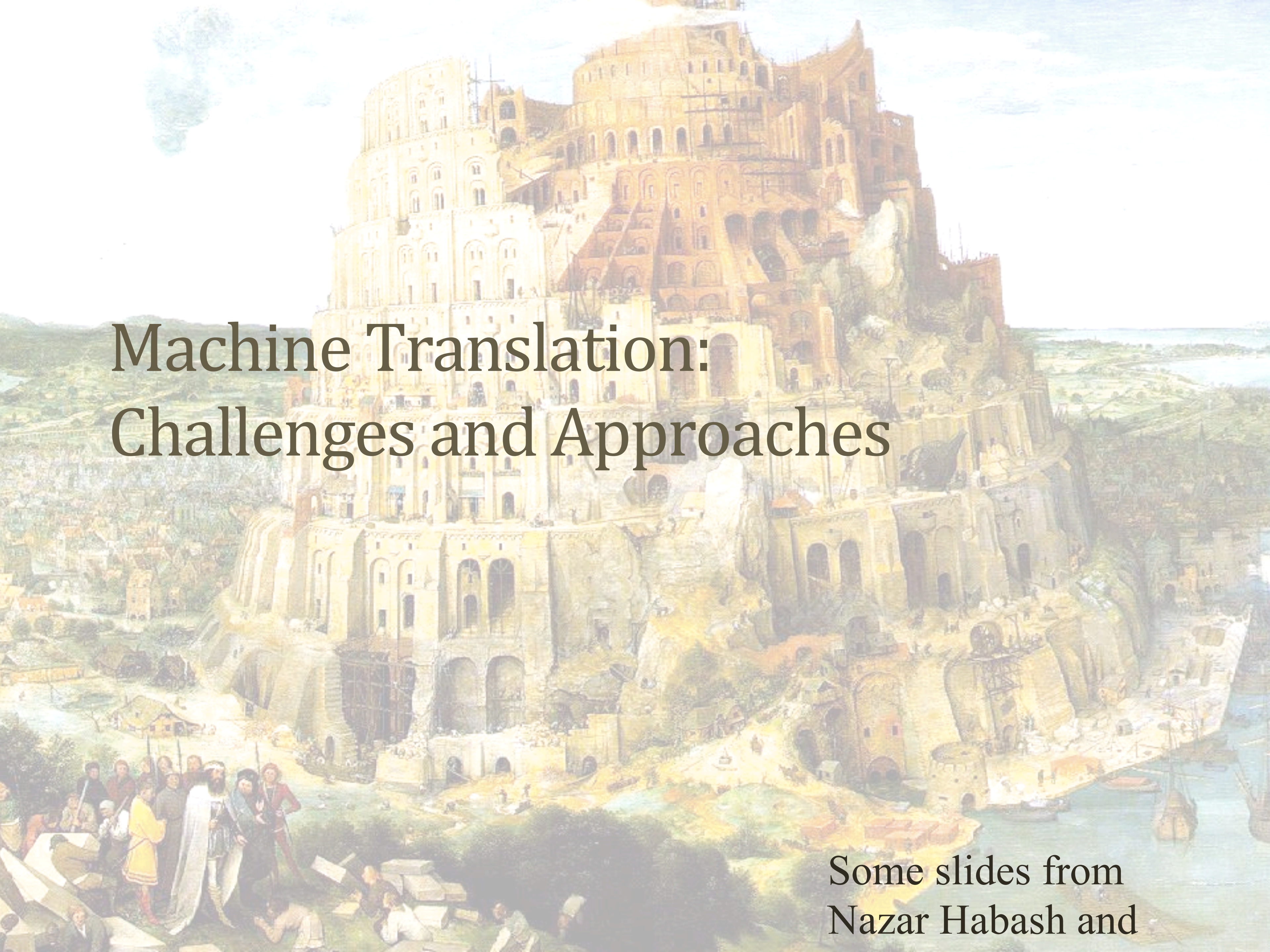


Machine Translation: Challenges and Approaches

Some slides from
Nazar Habash and

Announcements

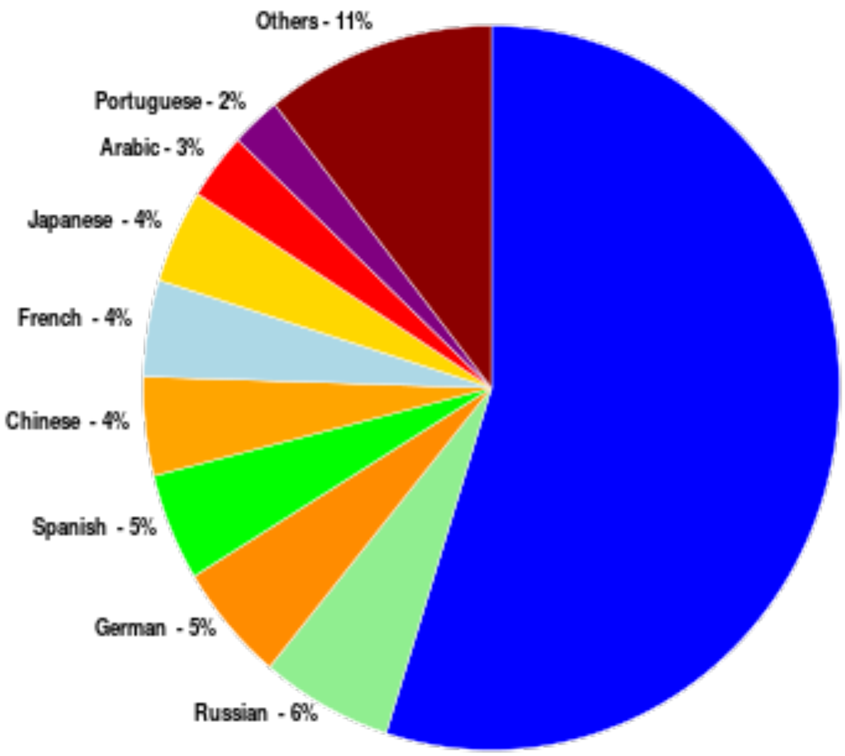
- Explanation of midterm grades at end of class (remind me!)
- Reading
  - Today: C 18.1-18.2 NLP
  - Next week: C 18.3, 18.4, NLP
- HW 2 will be returned next week
- My office hours today: 4:30-5:30

Semantic interpretation

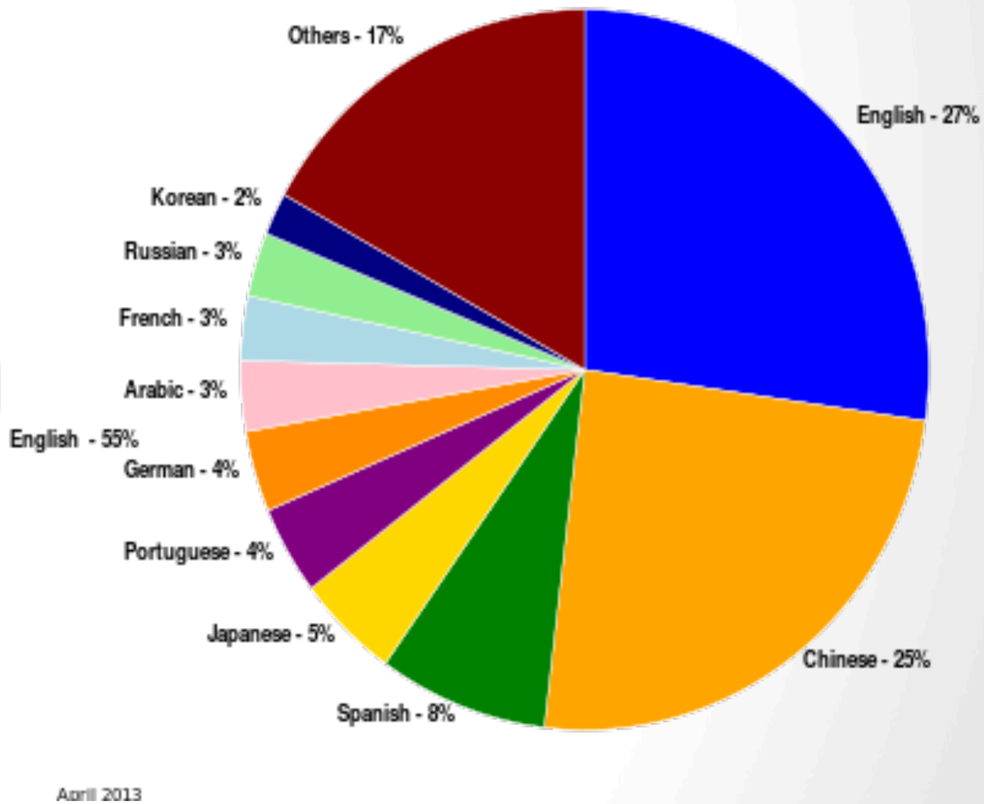
- Semantic role labeling, framenet parsers, AMT parsers
  - Take syntactic tree as input, produce a semantic representation as output
- Information extraction
  - Produce relations, events, entities
- Parsing directly into programming language (language as action)
  - Percy Liang: using language to represent if-then recipes (E.g., controlling smart phones)
  - Large online repository of english/code

Multilingual Users

- Content languages for websites



Percentage of Internet users by language



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

Google Translate

Yiddish
Yoruba
Zulu

Afrikaans	Bulgarian	Greek	German	Igno	Kurdish	Malayalam	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 (Nov 6th)
- MT Evaluation

Today

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

Multilingual Challenges

- Orthographic Variations
  - Ambiguous spelling
    - كتب الاولاد اشعارا كَتَبَ الأوْلادُ 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)

