6.863J Natural Language Processing
Lecture 23: Machine Translation 2

Instructor: Robert C. Berwick
berwick@ai.mit.edu

The Menu Bar

- Administrivia:
  - Final project!
  - Agenda:

    Machine Translation (MT) as a 'litmus test' or 'sandbox' (graveyard?) for putting together all of NLP
  - Practical systems: Phraselator; Systran (Babelfish); Logos,…
  - MT: the statistical approach – Star Trek view
  - Formalize what we did last time: Shake 'n Bake
  - Divide & conquer: 4 steps
    - Noisy channel model
    - Language Model
    - Translation model
    - Scrambling & Fertility

Alien languages: Alpha-centauri & Betelgeuse

```
1a. ok-voon ororok sprok .     2a. ok-drubel ok-voon anok plok sprok
1b. at-voon bichat dat .  2b. at-drubel at-voon pippat rrat dat .

3a. erok sprok izok hihok ghirok .    4a. ok-voon anok drok brok jok
3b. totat dat arrat vat    hilat . 4b. at-voon krat pippat sat lat .

5a. wiwok farok izok stok .    6a. lalok sprok izok jok stok .
5b. totat jjat quat cat . 6b. wat dat krat quat cat .

7a. lalok farok ororok lalok sprok izok enemok .
7b. wat jjat bichat wat dat vat eneat .

8a. lalok brok anok plok nok .9a. wiwok nok izok kantok ok-yurp .
8b. iat lat pippat rrat nnat .    9b. totat nnat quat oloat at-yurp .

10a. lalok mok nok yorok ghirok clok .
10b. wat nnat gat mat bat hilat .

11a. lalok nok crrrok hihok yorok zanzanok .
11b. wat nnat arrat mat zanzanat .

12a. lalok rarok nok izok hihok mok .
12b. wat nnat forat arrat vat gat .
```


We will build two things

- Assume word-word translation – though not same word order
- Use alignment of words to build translation dictionary
- Use translation dictionary to improve the alignment – because it eliminates some possibilities

To begin

```
1a. ok-voon ororok sprok .          2a. ok-drubel ok-voon anok plok sprok .
1b. at-voon bichat dat .            2b. at-drubel at-voon pippat rrat dat .

3a. erok sprok izok hihok ghirok .  4a. ok-voon anok drok brok jok .
3b. totat dat arrat vat    hilat .  4b. at-voon krat pippat sat lat .

5a. wiwok farok izok stok .         6a. lalok sprok izok jok stok .
5b. totat jjat quat cat .           6b. wat dat krat quat cat .

7a. lalok farok ororok lalok sprok izok enemok .
7b. wat jjat bichat wat dat vat eneat .

8a. lalok brok anok plok nok .      9a. wiwok nok izok kantok ok-yurp .
8b. iat lat pippat rrat nnat .      9b. totat nnat quat oloat at-yurp .

10a. lalok mok nok yorok ghirok clok .
10b. wat  nnat gat mat     bat hilat.

11a. lalok nok crrrok hihok yorok zanzanok .
11b. wat   nnat arrat mat   zanzanat .

12a. lalok rarok nok izok hihok mok .
12b. wat nnat forat arrat vat gat .
```

Translation dictionary:
ghiork – hilat
ok-drubel – at-drubel
ok-voon – at-voon
ok-yurp – at – yurp
zanzanok - zanzanat

OK, what does pairing buy us?

- Sentence 1: 2 possibilities left…
  1. ororok ↔ bichat & sprok ↔ dat
  2. ororok ↔ dat & sprok ↔ bichat

(But also: what if ororok untrans aux v…?)
Which is more likely?
Look for sentence w/ sprok but not ororok
Sentence (2a)
Link throughout corpus (1, 2, 3, 6, 7)
Sentence (2) now looks like a good place to crack…

Sentences 2, 3…

- S2: anok plok/pippat rrat
- S4: **4a. ok-voon anok drok brok jok .**
  4b. at-voon krat pippat sat lat .

