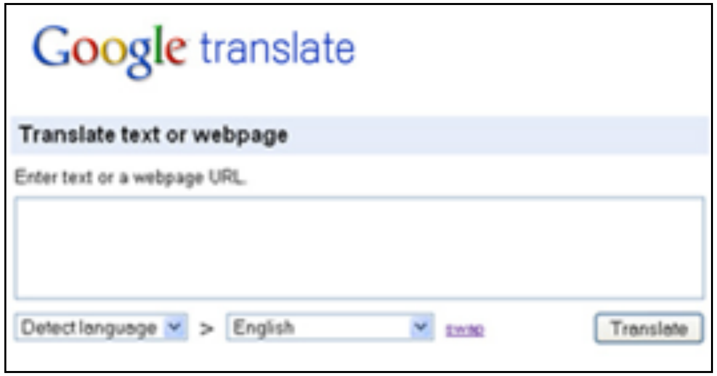
Machine Translation

Jason Baldridge
UT Austin
Language and Computers

Many slides used from Jim Martin, Kevin Knight, Katrin Erk, Markus Dickinson, and Detmar Meurers

Today: End-User Products



...

US Government funded HLT research

SRI International

IBM T. J. Watson

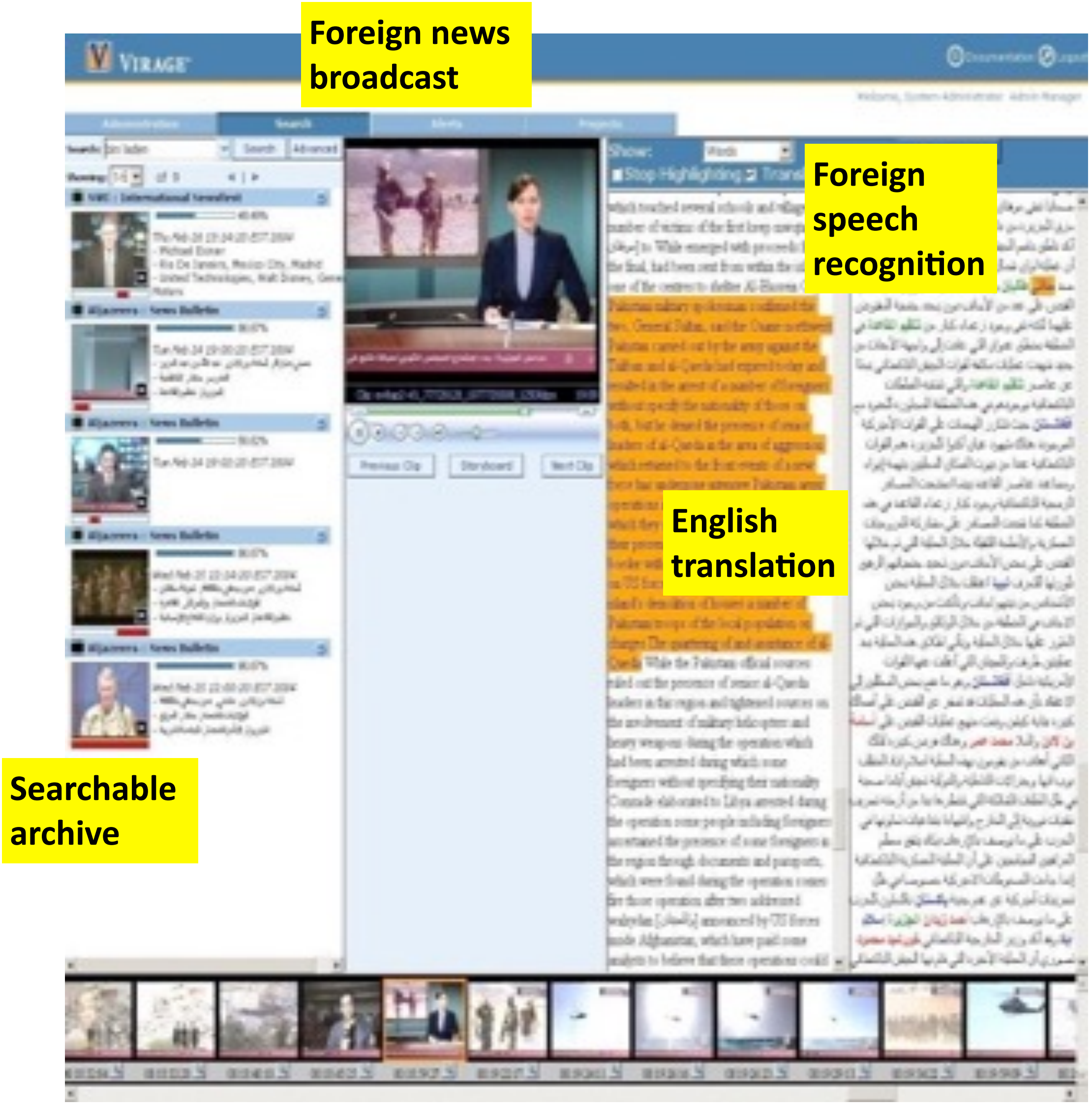
Columbia University

...

many research institutions + many universities + researchers, faculty, graduate students

+ US Government funding of computing systems research

Tools for Military & Intelligence



US Government funded research

Sources: Virage, BBN, Language Weaver, Voxtec

Hard for Computers!

olive oil

peanut oil

sesame oil

source:
Elaine Rich

Hard for Computers!

olive oil

peanut oil

sesame oil

baby oil?

source:
Elaine Rich

Hard for Computers!

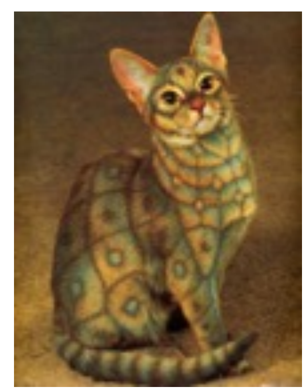
olive oil

peanut oil

sesame oil

baby oil?

why cats paint ≠ why paint cats



source:
Elaine Rich

Hard for Computers!

interest

bank loan interest?

concern, fascination?

part ownership?

Hard for Computers!

interest in IBM

bank loan interest?

concern, fascination?

part ownership?

Hard for Computers!

a financial **interest** in IBM

bank loan interest?

concern, fascination?

part ownership?

Humans do this effortlessly.

Hard for Computers!

company? spokesperson? building?

a financial **interest** in IBM

bank loan interest?

concern, fascination?

part ownership?

Humans do this effortlessly.

Thousands of Languages

MANDARIN	885,000,000
SPANISH	332,000,000
ENGLISH	322,000,000
BENGALI	189,000,000

TURKISH	59,000,000
URDU	58,000,000
MIN NAN (China)	49,000,000
JINYU (China)	45,000,000

HINDI	182,000,000
PORTUGUESE	170,000,000
RUSSIAN	170,000,000
JAPANESE	125,000,000
GERMAN	98,000,000

GUJARATI	44,000,000
POLISH	44,000,000
ARABIC	42,500,000
UKRAINIAN	41,000,000

WU (China)	77,175,000
JAVANESE	75,500,800
KOREAN	75,000,000
FRENCH	72,000,000
VIETNAMESE	67,662,000

ITALIAN	37,000,000
XIANG (China)	36,015,000
MALAYALAM	34,022,000
HAKKA (China)	34,000,000

TELUGU	66,350,000
YUE (China)	66,000,000
MARATHI	64,783,000
TAMIL	63,075,000

KANNADA	33,663,000
ORIYA	31,000,000
PANJABI	30,000,000
SUNDA	27,000,000

Source: Ethnologue

Machine translation and language variation

- Current examples of machine translation
- How can languages differ?
  - Do differences in language lead to differences in thought? The Sapir-Whorf hypothesis
- What makes machine translation hard?
- Evaluating machine translation

Machine translation today

- Google has a “Translate this page” button
- Portuguese -> English translation:
  - Folha de Sao Paulo online newspaper:http://www.folha.uol.com.br
  - Stories: http://contadoresdestorias.wordpress.com
- Exploring Machine Translation:
  - What language other than English do you speak? Do you know an online news page in that language?
  - Translation to English: What are the problems?

What makes MT hard?

- MT is a very difficult task because languages are vastly different. They differ:
  - Lexically: In the words they use
  - Syntactically: In the constructions they allow
  - Semantically: In the way meanings work
  - Pragmatically: In what readers take from a sentence.
- In addition, there is a good deal of real-world knowledge that goes into a translation.