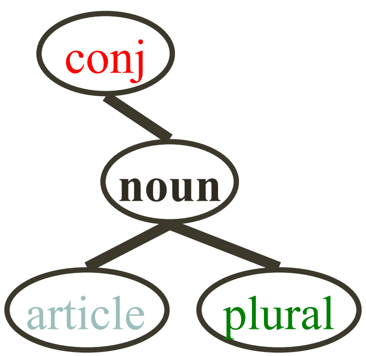
Multilingual Challenges

Morphological Variations

- Affixation vs. Root+Pattern

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

- Tokenization

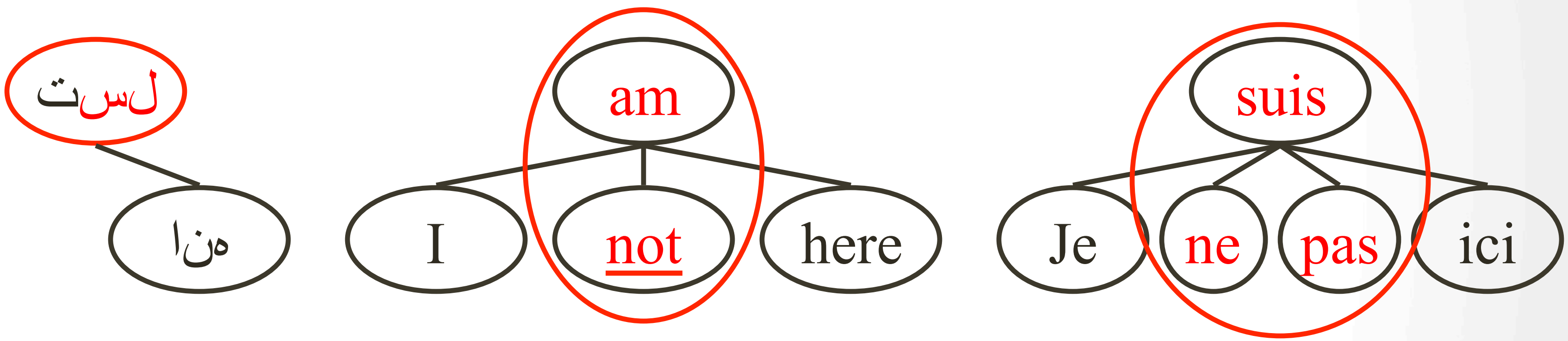


And the cars	→	*and the cars*
والسيارات	→	w Al SyArAt
Et les voitures	→	et le voitures

Slide from Nizar Habash

Translation Divergences

conflation



لست هنا
I-am-not here

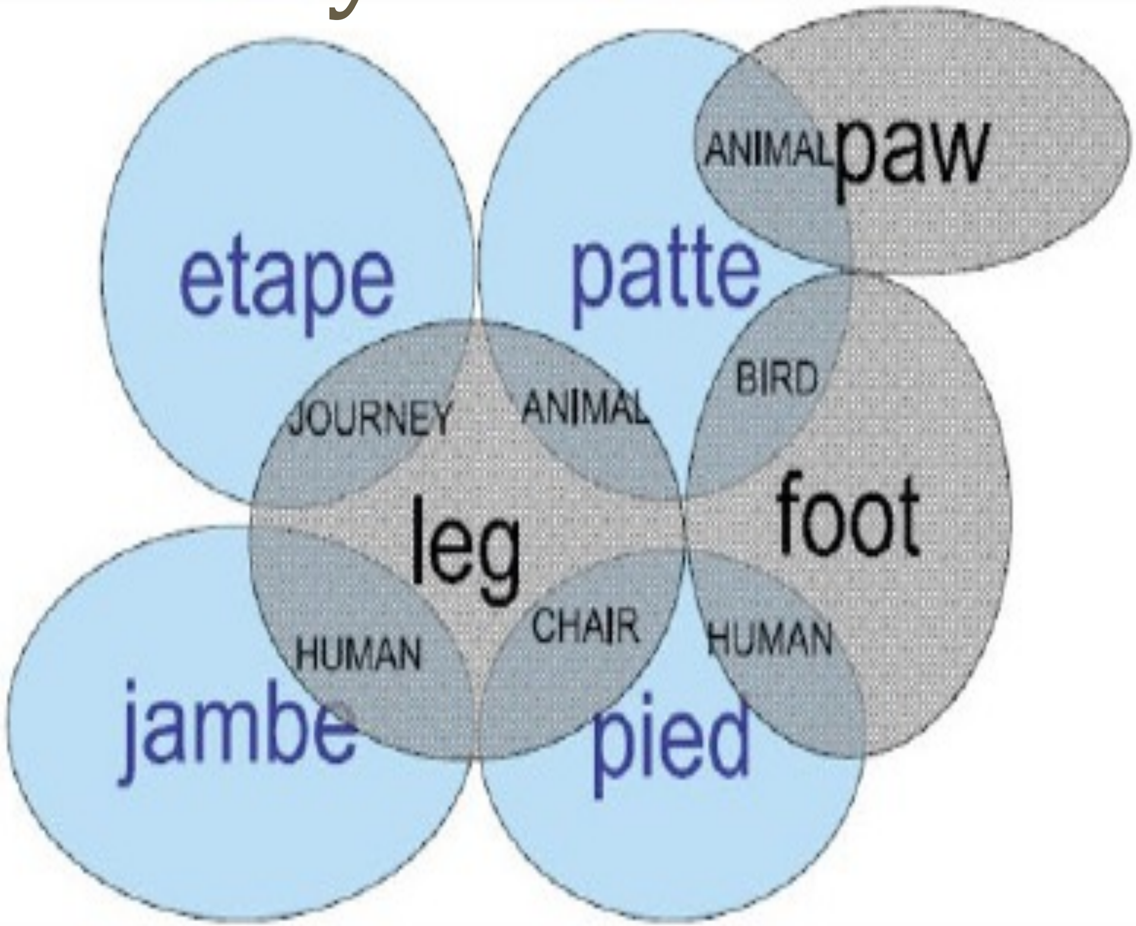
I am not here

Je ne suis pas ici
I not am not here

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*

Language Differences - vocabulary



[Example from Jurafsky and Martin]