Ok, anok ↔ pippat & plok ↔ rrat

S3: So far we have:

```
 erok sprok  izok hihok ghirok
totat  dat  arrat  dat hilat
```

Look at 8; 11; 3 & 12; 5, 6, 9

This suggests

```
erok sprok  izok hihok ghirok

totat  dat arrat  vat   hilat
```

Note:

- Aligning builds the translation dictionary
- Building the translation dictionary aids alignment
- "Decipherment"
- We shall see how this can be automated next time

The dictionary so far…

```
anok - pippat             ok-yurp - at-yurp
erok - total              ok-voon - at-voon
ghirok - hilat            ororok - bichat
hihok - arrat             plok - rrat
izok - vat/quat           sprok - dat
ok-drubel - at-drubel     zanzanok - zanzanat
```

Full dictionary

```
anok - pippat
brok - lat
crrok - none?
drok - sat
enemok - eneat
erok - total
farok - jjat
ghirok - hilat
hihok - arrat
izok - vat/quat
jok - krat
kantok - oloat
lalok - wat/iat
mok - gat ok-yurp - at-yurp
nok - nnat clok - bat
ok-drubel - at-drubel
ok-yurp - at-yurp
ororok - bichat
plok - rrat
rarok - forat
sprok - dat
stok - cat
wiwok - totat
yorok - mat
zanzanok - zanzanat
```


If you work through it you'll get all the pairs here, save 1: <u>crrrok</u>

- But you are suddenly abducted to the Federation Translation Center & presented with this sentence from Betelgeuse to translate into Alpha-Centaurian:
- **`iat lat pippat eneat hilat oloat at-yurp .`**

Translation B to A

- 13(B) iat lat pippat eneat hilat oloat at-yurp
- Consult dictionary – 7 words can be directly looked up
- iat lat pippat eneat hilat oloat at-yurp
- Many possible word orders for ‘felicitous’ translation!…how do we decide?

You are given this fragment of Alpha-C text & its bigrams… to help

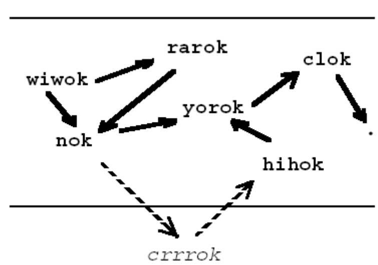
The translation (answer sheet)

- **iat lat pippat eneat hilat oloat at-yurp**
- **Word for word:**
  (13a) Lalok brok anok enemok ghirok kantok ok-yurp
  Lalok brok anok {enemok ghirok kantok ok-yurp}
  Lalok brok anok ghirok {enemok kantok ok-yurp}
  Lalok brok anok ghirok enemok {kantok ok-yurp}
  Final: Lalok brok anok ghirok enemok kantok ok-yurp

- **(14b) totat nnat forat arrat mat bat**
  erok? wiwok?

Now what? Wiwok to…?

Various possibilities



Wiwok…

(14a) Wiwok rarok nok crrrok hihok yorok clok…

How is this like/unlike ‘real’ translation

- Only 2 of the 27 AlphaC words were ambiguous
- Sentence length unchanged in all but one
- Sentences much shorter than typical
- Words & context -
- Output word order should be sensitive to input word order (J. loves M, M loves J)
- Data cooked
- No phrasal dictionary (amok plok = pippat rrat)

The actual sentences

1. Garcia and associates.
   Garcia y associados.
2. Carlos Garcias has three associates.
   Carlos Garcias tiene tres associados.
3. His associates are not strong.
   Sus associados no son fuertes.
4. Garcia has a company also.
5. Its clients are angry.
6. The associates are also angry.
7. The clients and the associates are enemies.

Statistical Machine Translation

- The fundamental idea of statistical MT is to let the computer learn how to do MT through studying the translation statistics from a bilingual corpus

What's the data? What are we doing?