Lexical differences between languages

- "I shot the sherriff":
  - No direct translation to German
  - erschiessen: shoot someone, and that person is dead afterwards
  - anschiessen: shoot someone, and that person is wounded but not dead
- "lend", "borrow" both translate to "leihen" in German
- German "Knopf" can be either "knob" or "button"

- Synonyms are words with the same meaning,
  - like strong/powerful
  - or couch/sofa
- Synonyms between two languages:
  - English book = Russian kniga
  - English music = Spanish música

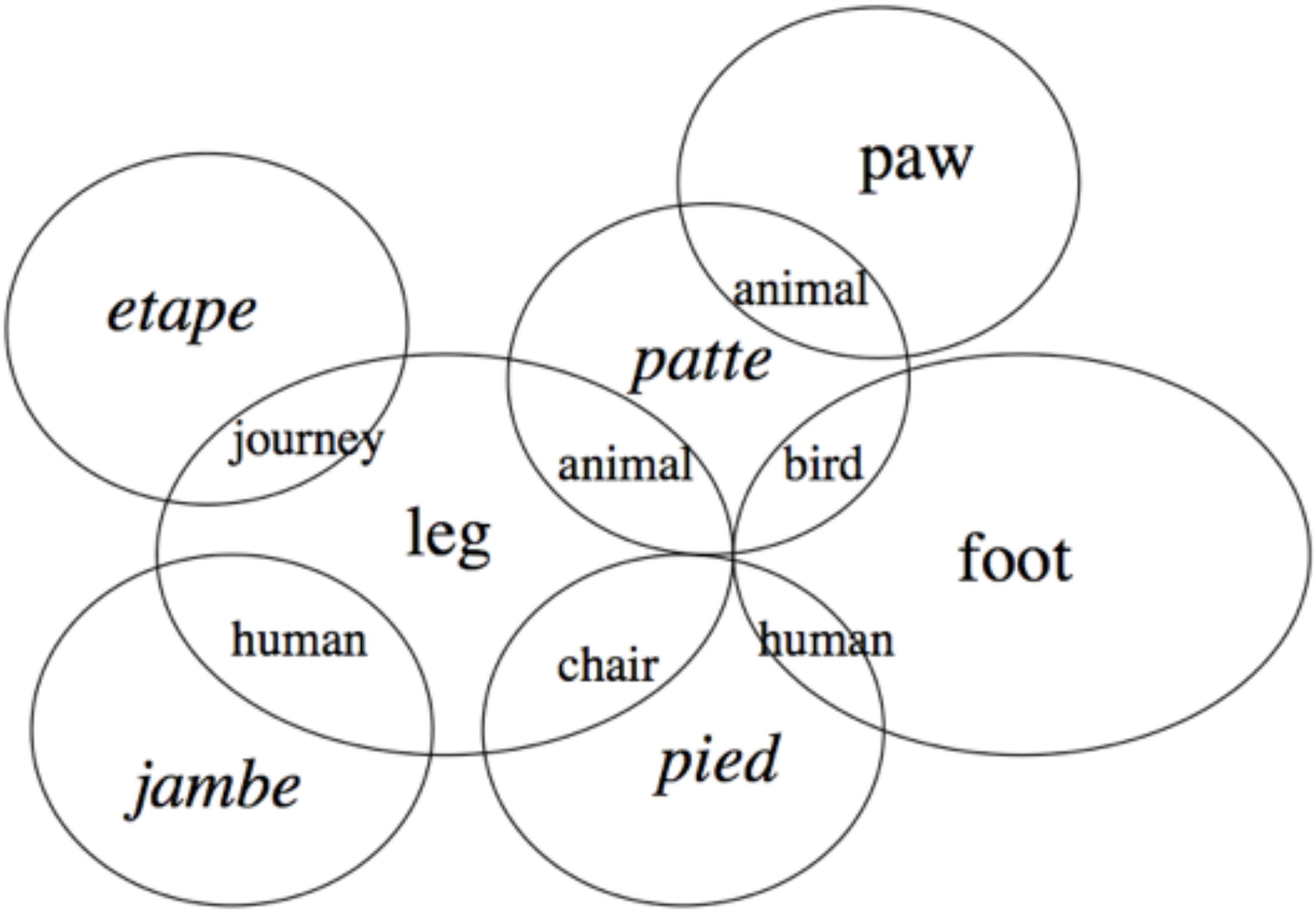
Lexical differences between languages: some terminology

- A hypernym is a more general term:
  - furniture is a hypernym of sofa
- Conversely, a hyponym is a more specific term:
  - sofa is a hyponym of furniture
- Hyponym/hypernym relations between languages:
  - shoot is a hypernym of both German anschiessen and erschiessen
  - English know is a hypernym of French savoir (know a fact) and connaître (be familiar with)
  - English hand and arm are hyponyms of Russian ruka

Lexical differences between languages

- Not always synonymy or hyponymy/hypernymy between languages: sometimes just meaning overlap
- English leg = étape (journey), jambe (human), pied (chair), patte (animal)
- foot = pied (human), patte (bird)
- paw = patte (animal)

Representing semantic overlap: Venn diagrams



Lexical differences between languages: lexical gaps

- Lexical gap: a concept has a word in one language, but not in another
- Translation will then usually require a whole phrase
- French gratiner means something like to cook with a cheese coating
- Hebrew stam means something like I'm just kidding or Nothing special.

Lexical gap example: Portuguese “saudade”

- “Saudade” is a word that means a deep longing or missing of someone or something, mixed with longing and fond remembrance, including some joy upon that remembrance. It is hard to sum up in a few English words -- for example here is the Wikipedia explanation which tries to do justice to the word:

Saudade (singular) or **saudades** (plural) (pronounced [sɐ.uˈdaðɨ] or [sawˈdaðɨ] in Portuguese[1], is a Portuguese language word difficult to translate adequately, which describes a deep emotional state of nostalgic longing for something or someone that one was fond of and which is lost. It often carries a fatalist tone and a repressed knowledge that the object of longing might really never return.

Saudade has been described as a "vague and constant desire for something that does not and probably cannot exist ... a turning towards the past or towards the future".[2] A stronger form of saudade may be felt towards people and things whose whereabouts are unknown, such as a lost lover, or a family member who has gone missing. It may also be translated as a deep longing or yearning for something which does not exist or is unattainable.

Saudade was once described as "the love that remains" or "the love that stays" after someone is gone. Saudade is the recollection of feelings, experiences, places or events that once brought excitement, pleasure, well-being, which now triggers the senses and makes one live again. It can be described as an emptiness, like someone (e.g., one's children, parents, sibling, grandparents, friends) or something (e.g., places, pets, things one used to do in childhood, or other activities performed in the past) that should be there in a particular moment is missing, and the individual feels this absence. In Portuguese, 'tenho saudades tuas', translated as 'I have *saudades* for you' means 'I miss you', but carries a much stronger tone. In fact, one can have 'saudades' of someone with which one is, but have some feeling of loss towards the past or the future.

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

Differences between languages: Multi-word expressions/idioms

- kick the bucket:
  - means die
  - word-by-word translation to another language will be wrong
  - Portuguese *chutar o balde*: literally "kick the bucket", but actually means to give up.
- Conversely, headline from spiegel.de:
  - Mit Tierschutz kraeftig Kohle machen translated as With strong animal protection make coal. Gibberish!
  - Kohle machen means make money

Differences between languages: collocations

- It's strong tea, not powerful tea
  even though strong and powerful are synonyms
- heavy smoker:
  - French grand fumeur (great smoker)
  - German starker Raucher (strong smoker)
- (Compare collocations and idioms)

Differences between languages: light verbs / support verbs

- take into account, take a walk:
  - "take" carries little or no meaning
  - the meaning of the phrase comes from the noun
  - In this case, "take" is a light verb
- Light verbs usually cannot be translated literally
  - Often there will be a <u>different</u> light verb in another language
  - English take a walk = French faire une promenade (make a walk)
  - English make an attempt = Dutch een poging doen (do an attempt)

Syntactic differences between languages

- Word order:
  - English: Subject - verb - object (SVO):
    John sees a squirrel.
    subject verb object
  - Japanese: SOV
  - Arabic: VSO

- Fixedness of word order:
  - English: Subject, object identified by their position relative to the verb
  - Other languages like Czech, Dutch, German, Latin: free(r) word order
    - Hans sieht den Weihnachtsmann.
      John sees Santa.
      subject verb object
    - Den Weihnachtsmann sieht Hans.
      Santa sees John
      object verb subject
  - works because of case marking:
    der: nominative case. den: accusative case

How language is used: Pragmatics

- Translation becomes even more difficult when we try to translate something in context.
  - *Thank you* is usually translated as *merci* in French, but it is translated as *s’il vous plaˆıt* ’please’ when responding to an offer.
- *Can you drive a stick-shift?* could be a request for you to drive my manual transmission automobile, or it could simply be a request for information about your driving abilities.

World knowledge

- Sometimes we have to use **real-world knowledge** to figure out what a sentence means.
    - *Put the paper in the printer. Then switch **it** on*.
- We know what *it* refers to only because we know that printers, not paper, can be switched on.

Evaluating MT systems

- We've seen some translation systems and we know that translation is hard.
- The question now is: How do we evaluate MT systems, in particular for use in large corporations as likely users?
  - How much change in the current setup will the MT system force?
  - How will it fit in with word processors and other software?
  - Will the company selling the MT system be around in the next few years for support and updates?
  - How fast is the MT system?
  - How good is the MT system (quality)?