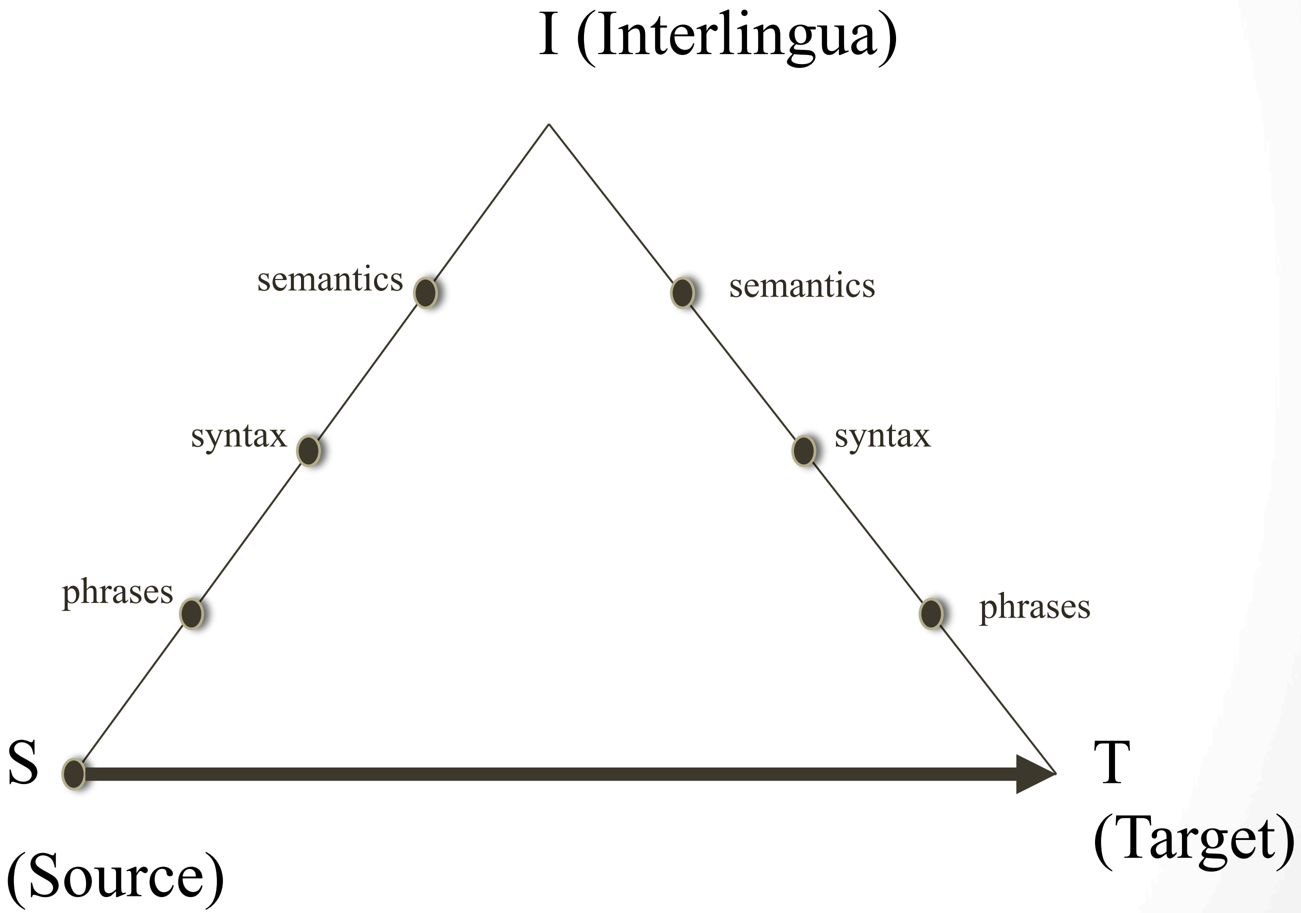
Language Differences - Syntax

- 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

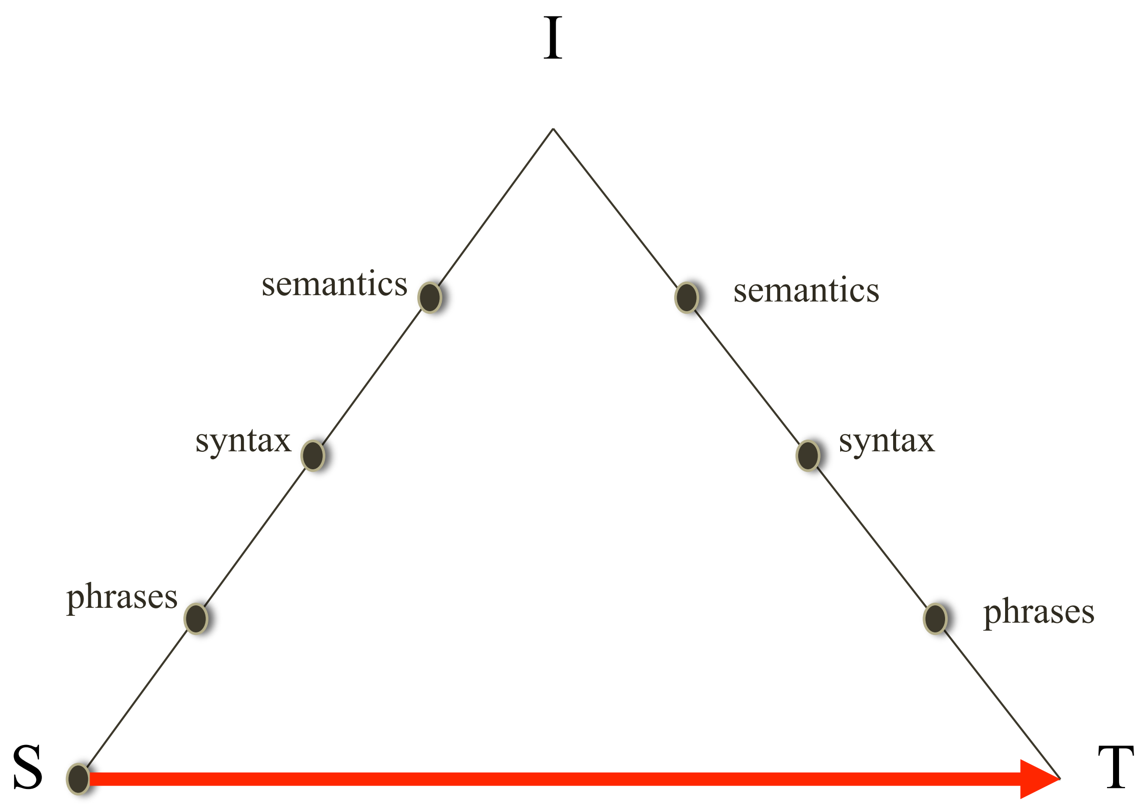
Today

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

MT Approaches
MT Pyramid

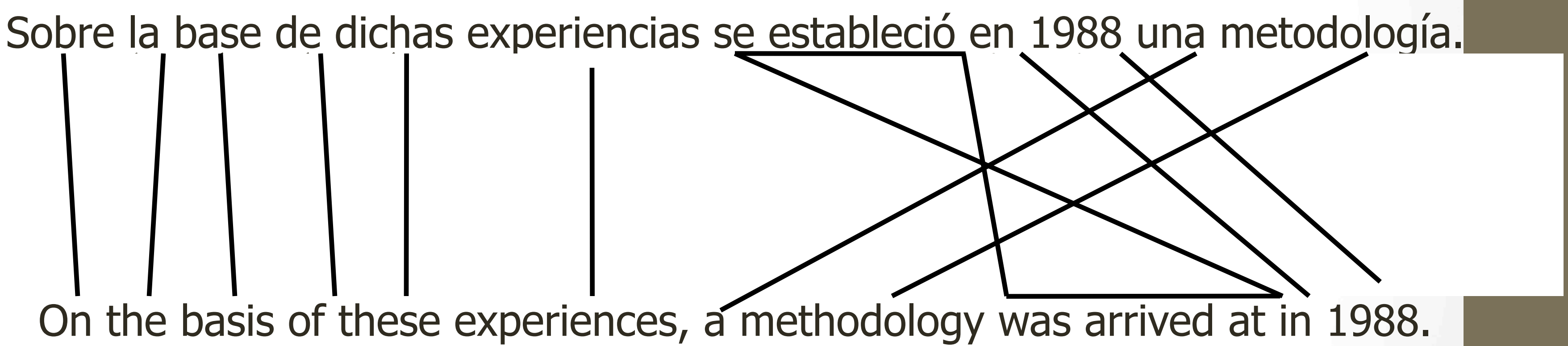


String-to-String Translation

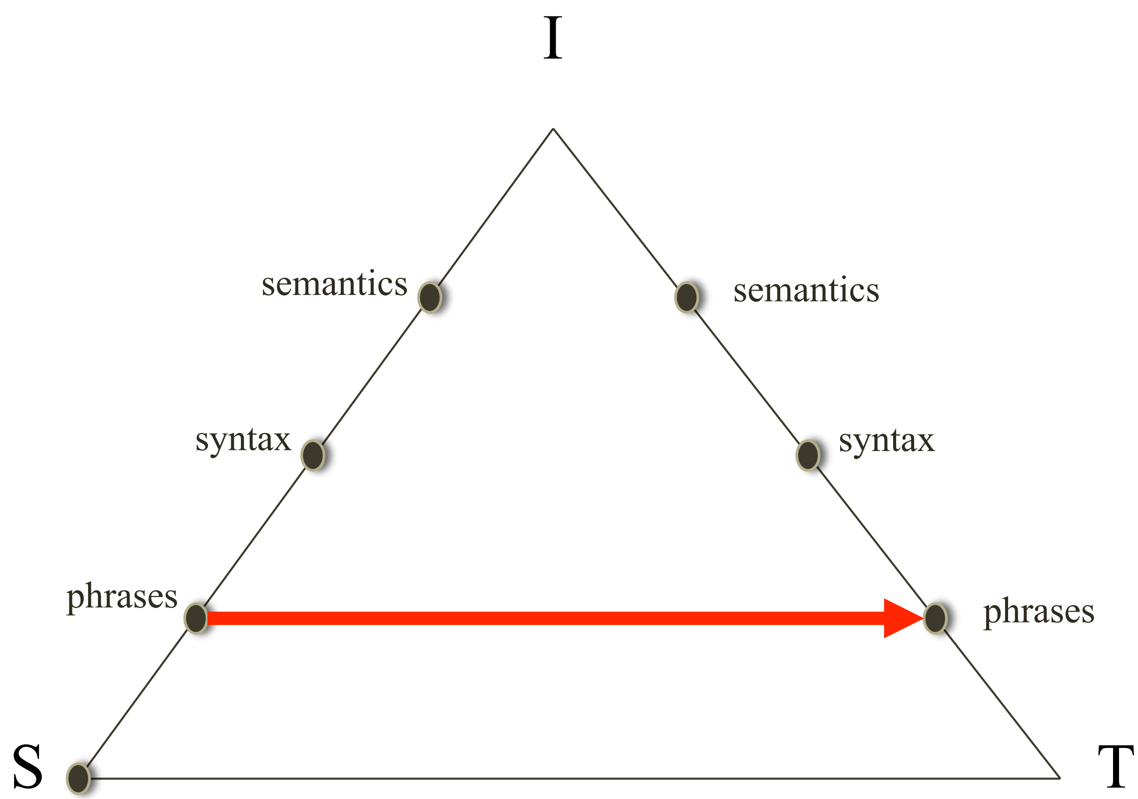


MT Approaches

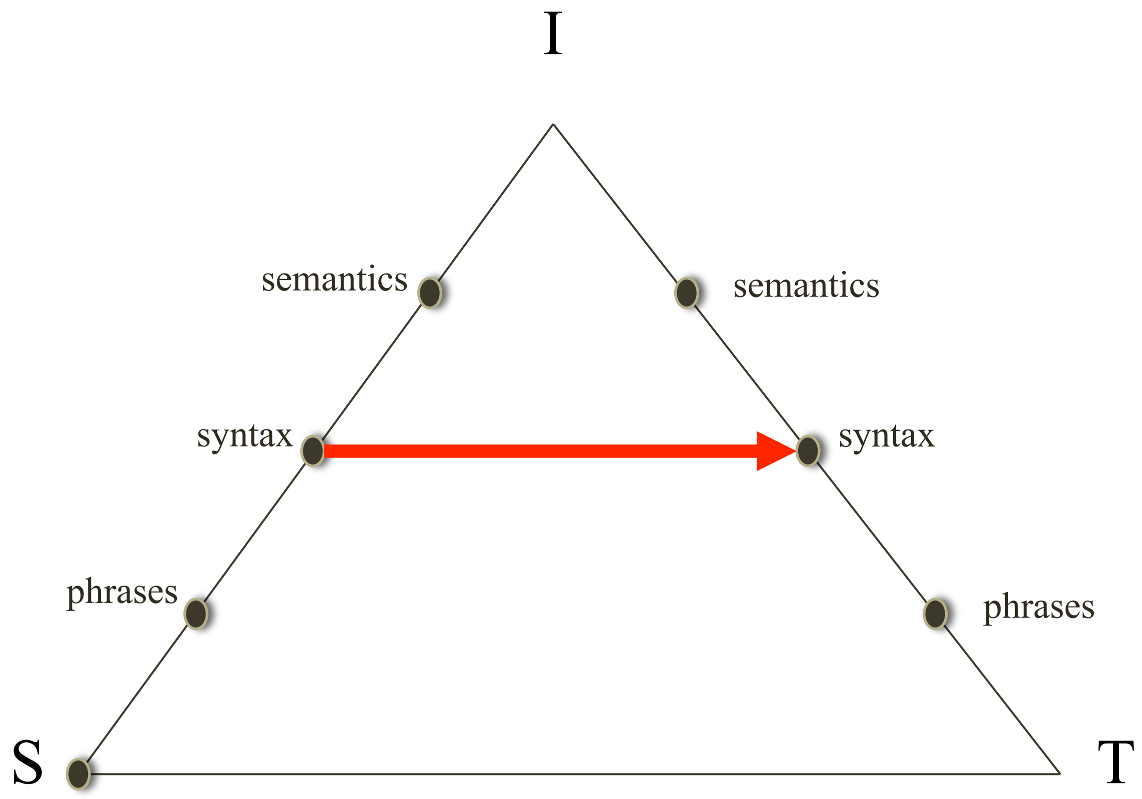
Gisting Example



Phrase-Based Translation

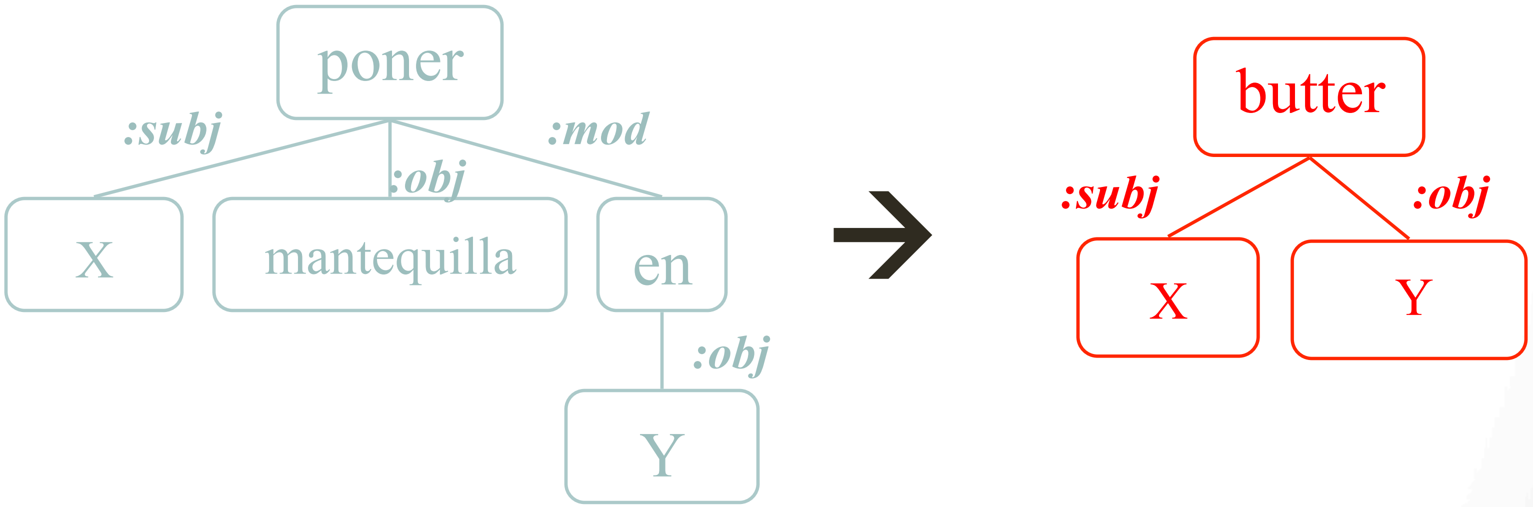


Tree-to-Tree Translation



MT Approaches