- Pairs of sentences that are translations of one another are used
- Learn parameters for a probability model
- Source, Target pairs (S,T)

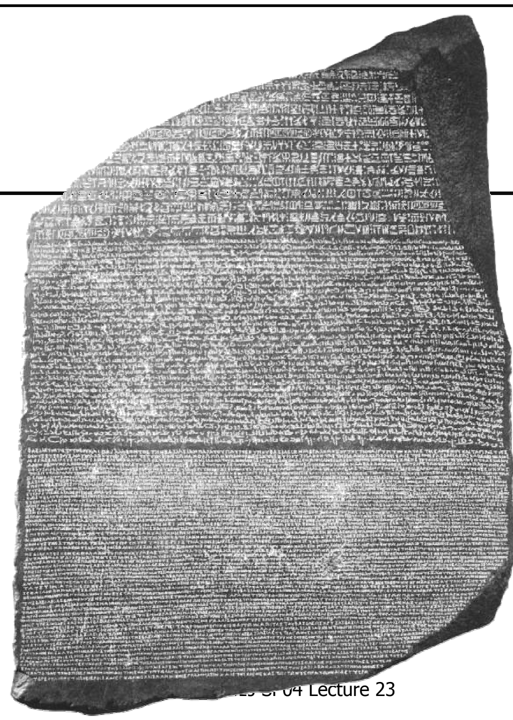
  Find pr distribution over (S,T)

Let’s see what this means

P(t|s) = P(t) x P(s|t)

Factor 1: Language Model

Factor 2: Translation Model



Statistical Machine Translation

- Warren Weaver (4 March 1947): (letter to Weiner)

Weaver, 1947

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'.

Example of what Weaver had in mind?

The proposal will not now be implemented

Les propositions ne seront pas mises en application maintenant

Example alignment

The proposal will not now be implemented

Les propositions ne seront pas mises en application maintenant

Honourable Members of the Senate,
Members of the House of Commons,
Ladies and Gentlemen:
Honorables sénateurs et sénatrices,
Mesdames et Messieurs les députés,
Mesdames et Messieurs,

My wife, Diana, and I were happy to welcome Her Majesty the Queen and the Duke of Edinburgh when they arrived in Canada last June and to be their hosts during their stay in the National Capital over Canada Day.
Ma femme et moi avons eu la joie d'accueillir Sa Majesté la Reine et le duc d'Édimbourg à leur arrivée au Canada en juin dernier et d'être leurs hôtes pendant leur séjour dans la région de la capitale nationale à l'occasion de la Fête du Canada.

As Governor General I have visited every province and territory, and I wish every Canadian could share that experience.
De plus, en tant que Gouverneur général, j'ai visité toutes les provinces ainsi que les territoires. C'est une expérience que je souhaite à tous les Canadiens.

We have to estimate these

- Training model from parallel aligned sentences (where do we get parallel texts; how do we align?)
- How much data needed?

So, how does English become French?

- Story 1. English gets converted to some sort of mental logic (predicate logic, or lexical-conceptual structures…), e.g., "I must not like ice-cream" into

  (obligatory (not (event like :obj ice-cream…))) blah blah blah

  Rest of story: how this gets mapped to French

Call this story interlingua

How does English become French?

- Story 2. English sentences gets syntactically parsed, into heads & modifiers, a binary tree say – phrases
- Then transformed into a French tree (a vine, say) – phrases swapped, english words replaced by french words.
- Call this syntactic transfer

How does English become French?

- Story 3. Words in English sentence replaced by French words, which are scrambled
- Zany!
- Heh: this is IBM Model 3 story

Like our alien system

- We will have two parts:

1. A bi-lingual dictionary that will tell us what e words go w/ what f words
2. A shake-n-bake idea of how the words might get scrambled around

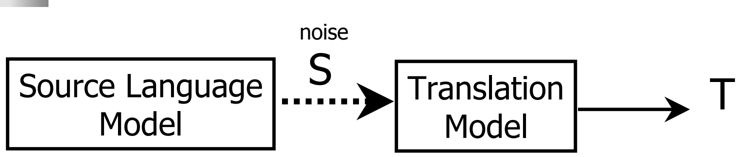
We get these from cycling between alignment & word translations – re-estimation loop on which words linked with which other words

What's the data? What are we doing?