Evaluating translation quality

- Two main components in evaluating quality:
  - **Intelligibility** = how understandable the output is
  - **Accuracy** = how faithful the output is to the input
    - A common (though problematic) evaluation metric is the BLEU metric, based on n-gram comparisons
- And some methods we can use to gauge these properties:
  - **Error analysis** = how many errors we have to sort through and how they affect intelligibility & accuracy
  - **Test suite** = a set of sentences that our system should be able to handle

The Sapir-Whorf hypothesis

- Edward Sapir (1884-1939),
Benjamin Whorf (1897-1941)
- Linguistic determinism: Language shapes thought
- Weaker version: A person’s thoughts are influenced by the language in which they express them
- Influence of language on thought
  - through vocabulary
  - but through grammar as well

- Influence of grammar on thought:
  - Whorf, work on Hopi
  - Whorf claimed that Hopi had neither words nor grammatical constructions referring to time
  - (Has since been disputed)
- Vocabulary influences that have been discussed in the literature:
  - The infamous “Eskimos have N words for snow” (See G. Pullum, “The great Eskimo vocabulary hoax”)
  - Color terms: work by Kay and Berlin

A different take on language influencing thought: Framing

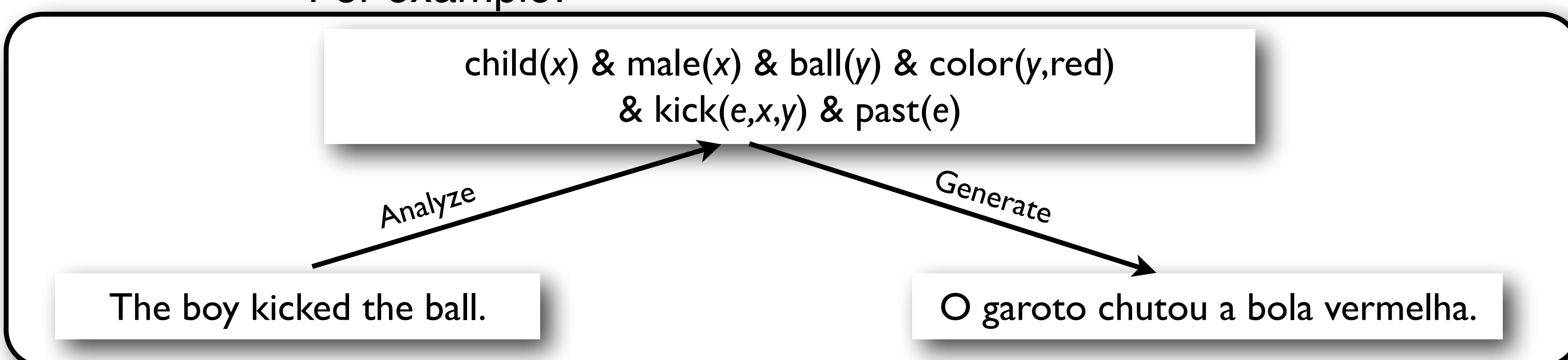
- Goffman, Lakoff
- Influencing opinion by modifying language used to refer to a concept
- Lakoff's example: "tax relief" instead of "tax burden" or "tax responsibilities"
- Compare "He blinked" and "He winked"
- Dihydrogen monoxide (DHMO), a potentially dangerous substance?
  - Against DHMO: http://www.dhmo.org/
  - For DHMO: http://www.armory.com/~crisper/DHMO/

How much do different languages have in common?

- Universal grammar: Are there principles of grammar that all languages share?
  - Human brain pre-wired for language?
- "Poverty of stimulus" argument: Do babies get enough "examples" of language to learn grammar?
- See also: S. Pinker, "The language instinct"

Interlinguas

- It seems like we should be able to create an **interlingua**, or language independent representation of meaning, to translate between languages.
  - An English sentence expresses some set of concepts and relations between them.
  - So does a Portuguese sentence.
  - So does a Chinese sentence....
- So, if we had a way to specify such representations, we could translate English sentences into this interlingua, and then go from that representation to any language we like. For example:



Interlinguas

- Unfortunately, this doesn’t work!
- There is no adequate representation of this nature.
- Hard to say exactly what should be represented. How fine-grained do we need to be?
  - e.g., Japanese distinguishes *older brother* from *younger brother*, so we have to disambiguate English *brother* to put it into the interlingua.
  - Then, if we translate into French, we have to ignore the disambiguation and simply translate it as *frere*, which simply means ’brother’.
- There are massive ambiguities which compound as we perform deeper and deeper levels of analysis.

Translation at different depths: Vauquois triangle

Linguistic meaning can be "transferred" at many different levels of analysis.

Most statistical MT does this, using words and word sequences (phrases).

Some systems do this (mostly, or all, research prototypes).

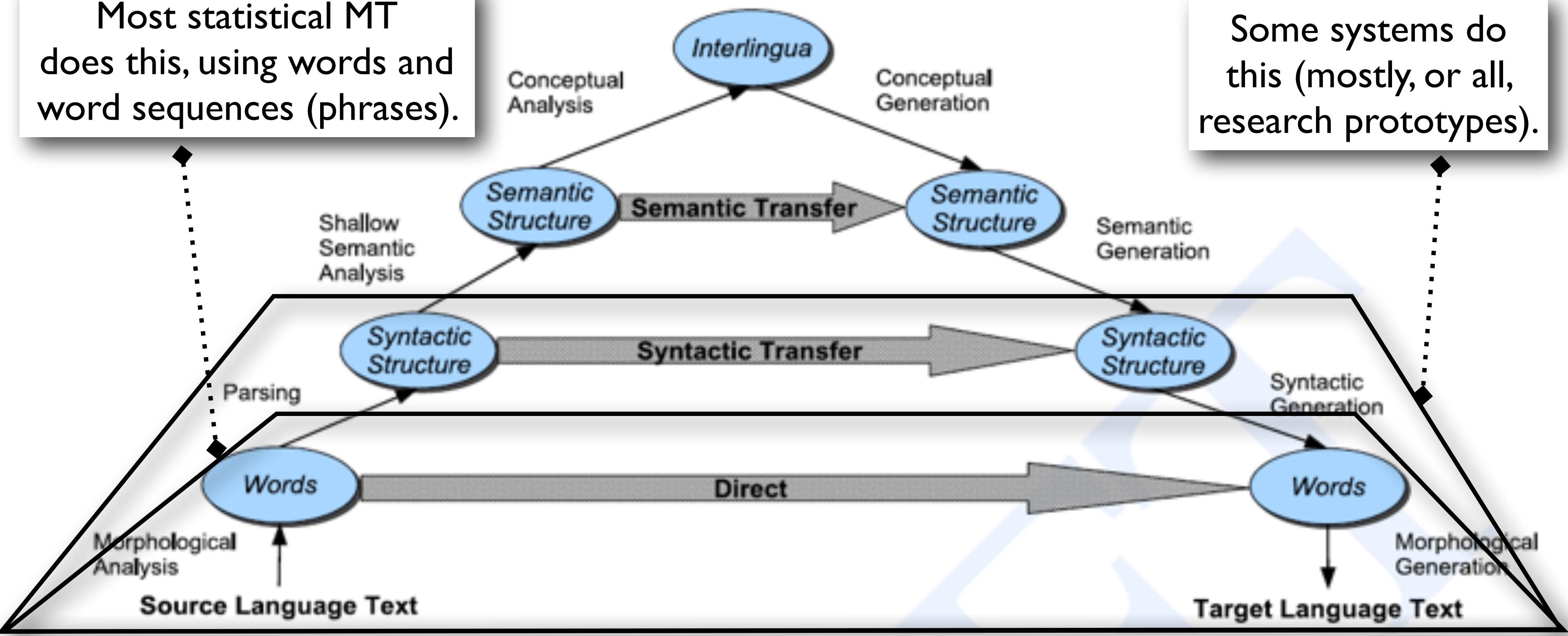


Figure 25.3 from Jurafsky and Martin (2008), *Speech and Language Processing*.

What Has Been Working?

- Focus on **common** linguistic phenomena
  - rather than obscure, difficult cases
- Have machines **learn** from online text and speech data
- **Manage uncertainty** with probabilistic models
- **Evaluate** accuracy of systems
- Develop **common tasks**, compare notes
- **Refine** models

Learning from Data

I have an **interest** in gardening.

I earned 5% **interest** last year.

Zuckerberg holds a controlling **interest**.

My **interest** is purely out of curiosity.

Human annotation of online text.

Learning from Data

concern, fascination

bank loan interest

part ownership

I have an **interest** in gardening.

I earned 5% **interest** last year.

Zuckerberg holds a controlling **interest**.

My **interest** is purely out of curiosity.

concern, fascination

Human annotation of online text.

Learning from Data

interest

bank loan interest?

concern, fascination?

part ownership?

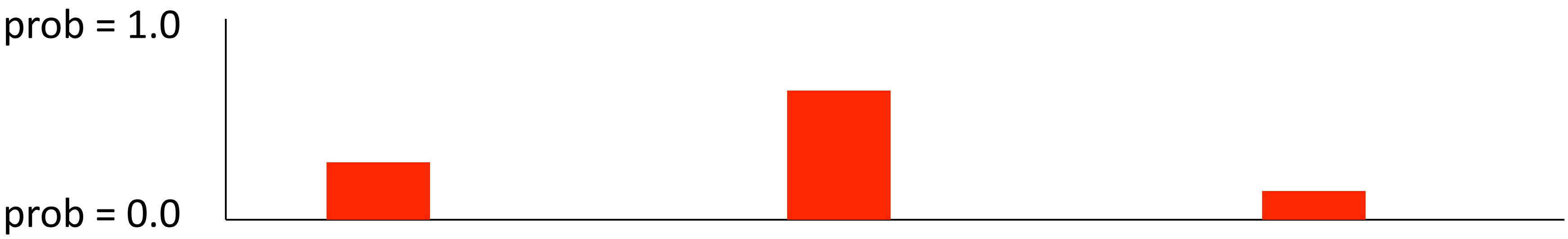
Learning from Data

interest

bank loan interest?

concern, fascination?

part ownership?