Transfer Example

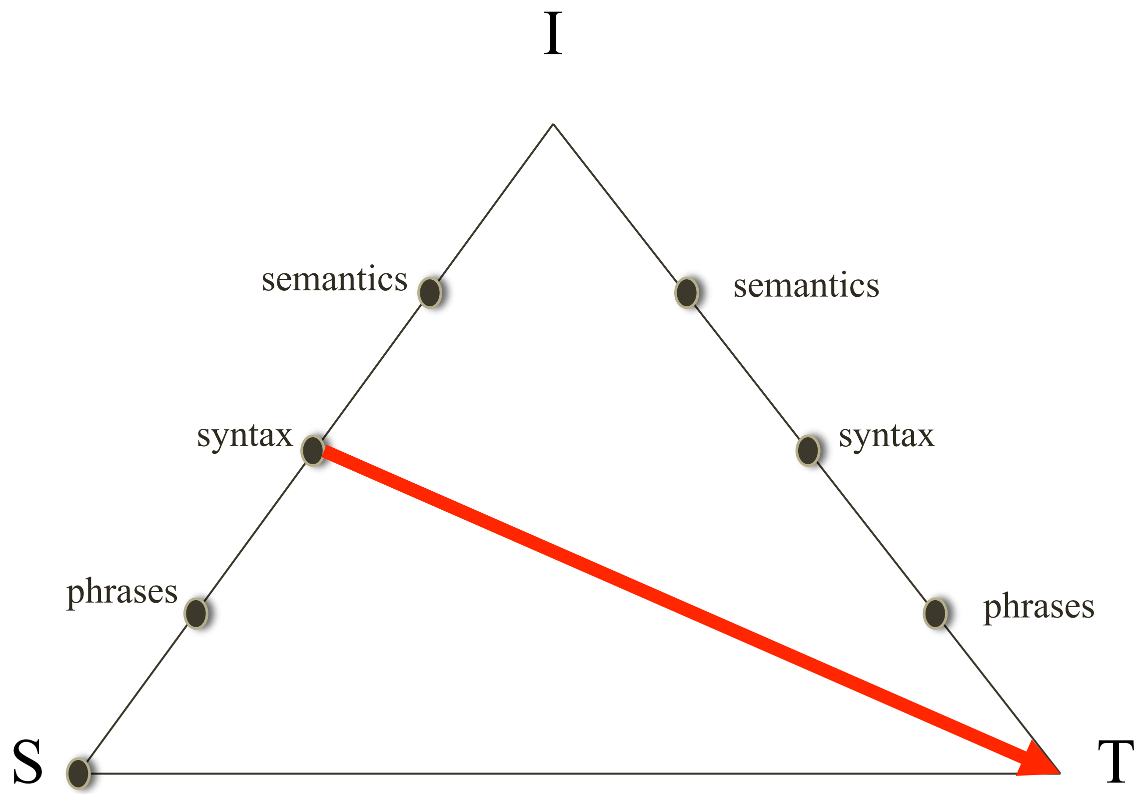
- Transfer Lexicon
  - Map SL structure to TL structure



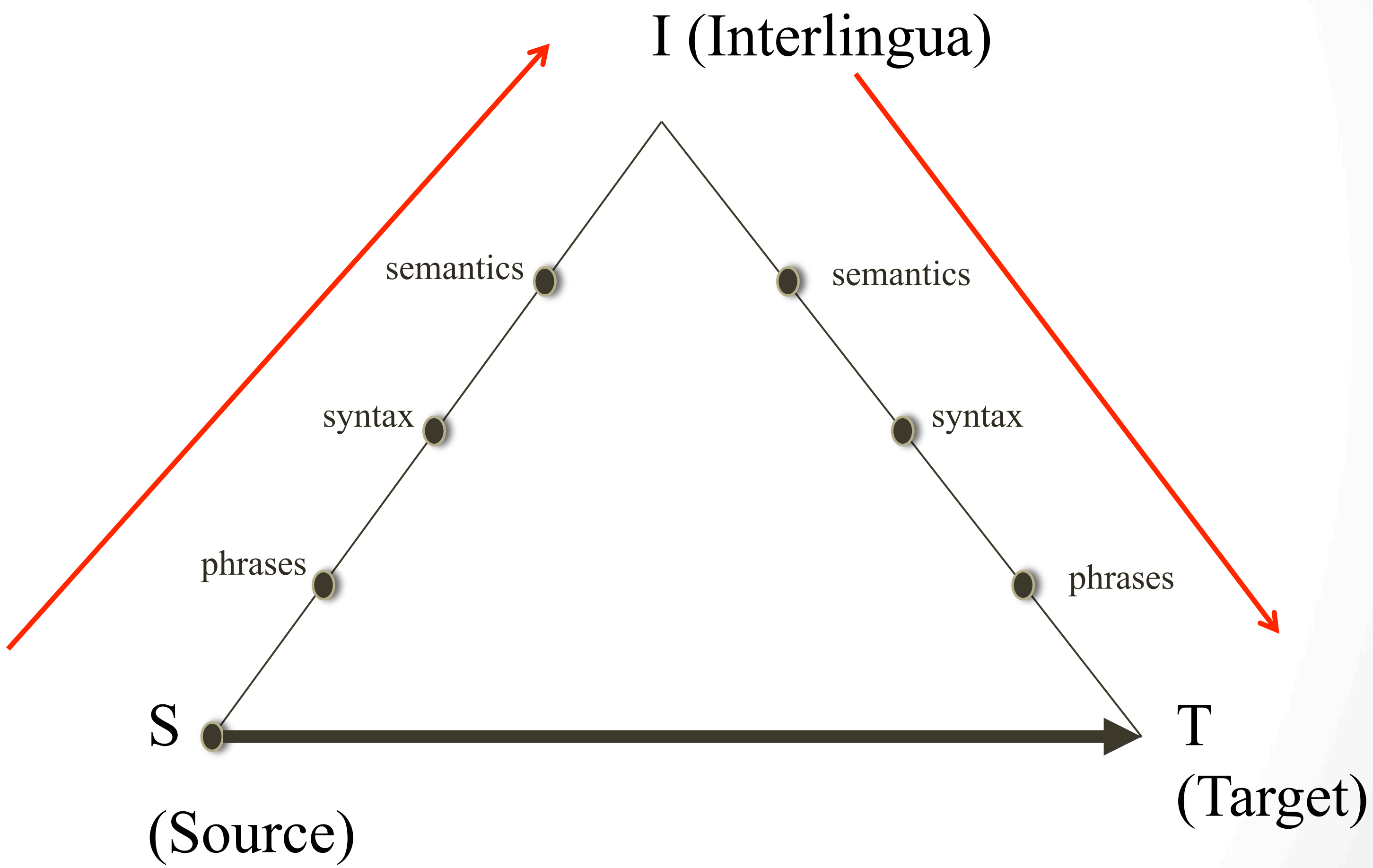
X puso mantequilla en Y

X buttered Y

Tree-to-String Translation



MT Approaches
MT Pyramid



AMR characteristics

- Rooted, labeled graphs
- Abstract away from syntactic differences
  - He described her as a genius
  - His description of her: genius
  - She was a genius according to his description
- Use Propbank framesets
  - “bond investor”: invest-01
- Heavily biased towards English