- Pairs of sentences that are translations of one another are used
- Learn parameters for a probability model
- Source, Target pairs (S,T)

Find pr distribution over (S,T)

Noisy channel model to the rescue



Find pr distribution over (S,T)

Noisy channel model

The program has been implemented

Noisy communication channel

La programme a été mis en application

French to English decoding (translation)

The program has been implemented

George Bush Model of translation

- Somewhere in the noisy channel between speaker's brain and mouth, the English sentence E got "corrupted" to its French translation F
- Crazy?
- No stranger than the view that an English sentence gets corrupted into an acoustic signal in passing from the person's brain to his mouth

We need to estimate pr's

- Need to know:
  - What people say in English (source)
  - How E gets turned into French (channel)
- What we see is F
- What we want to find is E

(this is like speech…!)

How do we do this?

- English sentence e, French sentence f
- An English sentence e can be translated to *any* French sentence f
- But some translations are more equal than others… (more likely)
- We use probabilities to measure this!

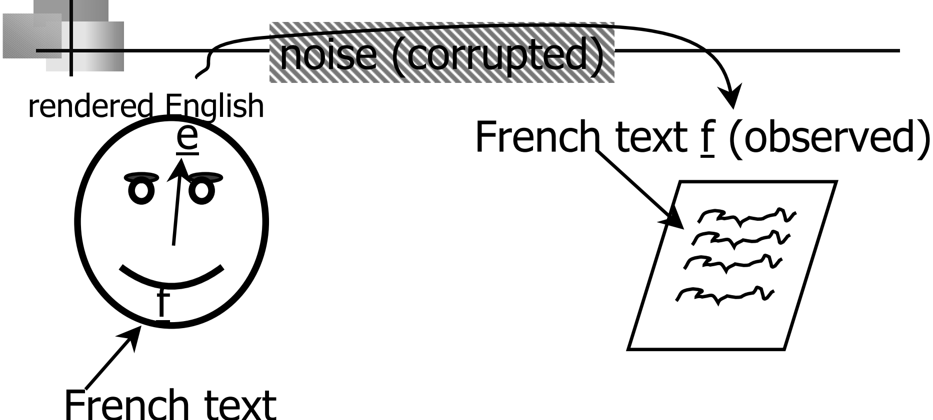
OK, to begin

- P(e)= pr of producing some English sentence e (e.g., “cheese-eating surrender monkeys”)
- P(e|f) = pr on encountering f, will produce e
- E.g., f= “Lincoln était un bon avocat”
  e= “cheese-eating surrender monkeys”

P(e|f) Not bloody likely!

Note: in general, e and f can be anything, not just words…

‘George Bush’ model of translation (noisy channel)



To be fair: perhaps the Jean Kerrie model

IBM "Model 3"

- First to do this, late 80s: Brown et al, "The Mathematics of Statistical Machine Translation", Computational Linguistics, 1990 (orig 1988 conference) – "Candide"
- We'll follow that paper

How to estimate?

- Formalize alignment
- Formalize dictionary in terms of P(f|e)
- Formalize shake-n-bake in terms of P(e)
- Formalize re-estimation in terms of the EM Algorithm
  - Give initial estimate (uniform), then up pr's of some associations, lower others

Bake-off – how to evaluate?

Tricky: not like speech (why?)

- Proposed measures…
  - Round-trip – ok, not always. E.g., "why in the world" → Sp → English → "why in the world" but
  - The Spanish is porqué en el mundo (???)

1. Compare human & machines –
2. Categorize as same; equally good; different meaning; wrong; (='fluency'); ungrammatical (= 'adequacy')
3. Humans take test based on translated text…

IBM toujours…

ISSUED: Apr. 23, 1996

FILED: Oct. 28, 1993

US PATENT NUMBER: 5510981

SERIAL NUMBER: 144913

INTL. CLASS (Ed. 6): G06F 17/28;

U.S. CLASS: 364-419.02; 364-419.08; 364-419.16; 381-043;