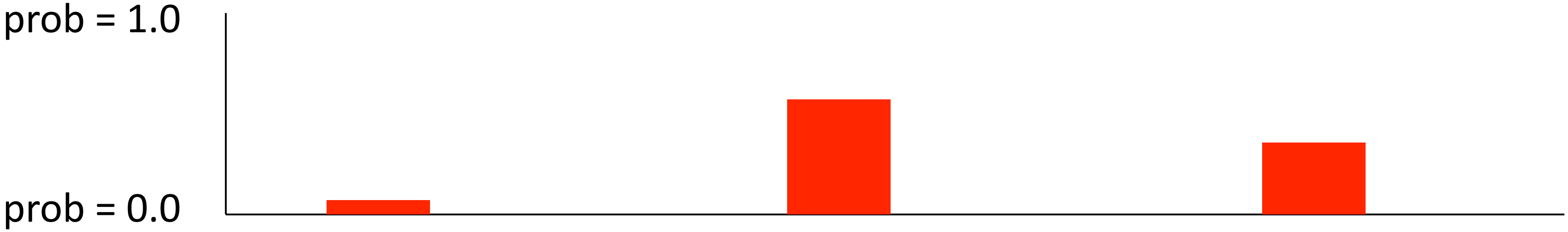
Learning from Data

interest in IBM

bank loan interest?

concern, fascination?

part ownership?



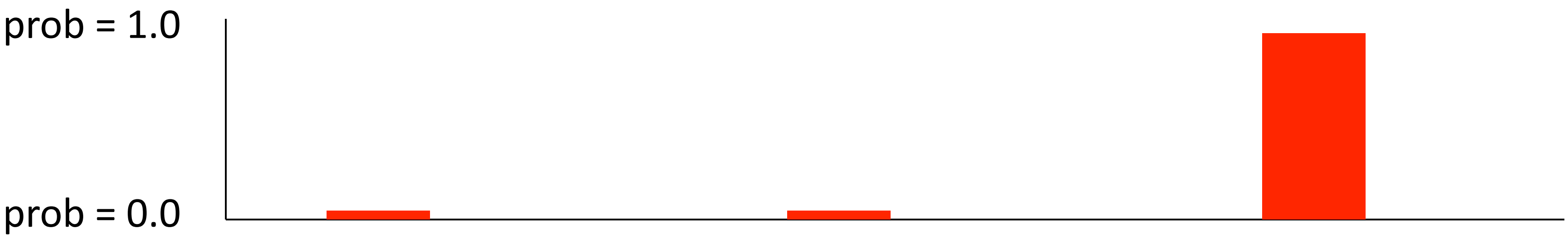
Learning from Data

a financial **interest** in IBM

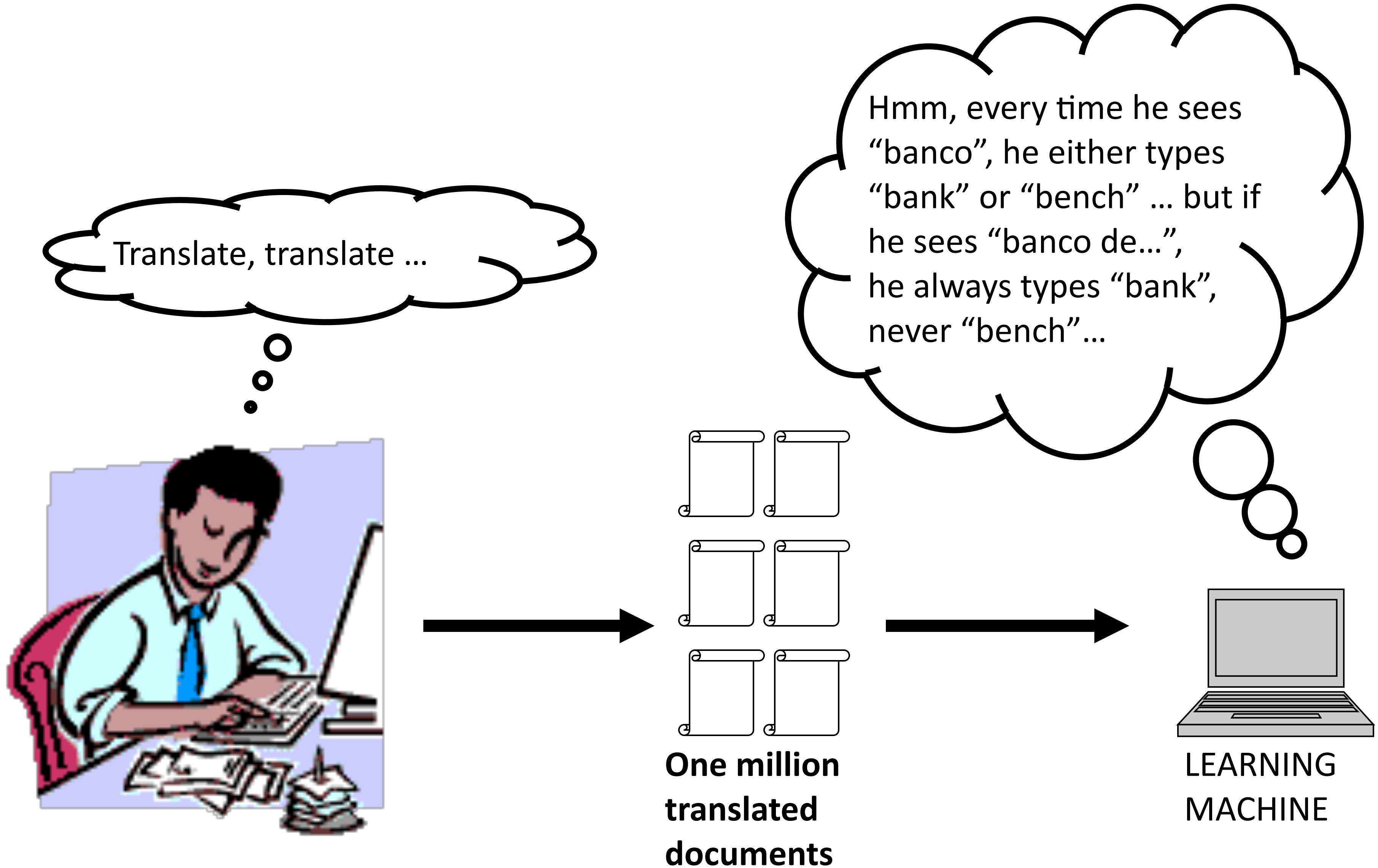
bank loan interest?

concern, fascination?

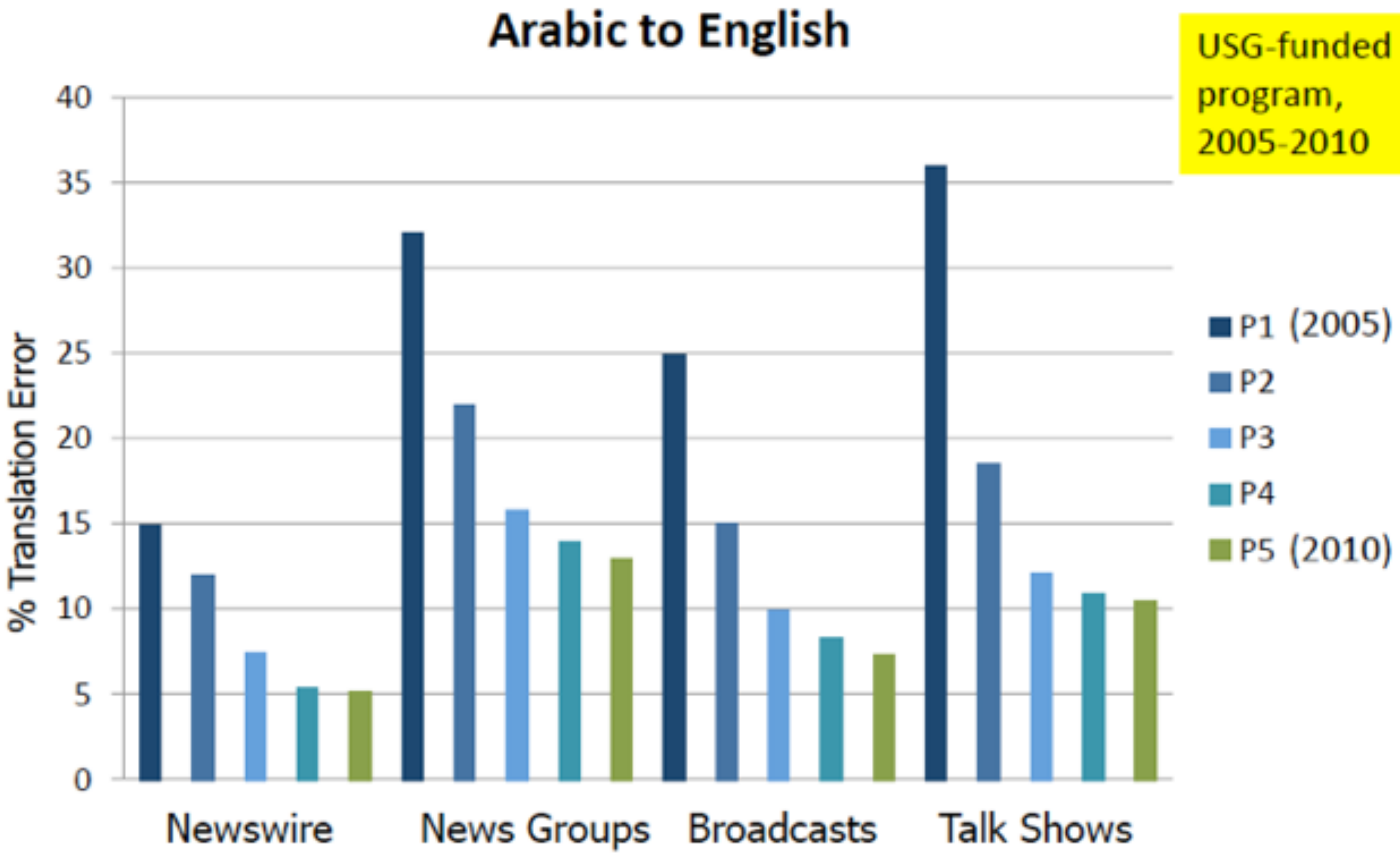
part ownership?