- Variables (or nodes) for entities, events, properties, states
- Leaf nodes are labeled with concepts:
  - (b/boy) an instance of the concept boy
- Relations link entities
  - (d/die-01 :location(p/park)): there was a death in the park
- AMR concepts
  - English words (e.g., boy), Propbank framesets (e.g., want-01) or special keywords (entity-types, quantities or conjunctions)

AMR relations

- ~100 relations
- Frame arguments
  - Arg0, arg1, arg2, arg3, arg4, arg5 (Propbank)
- General semantic relations
  - :Accompanier, :age, :beneficiary, :cause, :compared-to, :concession, :condition, :consistof, :degree, :destination, :direction, :domain, :duration, :employed-by, :example, :extent, :frequency, :instrument, :li, :location, :manner, :medium, :mod, :mode, :name, :part, :path, :polarity, :poss, :purpose, :source, :subevent, :subset, :time, :topic, :value.
- Relations for quantity
  - :quant, :unit, :scale
- Relations for date entity
  - :day, :month, :year, :weekday, :time, :timezone, :quarter,:dayperiod, :season, :year2, :decade, :century, :calendar, :era.
- Relations for lists
  - :op1, :op2, …. :op10
- Plus inverses (e.g., :arg0-of, :location-of)

General semantic relations

- Non-core relations
- (s :hum-02
  - :arg0 (s2 / soldier)
  - :beneficiary (g / girl)
  - :time (w / walk-01
      - :arg0 g
      - :destination (t/ town)))
- The soldier hummed to the girl as she walked to town.

Inverse relations

- In order to obtain rooted structures

- (s / sing-01
  - :arg0 (b /boy
    :source (c / college))

- The boy from the college sang.

- (b / boy
  - :arg0-of (s / sing-01)
    :source (c / college))

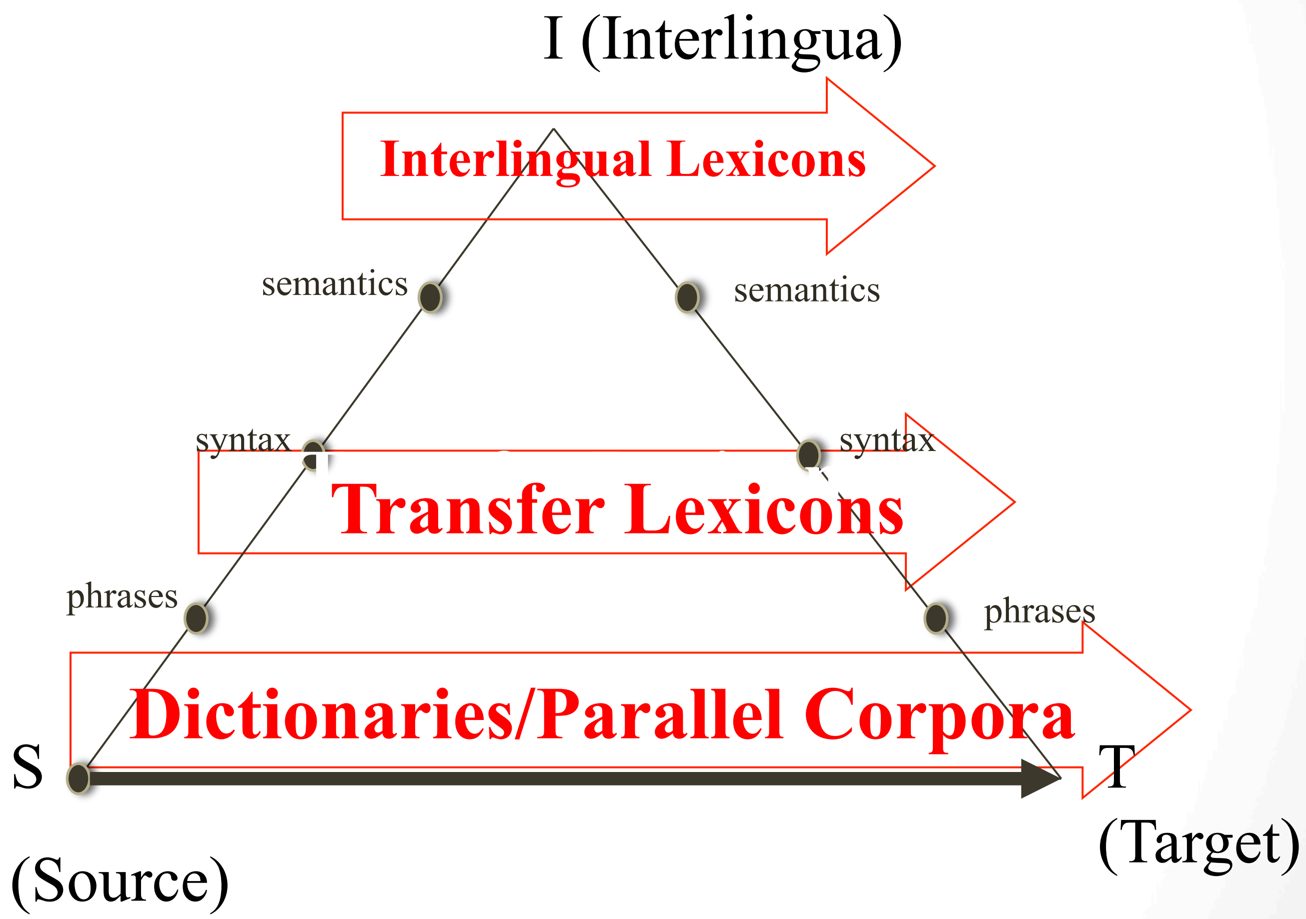
the college boy who sang

Modals and Negation

- Negation is represented with :polarity and modality is represented with concepts

- (g / go-01
        :arg0 (b / boy)
        :polarity -)
  The boy did not go.

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)”

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

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?

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)

a flower red

red flower a

flower red a

a red dog

Dog cat mouse

a red flower

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

Which phrases have high probability under p(f|e)