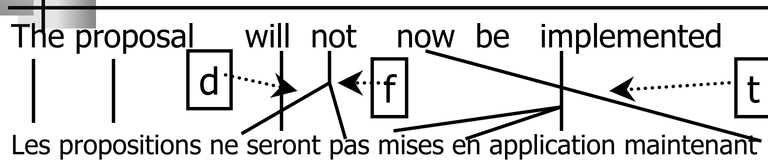
FIELD OF SEARCH: 364-419.02,419.08,419.16,200 MS File ; 381-43,51 ;

ABSTRACT: An apparatus for translating a series of source words in a first language…

The IBM series

- IBM1 – lexical probabilities only
- IBM2 – lexicon plus absolute position
- HMM – lexicon plus relative position
- IBM3 – plus fertilities
- IBM4 – inverted relative position alignment
- IBM5 – non-deficient version of model 4

Example alignment



4 parameters for P(f|e)

1. Word translation, t
2. Distortion (scrambling), d
3. Fertility, f

Spurious word toss-in, p

4 Parameters

- Word Translation, $t(f_j \mid e_i)$
- Distortion, scrambling, $d(a_j \mid j)$ $d(a_j \mid j\ m\ l)$
- Fertility, $phi(n \mid e_i)$
- Spurious word appearance, p_i
- Q: how much space?
- Other:
- Class-based alignment 50 classes
- Nondeficient alignments (nulls)

Bake-off Candide vs. Systran (Darpa) - 1995

	Fluency		Adequacy	
	1992	1993	1992	1993
Systran	47%	54%	69%	74%
Candide	51%	58%	58%	67%
Human		83%		84%

OK, now back to the game

How does English become French?

- Story 3. Words in English sentence replaced by French words, which are scrambled
- Zany!
- Heh: this is IBM Model 3 story

Decoupling by Bayes' Rule

- $P(e|f) = \dfrac{P(e) \times P(f|e)}{P(f)}$

- We want to *maximize* this quantity P(e|f), so we can simply maximize:

  $$P(e) \times P(f|e)$$

Q: What happened to P(f)?

A: the F sentence to translate is <u>fixed</u>

What's wrong with just comuting P(e|f) directly?

- We are extending from <u>words:</u>
  'sol' ↔ 'sun'
  'to pull the wool over someone's eyes' ↔ 'deitar areia para os olhos de alguém'

To <u>sentences:</u>

cheese eating surrender monkeys

fromage mangeant des singes de reddition

- What's wrong with this plan?
- Probably won't see a sentence match more than once, probably not at all!

In our case…

- What we see is f
- We want to find is e (the most likely translation e)
- In other words, compute:

$$\underset{e}{\text{argmax}}\ P(e|f)$$

What's wrong with this plan???

Why can't we just figure out P(e|f)?

So,

- If we compute P(e|f) directly, we had better be good – but there's no data….
- P(e|f) directly makes sense only if words in french are translations of words in english…
- A nice model for mutating bad french into bad english
- Note that it also gives no guarantee on the well-formedness of e!
- But: We can use Bayes' Rule to get good translations even if the pr estimates are crummy!

Why this order?

- If it seems backwards, it is
- Imagine you are building an English-French translator, but when you run it, you feed in French and ask, “what English would have caused this French sentence to come out”
- The right answer – a fluent English sentence (language model) that means what you think it means (translation model)

Why not just max p(e|f) directly?

- If we are translating french to english, then P(e|f) seems more intuitived
- For a given, fixed french sentence f, we would find the English sentence e that maximizes

$$P(e) \bullet P(e|f)$$

We solve this as $P(e|s) \approx P(e) \bullet P(f|e)$ [since french sentence is fixed, p(f) is fixed

- But why is figuring out P(f|e) any easier?