Learning to Translate



Machine Translation Error Rate



source: DARPA

Current Approaches

- Same old noisy channel model…
- If we're translating French to English the French we're seeing is just a weird garbled version of English
- There must have been some process that generated the French from the original English
- The key is to decode the garbles back into the original English by…
- Argmax P(E I F) by Bayes
- A very old idea

Warren Weaver (1947)

When I look at an article in Russian, I say to myself: This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.

Centauri/Arcturan Parallel Corpus [Knight, 1997]

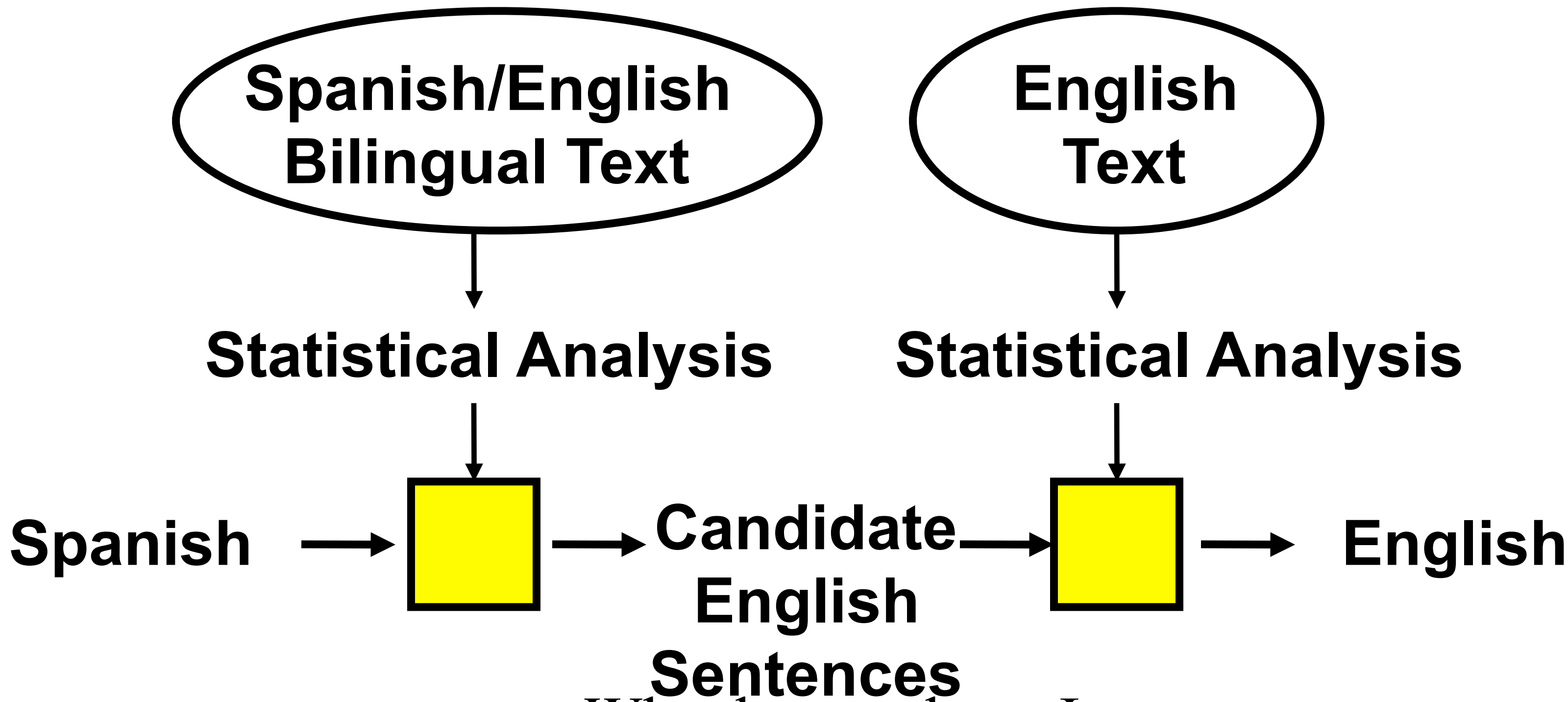
1a. ok-voon ororok sprok . 1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok . 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok . 12b. wat nnat forat arrat vat gat .

Translate this from Centauri to Arcturan:
lalok mok farok kantok ok-yurp crrrok hihok yorok clok jok

English/Spanish Parallel Corpus [Knight, 1997]

1a. Garcia and associates. 1b. Garcia y asociados .	7a. the clients and the associates are enemies . 7b. los clientes y los asociados son enemigos .
2a. Carlos Garcia has three associates . 2b. Carlos Garcia tiene tres asociados.	8a. the company has three groups . 8b. la empresa tiene tres grupos .
3a. his associates are not strong . 3b. sus asociados no son fuertes .	9a. its groups are in Europe . 9b. sus grupos estan en Europa .
4a. Garcia has a company also . 4b. Garcia tambien tiene una empresa .	10a. the modern groups sell strong pharmaceuticals. 10b. los grupos modernos venden medicinas fuertes .
5a. its clients are angry . 5b. sus clientes estan enfadados.	11a. the groups do not sell zenzanine . 11b. los grupos no venden zanzanina .
6a. the associates are also angry . 6b. los associados tambien estan enfadados.	12a. the small groups are not modern. 12b. los grupos pequenos no son modernos .

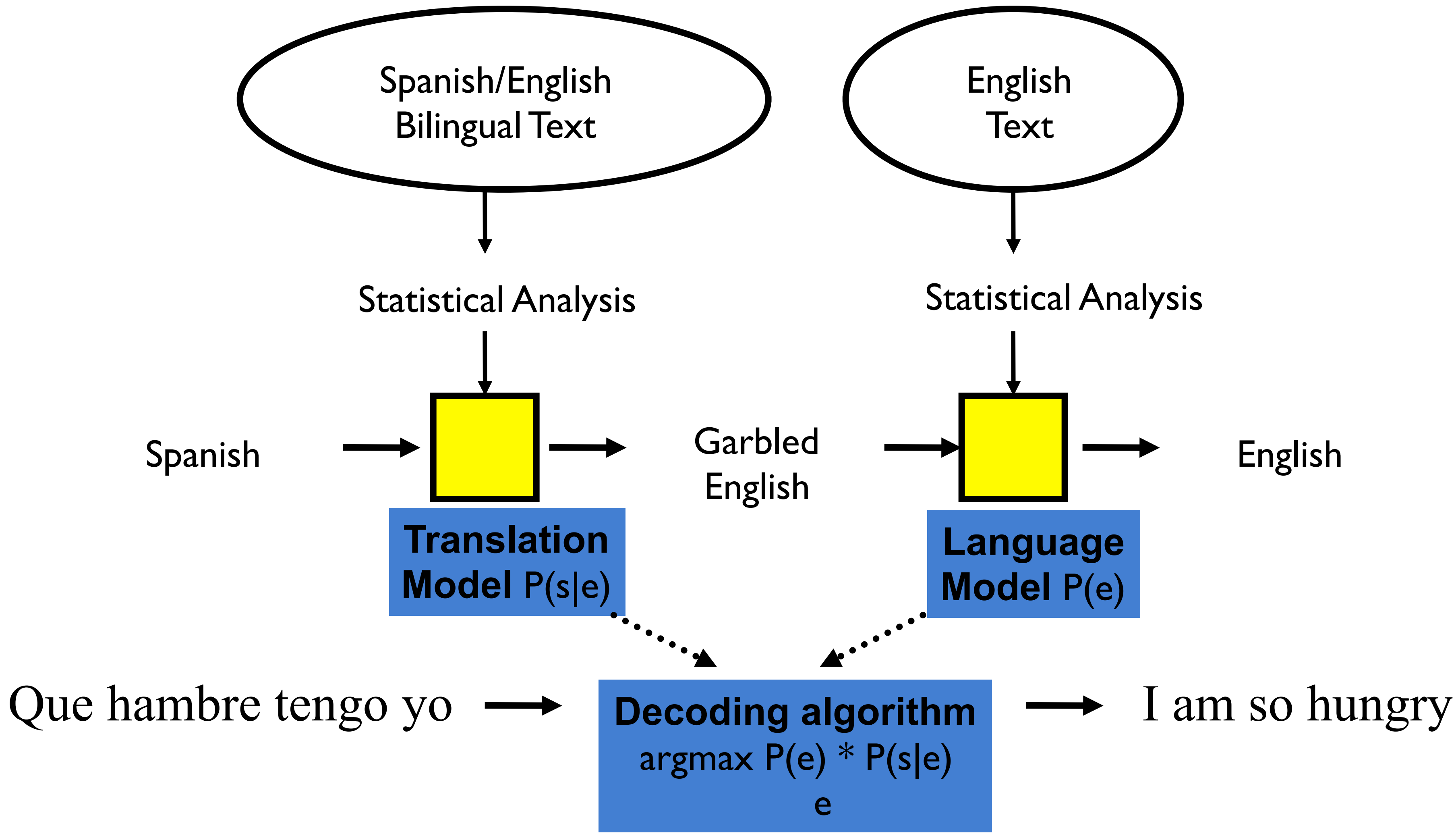
Translate this from English to Spanish:
the modern clients in Europe do not sell pharmaceuticals also .