A flower red

red flower a

flower red a

a red dog

dog cat mouse

a red flower

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

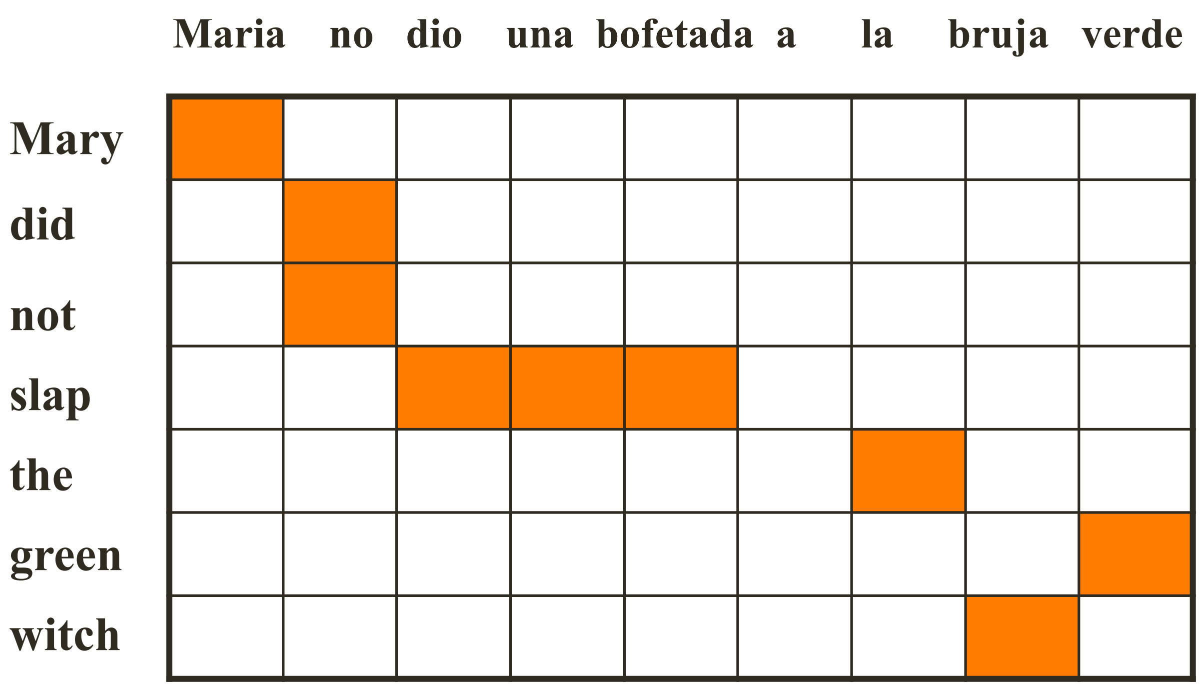
Slide from Radev

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

Statistical MT

Automatic Word Alignment

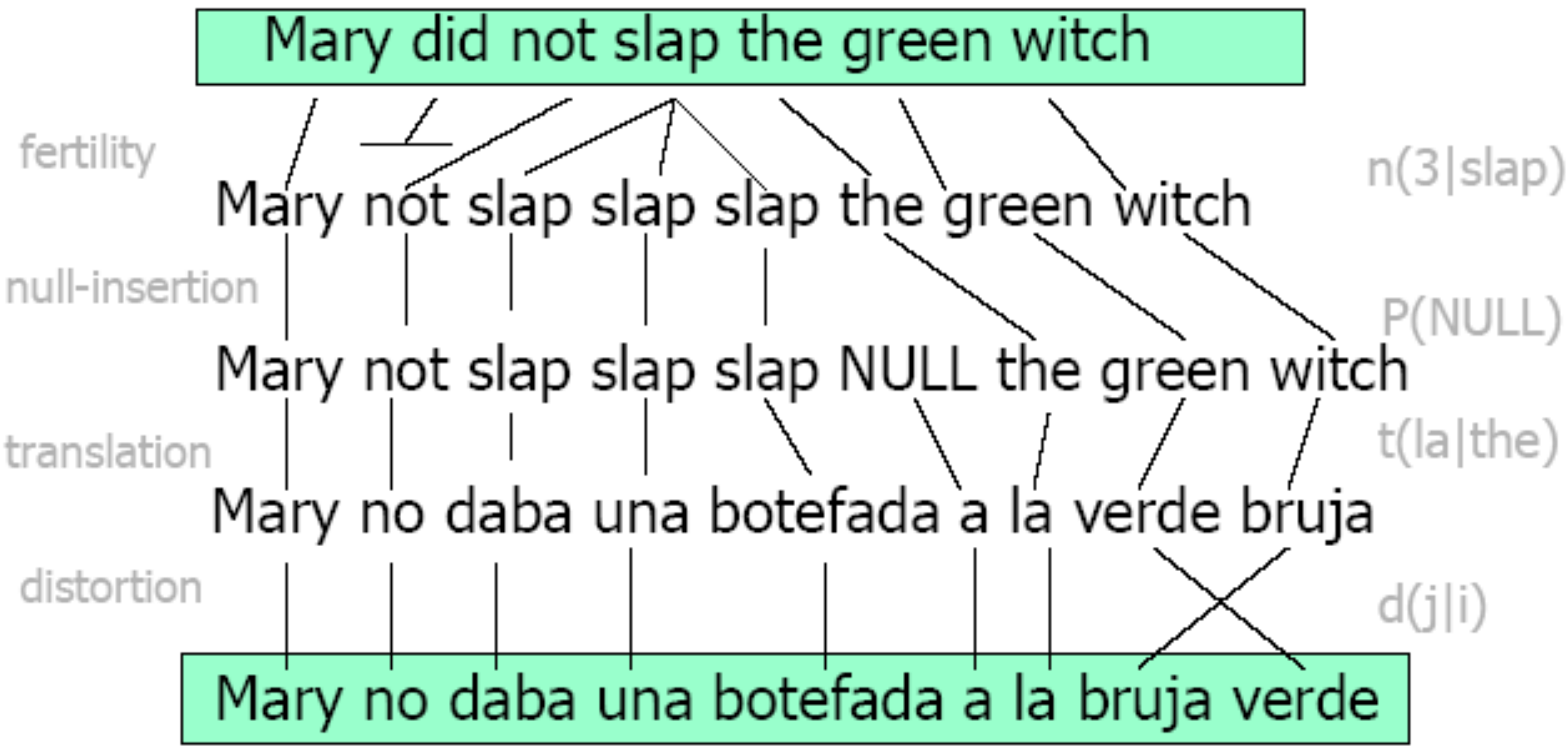
- GIZA++
  - A statistical machine translation toolkit used to train word alignments.
  - Uses Expectation-Maximization with various constraints to bootstrap alignments



What constraints might be used to bootstrap alignment?

Statistical MT

IBM Model (Word-based Model)



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

IBM's EM trained models (1-5)

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

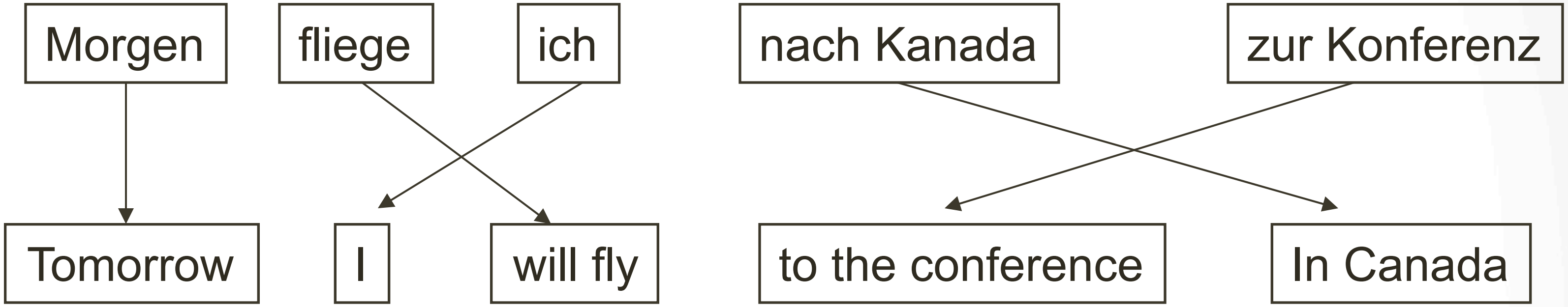
All are separate models to train!

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



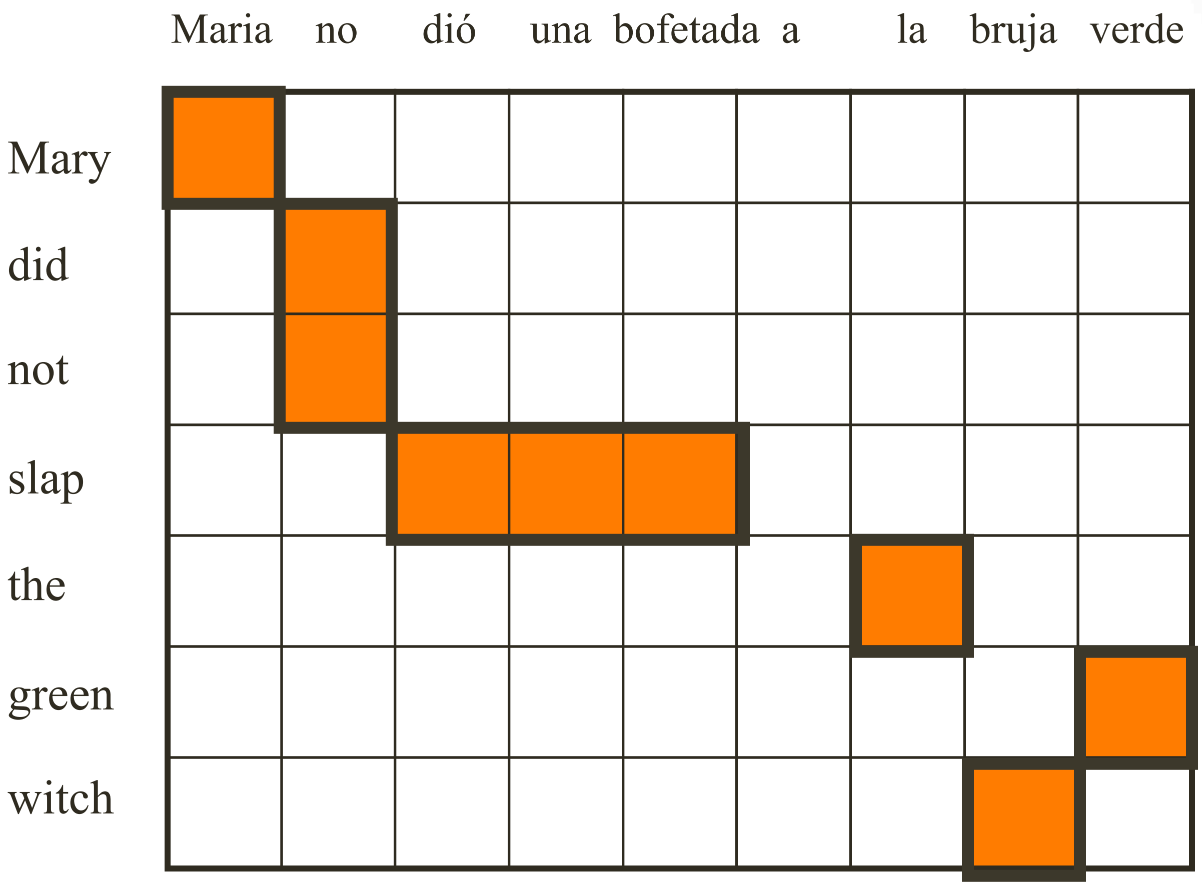
- Foreign input segmented in to phrases
  - "phrase" is any sequence of words
- Each phrase is probabilistically translated into English
  - P(to the conference | zur Konferenz)
  - P(into the meeting | zur Konferenz)
- Phrases are probabilistically re-ordered