Why not p(e|f)

- Answer: P(f|e) does not have to give good french translations
- P(f|e) can assign lots of probability weight to bad french sentences, as long as they contain the right words
- P(f|e) can be sloppy because P(e) will worry about word order
- If we try to figure out P(e|f) directly we need to get good english translations, all in one step

Estimating P(f|e)

- Given a sentence pair, P(f|e) is simply the product of the word translation probabilities between them irrespective of word order

Cheap and dirty P(s|e)

- Just product of individual translation probabilities!
- P(yo no comprendo | I don't understand)=

  P(yo | I) x
  P(yo | don't) x
  P(yo | understand) x
  P(no | I) x
  P(no | don't) x
  P(no | understand) x
  P(comprendo | I) x
  P(comprendo | don't) x
  P(comprendo | understand)

Bilingual corpus

- These can be estimated from a bilingual corpus: just retrieve all sentence pairs containing the word 'understand', count how many times 'comprendo' occurred, divide by total # of words in Spanish half of corpus
- Problems:
- P(comprendo | understand) will be too low (even if comprendo appears every time understand does, it's normalized by # words – so say .05)
- Worse: P(la | understand) too high, because 'la' is frequent, you'll often see it with 'comprendo' (remember, word order doesn't matter)

What is the fix?

- Decipherment
- ‘Understand’ might co-occur with both ‘la’ and ‘comprendo’ but if we have a <u>previous</u> link between ‘the’ and ‘la’, then we should weight towards ‘comprendo’.
- In turn, that would reduce translating ‘don’t’ as ‘comprendo’, because understand and con’t will co-occur
- After decipherment, P(comprendo | understand) should be close to 1; P(la | the), 0.4, with rest of wt there going to P(el the),
- Need to <u>re-estimate</u> – can’t keep aassuming previously established links (bootstrap - EM

Estimation maximization (EM)

- Key: word alignments
- Word alignment connects words in sentence pair s.t. each English word produces 0 or more French words, and each French word is connected to exactly one English word
- Longer sentence → more alignments possible
- Some are more reasonable than others, because they have more reasonable word translations
- Here is our revised approximation of P(f|e) or for spanish, P(s|e):

EM iteration for alignment

- Step 1: Assume all alignments for a given sentence pair equally likely (e.g., one S has 256 alignments, 1/256; another, 1 million)
- Step 2: Count up word pair connections in all alignments in all word pairs, weighted by the pr of the alignment in which it occurs (so short, less ambiguous sentences have more weight)
- Step 3: consider each English word in turn ('understand')

So, if this works…

- Our job has been reduced to three things:

1. Estimate the parameters for $P(e)$
2. Estimate the parameters for $P(f|e)$
3. Search the product space to maximize

Let's see what each of the pr quantities mean, and what role they play

Let's see what this means

$$P(e|f) = \underbrace{P(e)}_{\boxed{\text{Factor 1: Language Model}}} \times \underbrace{P(f|e)}_{\boxed{\text{Factor 2: Translation Model}}}$$

Factor 1: P(e), language model

- P(e) says that ‘John ate ice-cream’ has high pr, but ‘ate ice-cream John’ has lower pr
- Indeed, ungrammatical sentence – pr 0 (but this could be hard to figure out)
- P(e) is really <u>lowering</u> pr of ungrammatical S’s
- So really, this is like our alien language case (what part?)
- Several possible collections of words (‘bags’) – pick most probable sequence

Language model P(e)

- So in fact, we have to choose between many grammatical sentences, e.g.,
- Which of these is better translation?

  Fred viewed Sting in the television

  Fred saw Sting on TV

- So, we are back to N-grams again!
- This will let us model <u>word order</u>

Language model & N-grams

- In general – next word could depend on all preceding context
- But there are too many parameters to estimate, so we use just bigrams or trigrams
- To find pr for a whole sentence, multiply conditional pr's of the n-grams it contains

Language and N-gram example

- P(I found riches in my backyard)=
  P(I | start-of-sentence) x
  P(found | I) x
  P(riches | found) x
  P(in | riches) x
  P(my | in) x
  P(backyard | my) x
  P(end-of-sentence | backyard)
- Note how this will be higher pr than "my I in riches backyard found" – just as we want

Language model P(e)

- So, if this does word order…
- Question: restore order for

  actual the hashing is since not collision-free usually the is less perfectly the of somewhat capacity table
- Question: what knowledge are you using?
- Amazingly, this alone can be used to restore scrambled English sentences (63-80%)
- Question: restore order for

  loves John Mary

A final use of P(e)

- Choose between alternative translations

  I found the riches in my backyard

  I found the riches on my backyard

In Spanish, 'in' and 'on' correspond to 'en'

We can use trigram counts to tell the difference and select the higher pr one…

Problemes? Problemos?