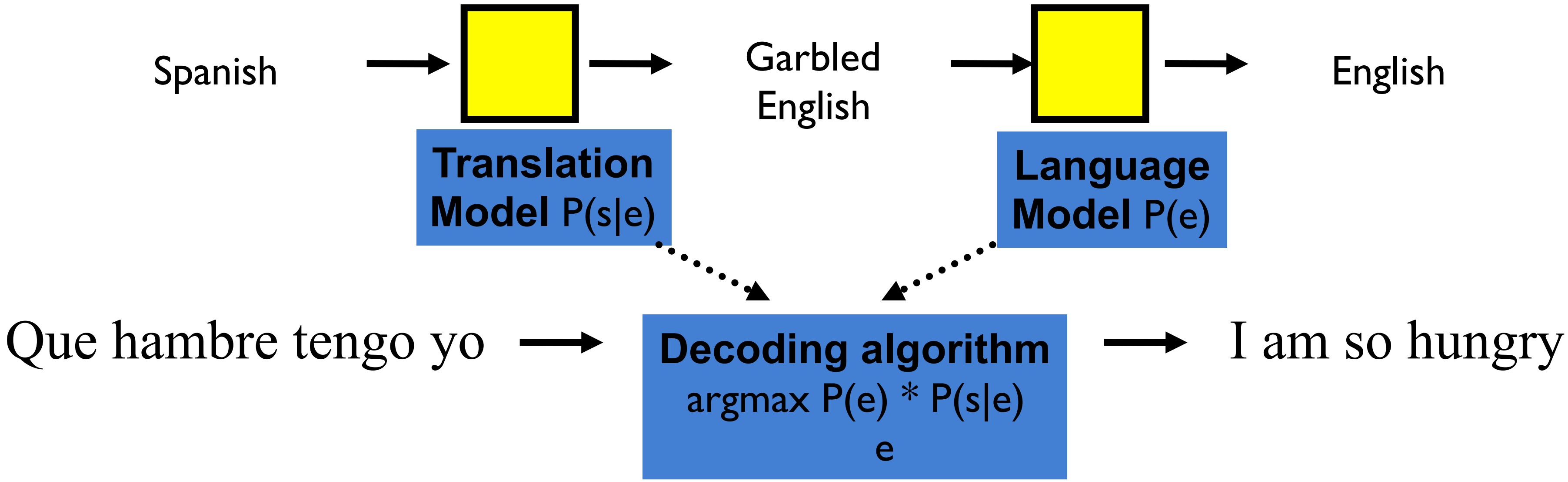
What hunger have I,
Hungry I am so,
I am so hungry,
Have I that hunger …

Que hambre tengo yo → (candidates above) → I am so hungry

Statistical MT Systems



Bayes Rule/Noisy Channel



Given a source sentence s, the decoder should consider many possible translations … and return the target string e that maximizes

$$P(e \mid s)$$

By Bayes Rule, we can also write this as:

$$P(e) \times P(s \mid e) \,/\, P(s)$$

and maximize that instead. P(s) never changes while we compare different e's, so we can equivalently maximize this:

$$P(e) \times P(s \mid e)$$

Three Sub-Problems of Statistical MT

- Language model
  - Given an English string e, assigns P(e) by formula
  - good English string -> high P(e)
  - random word sequence -> low P(e)
- Translation model
  - Given a pair of strings <f,e>, assigns P(f | e) by formula
  - <f,e> look like translations -> high P(f | e)
  - <f,e> don't look like translations -> low P(f | e)
- Decoding algorithm
  - Given a language model, a translation model, and a new sentence f … find translation e maximizing P(e) * P(f | e)

Translation Model?

Generative story:

Mary did not slap the green witch

Source-language morphological analysis

Source parse tree

Semantic representation

Generate target structure

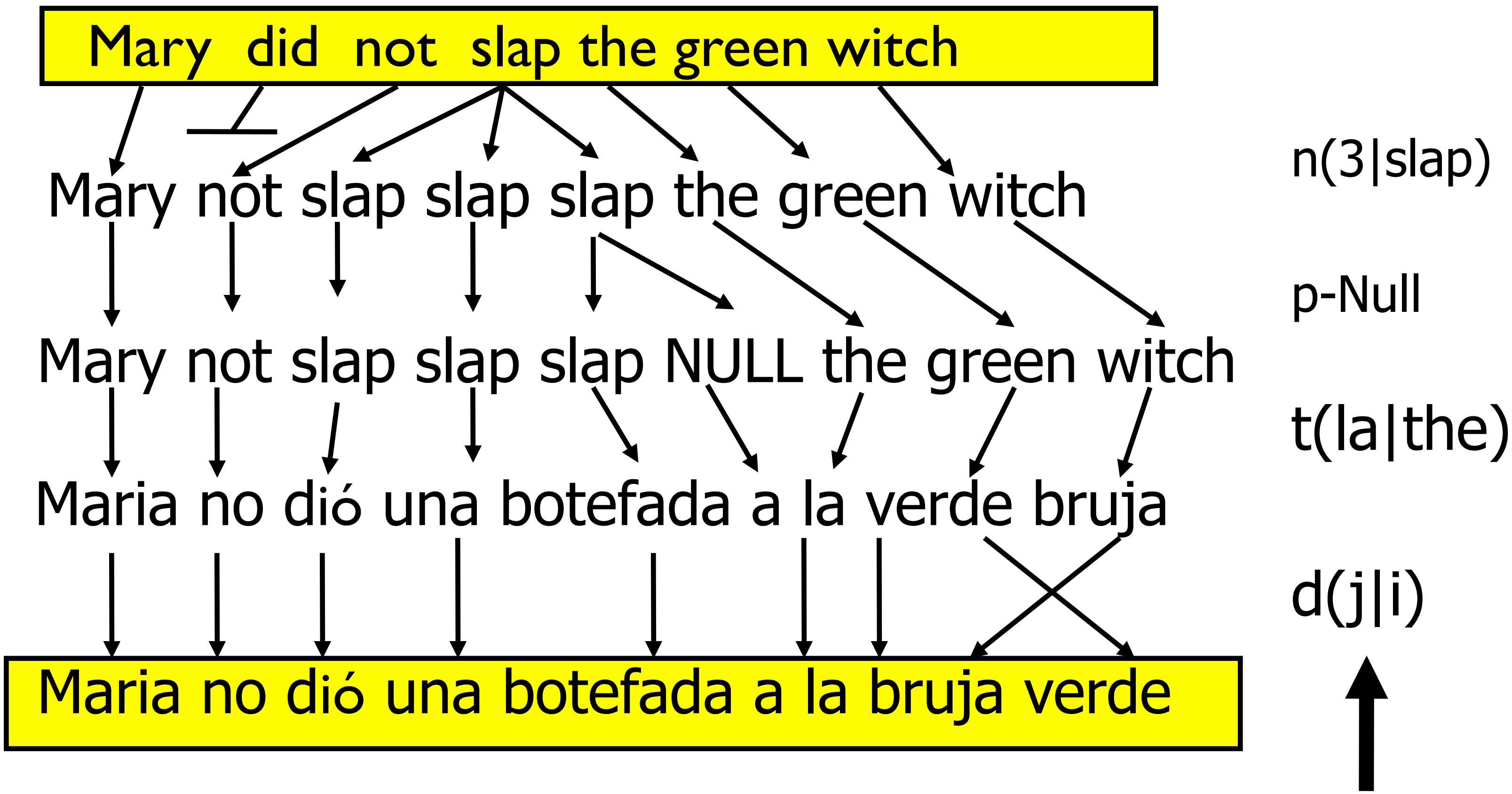
Maria no dió una botefada a la bruja verde

Way too hard.

The Classic Translation Model

Word Substitution/Permutation [IBM Model 3, Brown et al., 1993]

Generative story:



Parts List

- We need probabilities for
  - n (x|y): The probability that word y will yield x outputs in the translation… (fertility)
  - p_{null}: The probability of a null insertion
  - $t(t_{word} \mid s_{word})$: The actual word translation probability table; the probability of a target word given a source word.
  - d(j|i) the probability that a word at position i will make an appearance at position j in the translation
- Every one of these can be learned from a word aligned corpus.
- We'll look at how to determine the translation probabilities.

Translation probabilities from word-aligned parallel texts.

- Say you had the English/Portuguese parallel corpus given below.
- The words can be aligned such that the English words expressing a concept are connected to the Portuguese words expressing the same concept.

```
1a. is he tall ?

1b. ele e alto ?


2a. I saw you when you were walking in the tall building .

2b. eu te vi quando voce estava andando no predio alto .


3a. is the tall dog that is with you yours ?

3b. o cachorro alto que esta com voce e seu ?
```

Translation probabilities from word-aligned parallel texts.

- The probability of a Portuguese word given an English word can simply be read off the alignments.

```
1a. is he tall ?          2a. I saw you when you were walking in the tall building .

1b. ele e alto ?          2b. eu te vi quando voce estava andando no predio alto .


3a. is the tall dog that is with you yours ?

3b. o cachorro alto que esta com voce e seu ?
```

$p(f=\text{"ele"} \mid e=\text{"he"}) = 1/1 = 1.0$

$p(f=\text{"voce"} \mid e=\text{"you"}) = ??$

$p(f=\text{"alto"} \mid e=\text{"tall"}) = 3/3 = 1.0$

$p(f=\text{"e"} \mid e=\text{"is"}) = 2/3 = .67$

$p(f=\text{"esta"} \mid e=\text{"is"}) = 1/3 = .33$

Note: .67 + .33 = 1.0

Parts List

- Every one of these can be learned from a sentence aligned corpus…
  - Ie. A corpus where sentences are paired but nothing else is specified
- And the EM algorithm

What if we don’t have word alignments?