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

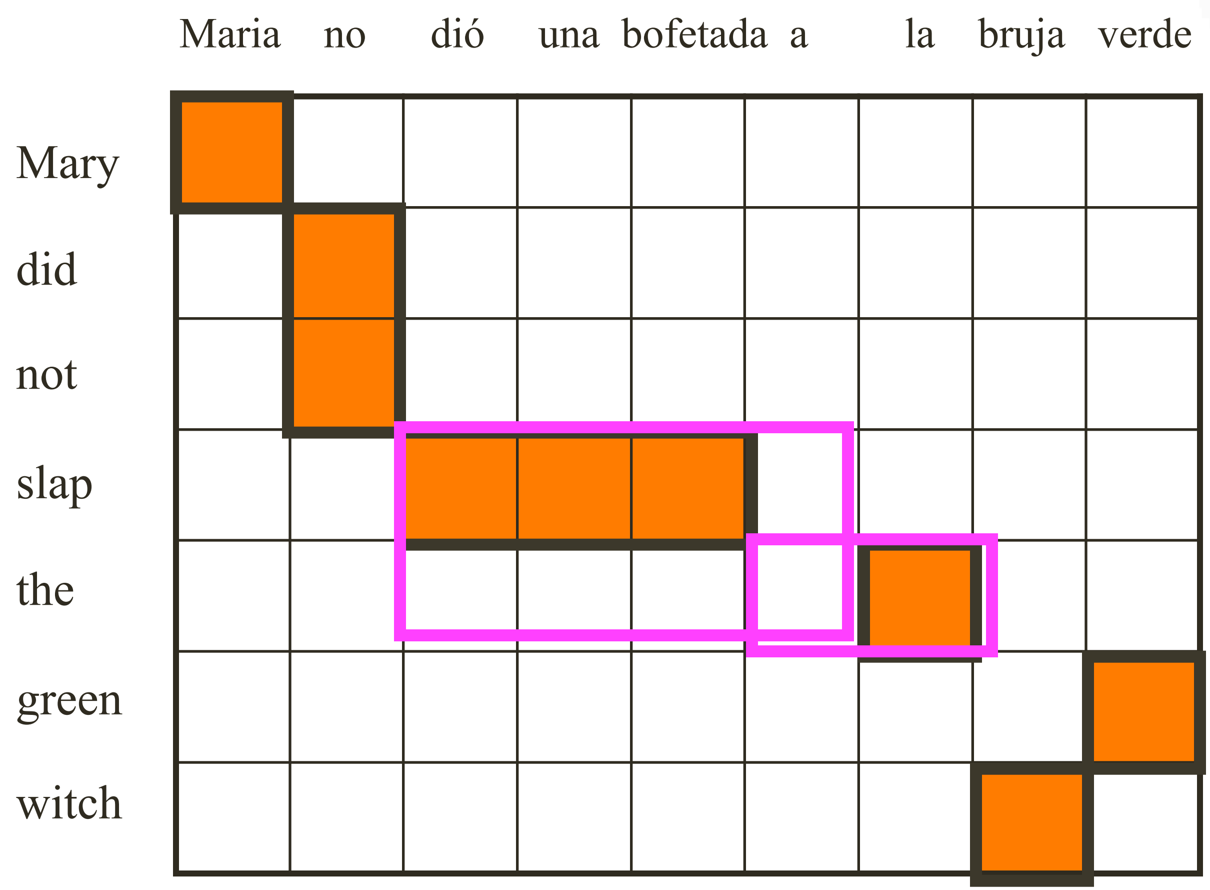
Word Alignment Induced Phrases



(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

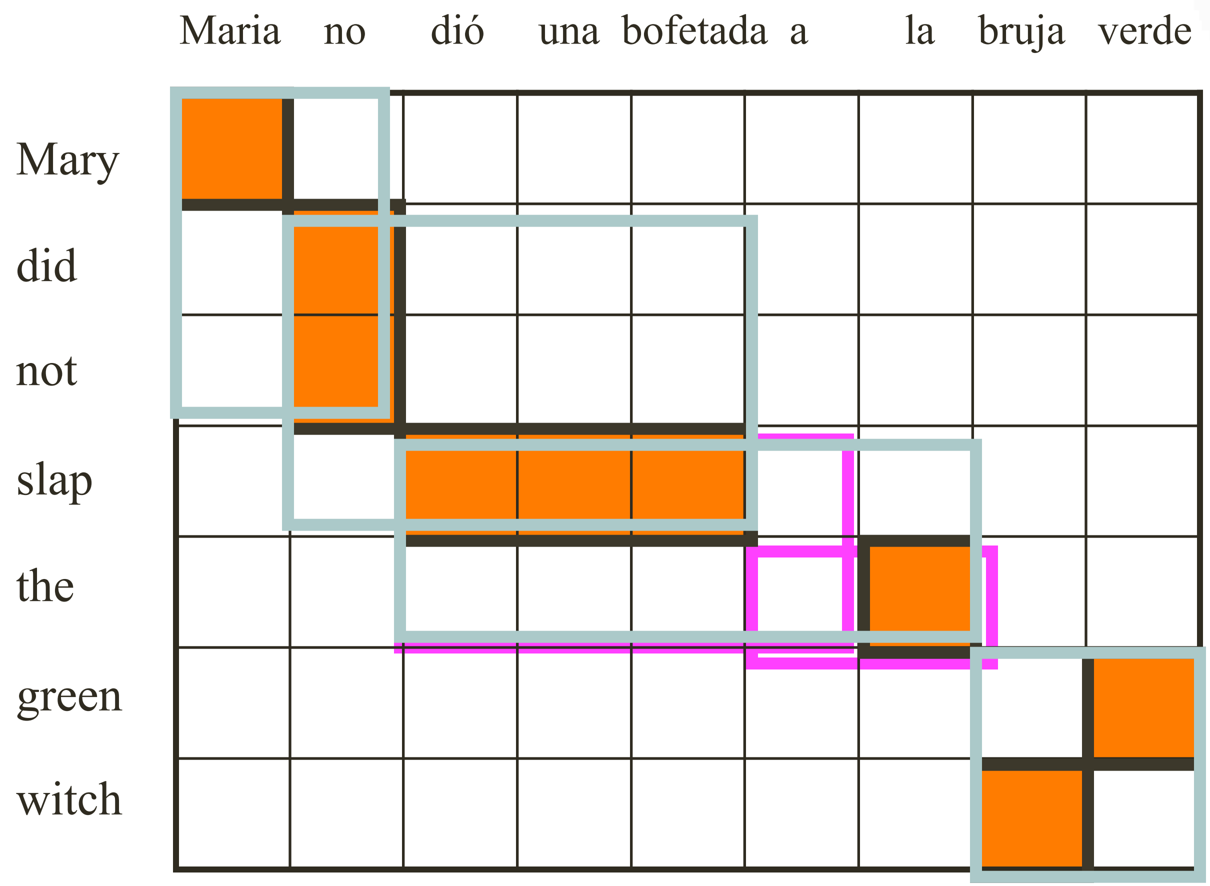


(Maria, Mary) (no, did not) (slap, dió una bofetada) (la, the) (bruja, witch) (verde, green)

(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



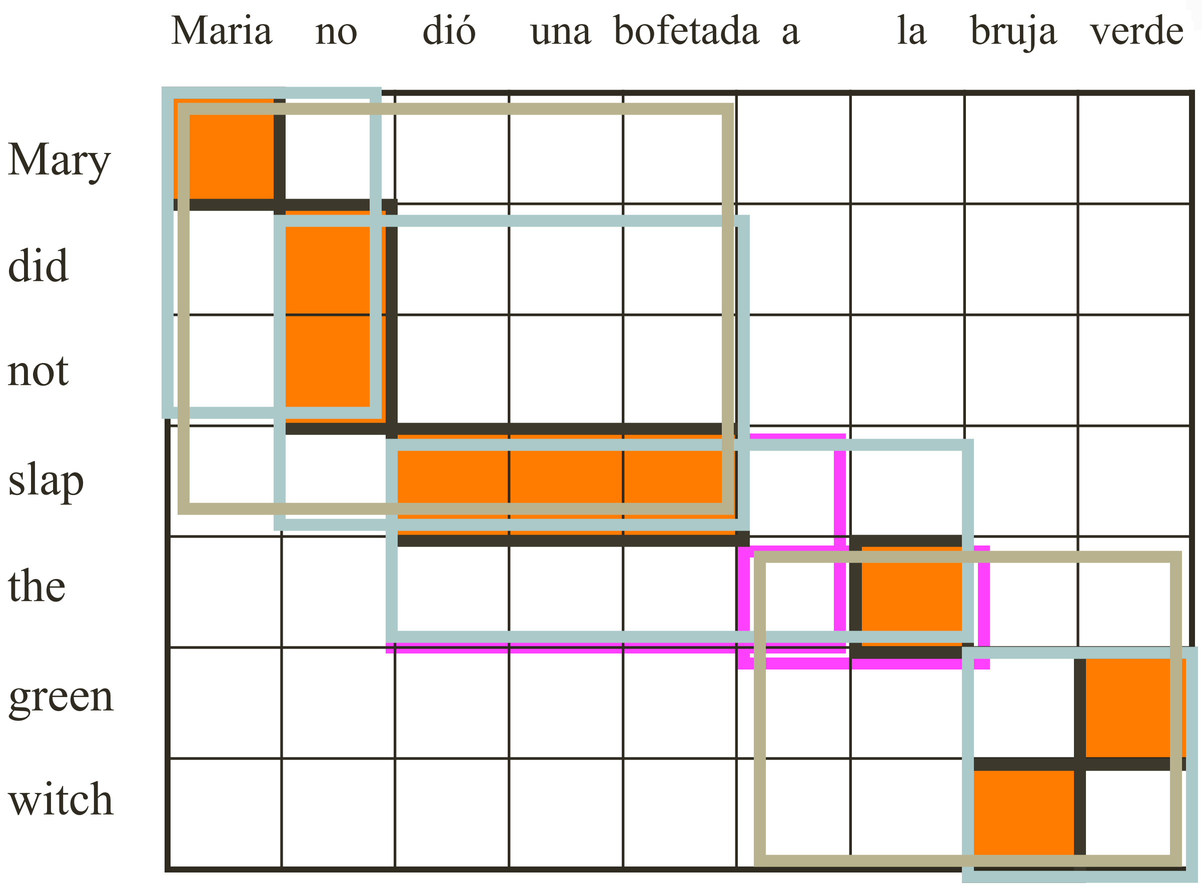
(Maria, Mary) (no, did not) (slap, dió una bofetada) (la, the) (bruja, witch) (verde, green)

(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)