The estimation catch

- Where do these pr numbers come from?
  Which has higher pr:
  'I hate ice-cream', or 'I like ice-cream'?
  use Google!
- What happens when P(y | x) is zero? (not observed in training)
- The whole product would be zero
- Bad, because then "I like cheese eating monkeys" = same pr as "like I monkeys cheese eating"

Estimation

- Acute issue for trigrams - `found riches in’ probably never seen
- Solution: smoothing (see textbook & next lecture - large literature on this)

Problems…

- Won’t always work – consider
  - Underline it
  - Emphasize it
- English might prefer the first, but must look at Spanish – ‘subrayar’ translates as both, but mostly as ‘underline’; Spanish uses ‘accentuar’ for emphasis
- But this means we need to look at <u>connections</u> between 2 languages, ie, P(f|e) that bridge between them, not <u>just</u> in English… that is the job of the <u>Translation Model</u>

Language model & translation model

- Factoring knowledge out this way makes estimation easier
- Since P(e) takes care of word order, the translation model, P(f|e) doesn't have to worry about this – it can give crummy pr estimates, it can be sloppy, as long as it has the right words
- But as we've seen, P(e) can't do all the work for this…

Translation model P(f|e)

- What was it in our alien example?
- It was the bilingual dictionary
- What does it do?
- Ensure the words of e express the ideas of f
- So, responsibility is divided between P(e) and P(f|e)

An example (Spanish)

Spanish-English

- P(e) x P(s|e) to get P(e|s) – assume 'subrayar' input…

1. <u>Underline it.</u>
   P(underline) x
   P(it | underline) x
   P(subrayar | underline)
2. <u>Emphasize it.</u>
   P(emphasize) x
   P(it | emphasize) x
   P(subrayar | emphasize)

- (1) is preferred because 'underline' is common *<u>and</u>* it is usually translated as 'subrayar'

Language model can give crummy pr's

- As long as it has the right words
- This gives some measure of robustness
- Example – all of these could have roughly the same pr, despite being lousy translations…

Lousy translations

- P(Yo no comprendo|I don't understand)
- P(Comprendo yo no | Don't understand I)
- P(No yo comprendo | I don't understand)
- P(Comprendo yo no | I don't understand)
- P(Yo no comprendo | I understand don't)
- P(Yo no comprendo | Understand I don't)

- In fact, this gives a first-cut way to estimate P(f|e)! Do you see how?

Cheap and dirty P(s|e)

- Just product of individual translation probabilities!
- P(yo no comprendo | I don't understand)=

  P(yo | I) x
  P(yo | don't) x
  P(yo | understand) x
  P(no | I) x
  P(no | don't) x
  P(no | understand) x
  P(comprendo | I) x
  P(comprendo | don't) x
  P(comprendo | understand)

Any problemos?

- Si…
- P(comprendo | understand) will be too low
- P(la | understand) will be too high – just because la is frequent in Spanish
- Use our method for alien languages!
- If we have previously established a link between 'the' and 'la', then we should boost 'comprendo'
- That will reduce translation of 'don't' as 'comprendo' because that will co-occur only when 'understand' is already nearby
- P(comprendo | understand) should work out close to 1, and P(la | the) say 0.4, rest going to P(el | the)…

In other words…

- Use alignments to assist with P(e), P(f|e)

- Use P(e) to assist with alignments

problème de poulet et d'oeufs et
problema del pollo y del huevo y
problema dell'uovo e del pollo e
Huhn- und Eiproblem und

Problemos

- Alignments help us get the translations
- Translations help us get the alignments…

- Where do we start???