- Word alignments require human annotation (costs money), so we’d prefer to learn model parameters without them.
- But, this is a type of chicken-and-egg problem:
  - if we have word alignments we can estimate the parameters of the model
  - if we have parameters, we can estimate the word alignments
- The answer: the expectation-maximization algorithm applied to sentence aligned parallel texts.

Expectation Maximization (EM) algorithm

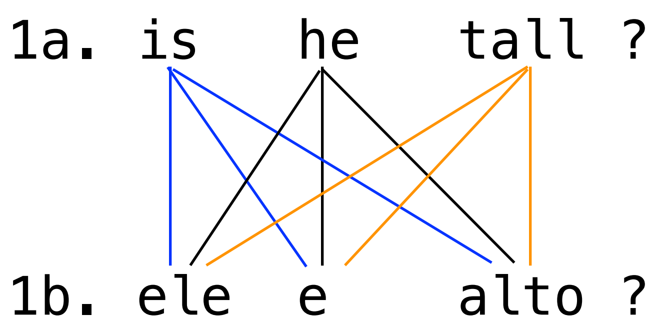
- The EM algorithm works forwards and backwards to estimate the model parameters:

1. initialize model parameters (e.g. uniform or random)
2. (re-)assign probabilities to the missing data
3. (re-)estimate model parameters from completed data (weighted counts)
4. iterate, i.e., repeat steps 2 & 3 until you hit some stopping point

- The k-means algorithm we looked at for authorship attribution is actually a form of EM.
  - We guessed at initial centroids.
  - We then re-estimated the centroids based on the how the initial centroids clustered the data.

EM for learning word alignments

- We assume that sentences are aligned.
- We treat each sentence as an unordered collection of words (the bag-of-words we've seen before).
- We use a simple heuristic to generate initial translation probabilities: If a word appears in a sentence, then we act as if it is aligned with every word in the translated sentence.
- Here's what it looks like:



Translation probabilities from sentence-aligned parallel texts.

```
1a. is he tall ?         2a. I saw you when you were walking in the tall building .

1b. ele e alto ?         2b. eu te vi quando voce estava andando no predio alto .


3a. is the tall dog that is with you yours ?

3b. o cachorro alto que esta com voce e seu ?
```

Examples

p(f=“ele”|e=“he”) = 1/3 = .33

p(f=“e”|e=“is”)
= (1+1+1)/(3+9+9) = 3/21 = .14

p(f=“ele”|e=“is”)
= 1/(3+9+9) = 1/21 = .05

Is that right?!

Answer: it isn’t right or wrong -- what matters is (1) relative probability and (2) this is just a starting point!

There is a simple refinement to make these values better. Can you spot it?

Answer: use counts that encode the length of the sentence so that short sentences are more convincing evidence of translation equivalence, e.g.

```
p(f=“e”|e=“is”)
  = (1/3 + 1/9 +1/9)/(3/3 + 9/9 + 9/9)
  = (5/9) / 3
  = 5/27
  = .19
```

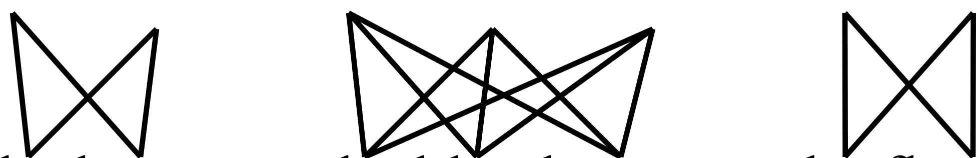
(This won’t be on the test, but forms the basis for what we’ll discuss next.)

Translation probabilities from sentence-aligned parallel texts.

- With more sentences, the probability of the correct translations will rise, because words and their translations are more likely to co-occur in sentence pairs than are random word pairs.
  - E.g., *e* and *esta* will tend to occur more often, compared to *ele* “he”, *eu* “I”, *voce* “you”, etc., in sentences whose translation includes “is”.
- Also, these are just initial weights for the EM procedure, which will iteratively re-estimate them and converge to much better values.
- A high-level view of this is given in the following slides.

Word Alignment

… la maison … la maison bleue … la fleur …

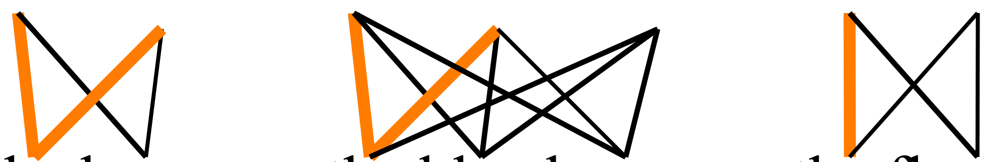


… the house … the blue house … the flower …

- Assume that All word alignments equally likely.
- That is, that all P(french-word | english-word) are equal
- Recall that we want P(f|e)

Word Alignment

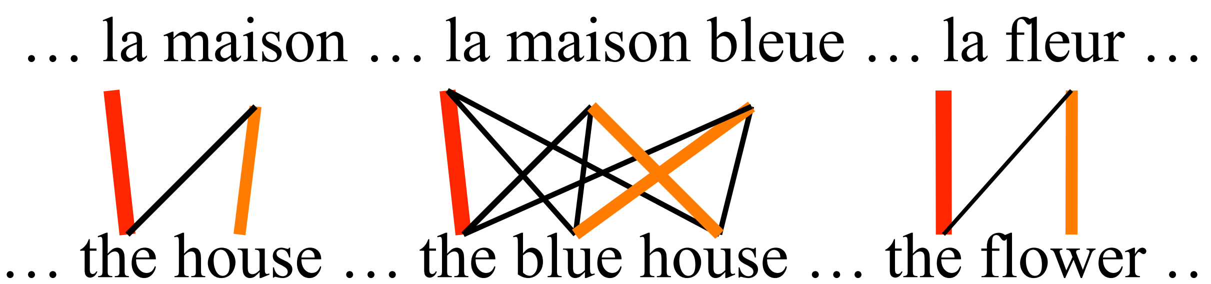
… la maison … la maison bleue … la fleur …



… the house … the blue house … the flower …

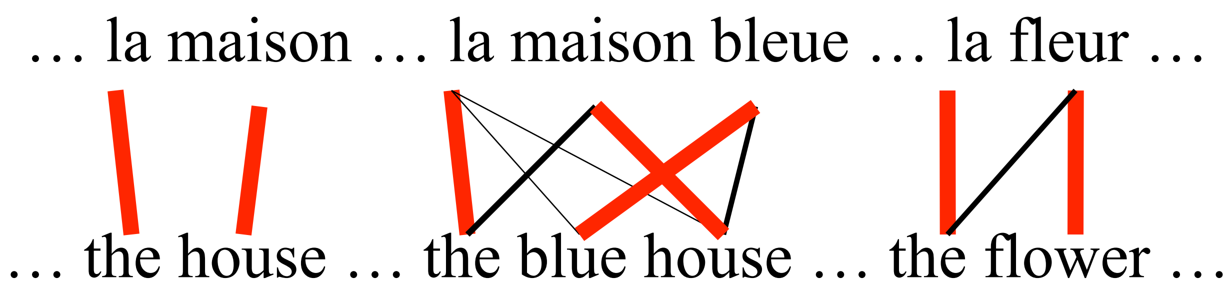
"la" and "the" observed to co-occur frequently, so P(la | the) is increased.

Word Alignment



Connections between e.g. fleur and flower are more likely (pigeon hole principle).

Word Alignment



settling down after another iteration

Inherent hidden structure revealed by EM training!

… la maison … la maison bleue … la fleur …

… the house … the blue house … the flower …

$$p(\text{la}|\text{the}) = 0.453$$
$$p(\text{le}|\text{the}) = 0.334$$
$$p(\text{maison}|\text{house}) = 0.876$$
$$p(\text{bleu}|\text{blue}) = 0.563$$

Parts List

- Given a sentence alignment we can induce a word alignment
- Given that word alignment we can get the p, t, d and n parameters we need for the model.
- Ie. We can argmax P(e|f) by max over P(f|e)*P(e)… and we can do that by iterating over some large space of possibilities.