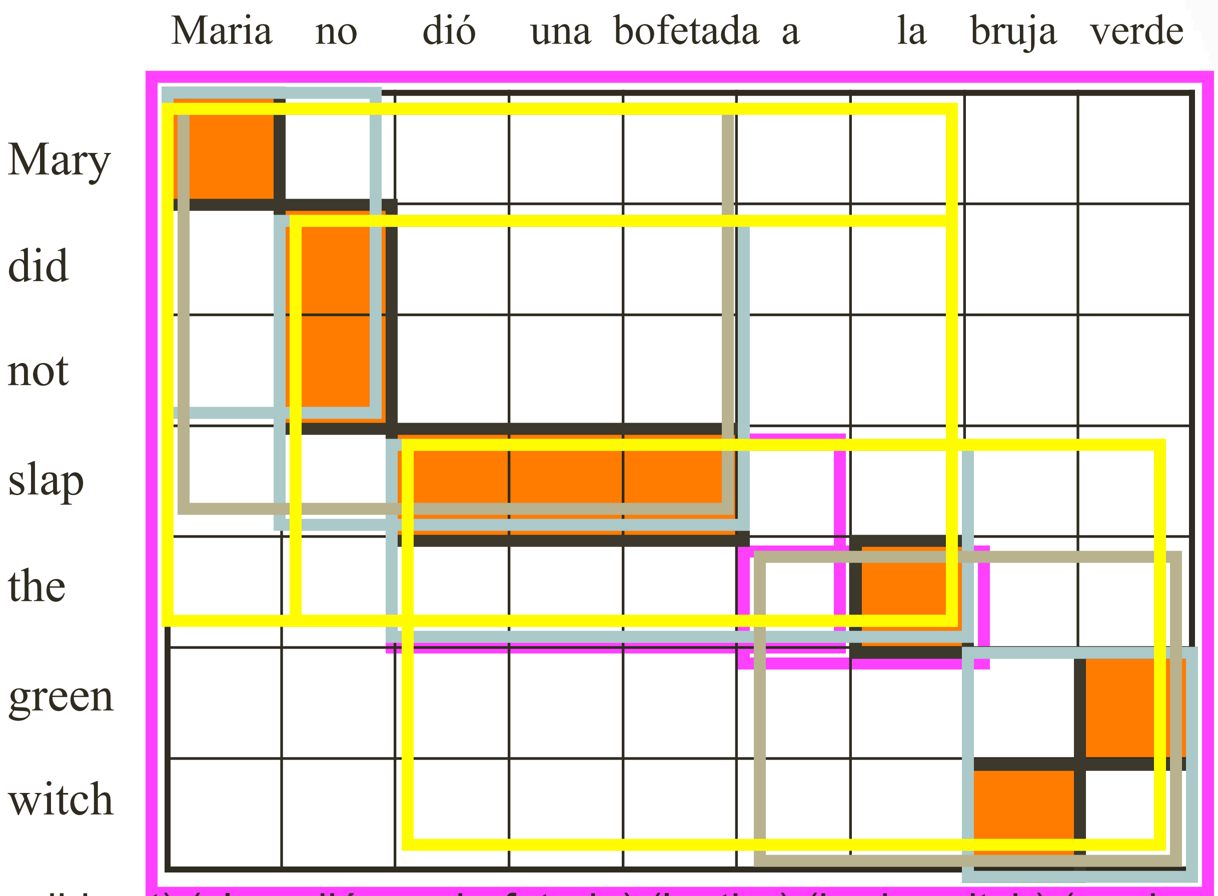
(Maria no, Mary did not) (no dió una bofetada, did not slap), (dió una bofetada a la, slap the)

(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)

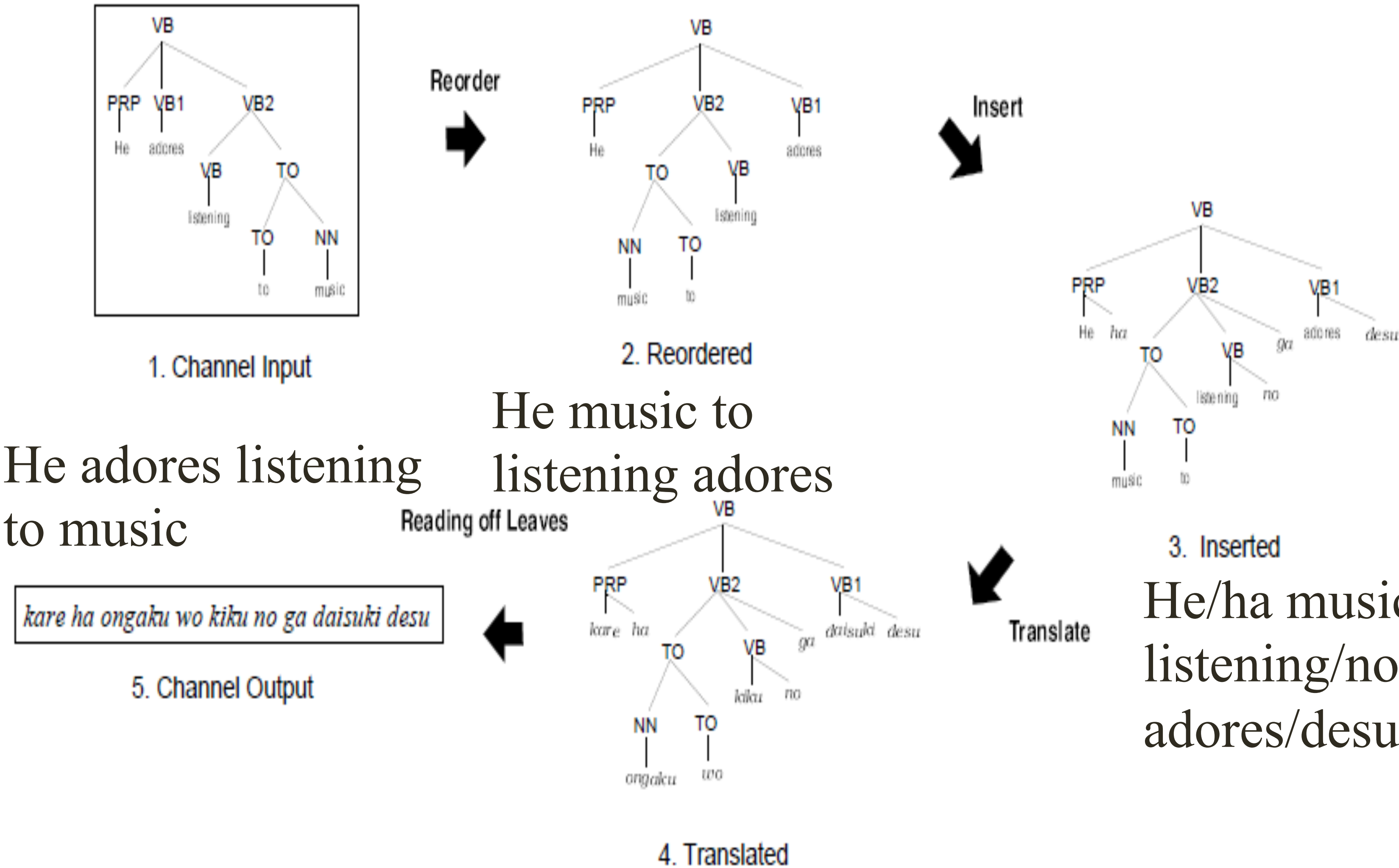
(a la bruja verde, the green witch) …

(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 200

Clause restructuring (Collins et al.)

- Ich werde Ihnen den Report aushaendigen … damit Sie den eventuell uebernehment koennen.
- I will pass_on to_you the report, so_that you can adopt that perhaps
- Google translate: I will give you the report ... so that you can take over the eventuality.
- 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

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

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

MT Approaches
Practical Considerations

- Resource 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

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*

Human-based Evaluation Example
Adequacy 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

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

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

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

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

colorless green
green ideas
ideas sleep
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 = 0.8
Bigram precision = 2 / 4 = 0.5

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

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*.

Bleu scores 2019 (teams in WMT)

Portuguese - > Spanish

BLEU	TER
66.6	19.7
59.9	25.3
59.1	25.5
58.6	25.1
58.4	25.3
56.9	26.9
54.9	28.4
54.8	29.8
52.3	32.9
52.2	32.8
51.9	30.5
49.7	32.1

Spanish-> Portuguese

BLEU	TER
64.7	20.8
62.1	23.0
53.3	29.1
52.8	28.6
52.0	29.4
51.0	33.1
47.9	33.4
46.1	36.0
46.1	35.9
45.5	35.3
44.0	37.5

Bleu scores 2019 (teams in WMT)

Hindi -> Nepali

BLEU	TER
53.7	36.3
11.5	79.1
11.1	79.7
08.2	77.1
08.2	77.2
03.7	-
03.6	-
03.5	-
03.1	-
02.8	-
02.7	-
01.6	-
01.4	-

Nepali -> Hindi

BLEU	TER
49.1	43.0
24.6	69.1
12.1	76.2
09.8	91.3
09.1	88.3
09.1	88.4
04.2	-
03.6	-
02.7	-
01.4	-
0	-
0	-

Table 17: Europarl v9 Parallel Corpus

	Czech ↔ Polish		Spanish ↔ Portuguese	
sentences	631372		1811977	
words	12526659	12641841	47832025	46191472

Table 18: Wiki Titles v1 Parallel Corpus

	Czech ↔ Polish		Spanish ↔ Portuguese	
sentences	248645		621296	
words	551084	554335	1564668	1533764

Table 19: JRC-Acquis Parallel Corpus

	Czech ↔ Polish		Spanish ↔ Portuguese	
sentences	1311362		1650126	
words	21409363	21880482	35868080	33474269

Table 20: News Commentary v14 Parallel Corpus

	Spanish ↔ Portuguese	
sentences	48168	
words	1271324	1219031

Table 21: GNOME, Ubuntu, KDE Parallel Corpus

	Hindi ↔ Nepali	
sentences	65505	
words	253216	222823

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

Midterm grades

- Available after class
- Mean: 69
- Median: 70.25
- Max: 95.5
- Min: 32.5
- STDEV: 14.04
- Will be curved