For the example…

- Yo no comprendo / I don't understand
- There are six possible alignments (for now…assuming no null maps, etc)
- All possible word combinations…

- This is just like our alien language case

Like this…

- P(yo no comprendo | I don't understand)=
  P(Alignment1) x P(yo | I) x P(no | don't) x
  P(comprendo | understand)

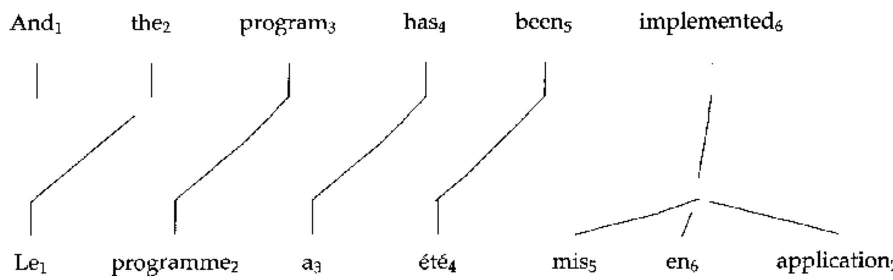
+P(Alignment2) x P(yo | don't) x P(no | I) x
P(comprendo | understand)

+P(Alignment3) x P(yo | understand) x P(no | I) x

…

+P(Alignment6) x P(yo | understand) x P(no | don't)
P(comprendo | I)

Example



More problemos…

- Can't assume direct word-for-word translation – some sentence pairs are different lengths
- An English word might correspond to more than one French word, or none at all
- So we model this -

Procrustean bed

- For each word e_i in the sentence, $i= 1, 2, \ldots, l$) we choose a fertility $\phi(e_i)$, equal to 0, 1, 2,…
- This value is dependent solely on the English word, not other words or the sentence, or the other fertilities
- For each word e_i we generate $\phi(e_i)$ French words – not dependent on English context
- The French words are permuted ('distorted') – assigned a position slot (this is the scrambling phase)
- Call this a distortion parameter $d(i|j)$

Summary of components

- The language model: $P(e)$
- The translation model for $P(f|e)$
  - Word translation t
  - Distortion (scrambling) d
  - Fertility ϕ
- (really evil): (for next time)
- Maximize (A* search) through product space

What's the input data? Aligned S's

The high turnover rate was largely due to an increase in the sales volume.
Employment and investment levels have also climbed.
Following a two-year transitional period, the new Foodstuffs Ordinance for Mineral Water came into effect on April 1, 1988.
Specifically, it contains more stringent requirements regarding quality consistency and purity guarantees.

La progression des chiffres d'affaires résulte en grande partie de l'acroissement du volume des ventes.
L'emploi et les investissements ont également augmenté.
La nouvelle ordonnance fédérale sur les denrées alimentaires concernant entre autre les eaux minérales, entrée en vigueur le 1er avril 1988 après une période transitoire de deux ans, exige surtout une plus grande constance dans la qualité et une garantie de la pureté.

What's the data?

- Hansard – Canadian Parliament since early 1800s, dual language
- 100M words, > 1M sentences
- Each on separate tape (!)
- Corresponding sentences not marked, paragraphs missing
- We want this – how to we get to it?

Issues with alignment

- Clues include…
  - French sentences usually in same order as English sentences (but word order difft)
  - Short French sentences $\leftrightarrow$ short English sentences, and v.v.
  - Corresponding French and English sentences often contain many of the same character sequences (why?)

Weaver knew…

Think, by analogy, of individuals living in a series of tall closed towers,
all erected over a common foundation. When they try to communicate with one another,
they shout back and forth, each from his own closed tower. It is difficult to make the sound
penetrate even the nearest towers, and communication proceeds very poorly indeed.

But, when an individual goes down his tower, he finds himself in a great open basement,
common to all the towers. Here he establishes easy and useful communication with the persons
who have also descended from their towers.

Thus it may be true that the way to translate from Chinese to Arabic, or from Russian to
Portuguese, is not to attempt the direct route, shouting from tower to tower. Perhaps the way
is to descend, from each language, down to the common base of human communication—
the real but as yet undiscovered universal language—and—then re-emerge by whatever particular
route is convenient.

Lost in the translation

- Proust: "A la recherche du temps perdu: "Depuis longtemps…
- Translation: "For a long time I would go to bed early…"
- Last word in book: depuis