Decoding

Que	hambre	tengo	yo
what	hunger	have	I
that	hungry	am	me
so		make	
where			

what hungry make me

Decoding

Que	hambre	tengo	yo
what	hunger	have	I
that	hungry	am	me
so		make	
where			

so hungry have me

Decoding

Que	hambre	tengo	yo
what →	hunger →	have →	I
that	hungry	am	me
so		make	
where			

what hunger have I

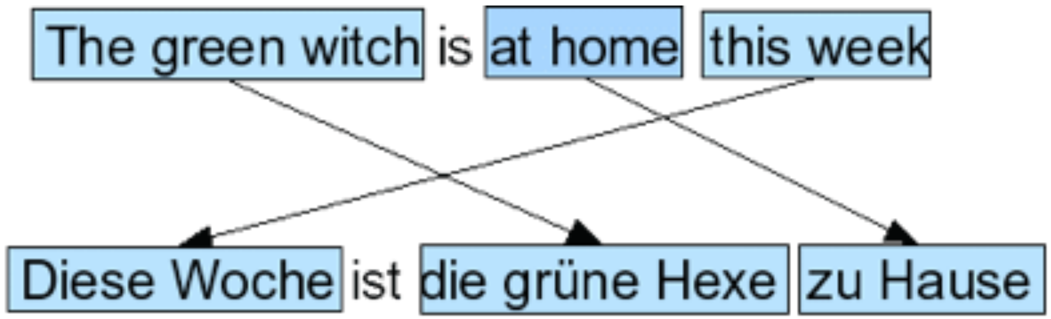
Que	hambre	tengo	yo
what	hunger	have	I
that	hungry	am	me
so		make	
where			

I am so hungry

Flaws of Word-Based MT

- Multiple English words for one French word
  - IBM models can do one-to-many (fertility) but not many-to-one
- Phrasal Translation
  - “real estate”, “note that”, “interest in”
- Syntactic Transformations
  - Languages with differing word orders (SVO vs. VSO)
  - Translation model penalizes any proposed re-ordering
  - Language model not strong enough to force the verb to move to the right place

Intuition of phrase-based translation (Koehn et al. 2003)



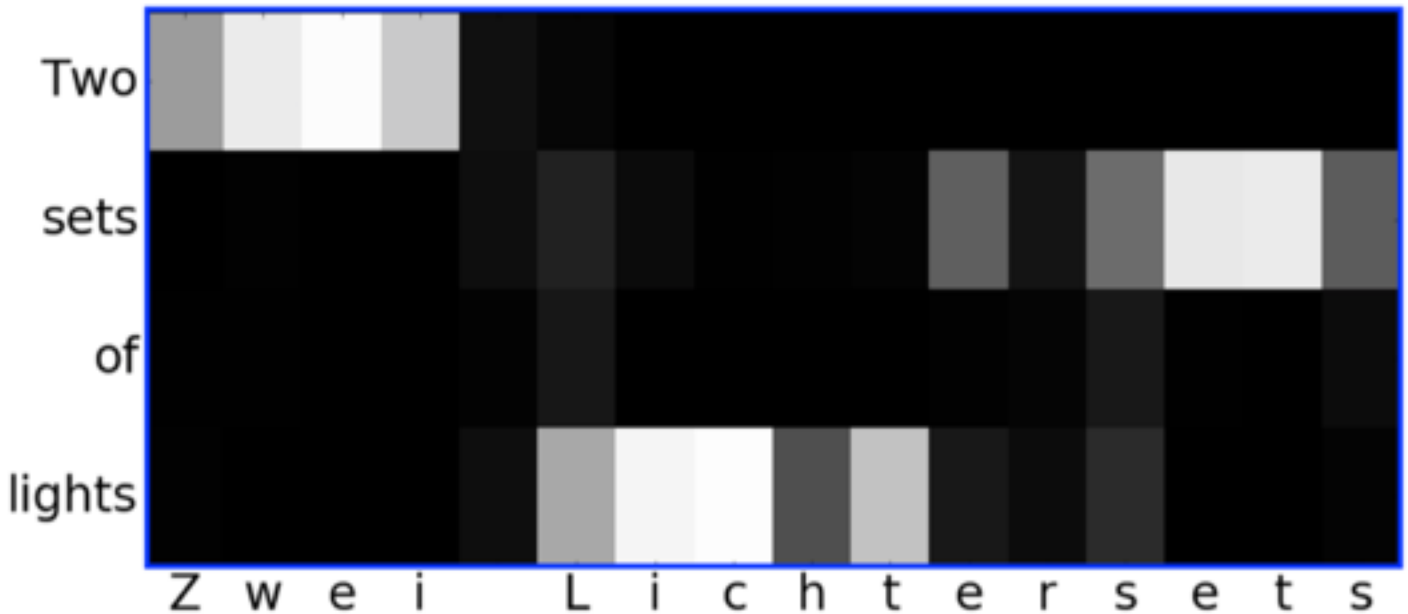
- Generative story has three steps
  1) Group words into phrases
  2) Translate each phrase
  3) Move the phrases around

Advantages of Phrase-Based SMT

- Many-word-to-many-word translations can handle non-compositional phrases (e.g., “real estate”)
- Local context is very useful for disambiguating
  - “Interest rate” → …
  - “Interest in” → …
- The more data, the longer the learned phrases
  - Sometimes whole sentences
  - Interesting parallel to concatenative synthesis for TTS

Character level neural models

- New MT architectures are now coming out that use deep learning and operate at the level of characters.
- Advantages: words are no longer unanalyzable chunks, and overall system has fewer components.

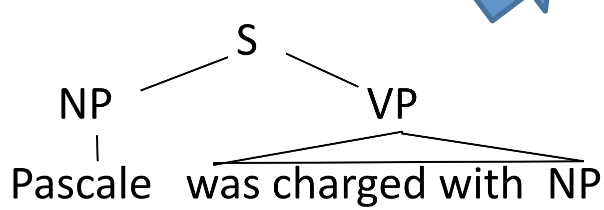


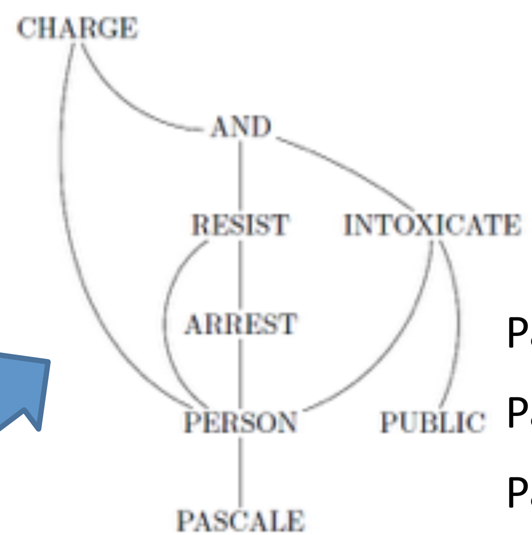
Character-level alignments: Chung et al (2016), http://arxiv.org/abs/1603.06147

Lots of Progress, But …

- Machines make lots of errors
- Machines need a **deeper understanding** of what they read and hear

Pascale was charged with public intoxication and resisting arrest.





Pascale was charged

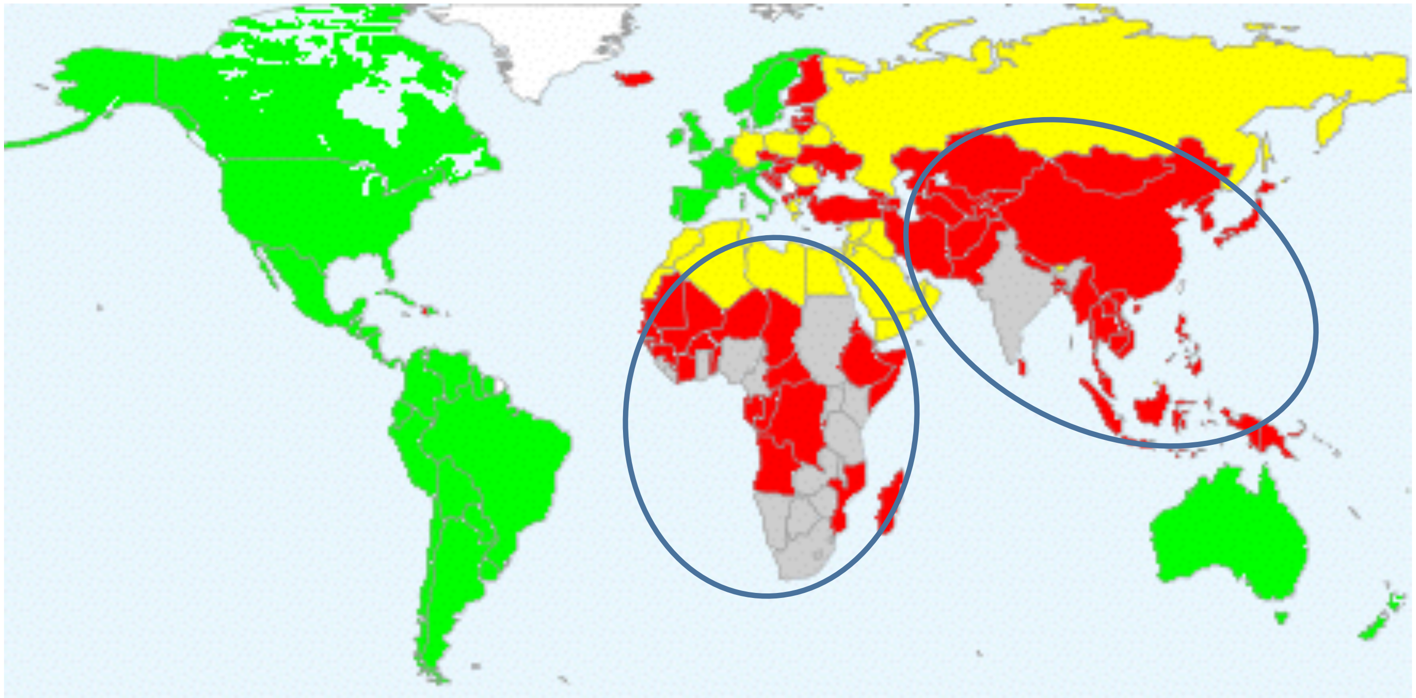
Pascale was intoxicated (?)

Pascale was resisting (?)

Pascale was being arrested

Things We Can’t Do Yet

Let's Not Forget …



Final note: http://news.bbc.co.uk/2/hi/uk_news/wales/7702913.stm

- Translation can go dramatically wrong even with humans

The Welsh text says:
I am not in the office at the moment. Send any work to be translated.