CS 188: Artificial Intelligence
Spring 2006

Lecture 28: Machine Translation
5/2/2006

Dan Klein – UC Berkeley

Machine Translation: Examples

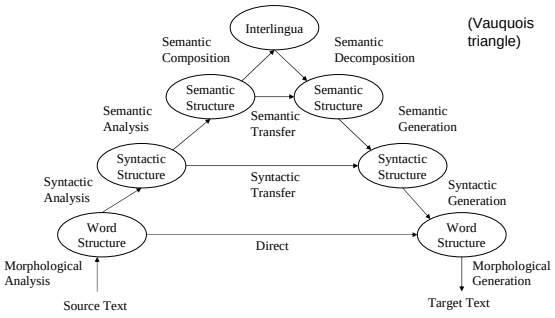
Atlanta, preso il killer del palazzo di Giustizia

ATLANTA - La grande paura che per 26 ore ha attanagliato Atlanta è finita: Brian Nichols, l'uomo che aveva ucciso tre persone a palazzo di Giustizia e che [illegible], s'è consegnato alla polizia, dopo avere cercato rifugio nell'alloggio di una donna in un complesso d'appartamenti alla periferia della città. Per tutto il giorno, il centro della città, sede della [illegible] dei Giochi 1996, cuore di una popolosa area metropolitana, era rimasto paralizzato.

Atlanta, taken the killer of the palace of Justice

ATLANTA - The great fear that for 26 hours has gripped Atlanta is ended: Brian Nichols, the man who had killed three persons to palace of Justice and that [illegible] s' is delivered to the police, after to have tried shelter in the lodging of one woman in a complex of apartments to the periphery of the city. For all the day, the center of the city, center of the [illegible] and of Giochi 1996, heart of one popolosa metropolitan area, was remained paralyzed.

Levels of Transfer



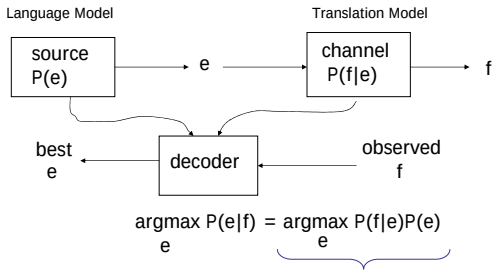
General Approaches

- Rule based approaches
  - Expert system style rewrite systems
  - Interlingua methods (analyze and generate)
  - Lexicons come from humans or dictionaries
  - Can be very fast, and can accumulate a lot of knowledge over time (e.g. Systran)
- Statistical approaches
  - Noisy channel systems
  - Lower-level transfer
  - Lexicons discovered using parallel corpora
  - Require little human declaration of knowledge

The Coding View

- "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 (1955:18, quoting a letter he wrote in 1947)

MT System Components



Finds an English translation which is both fluent and semantically faithful to the French source

Language Models

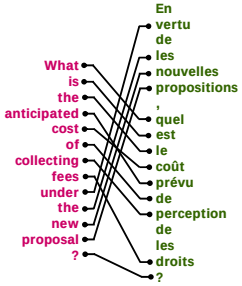
- Language Models
  - Any probabilistic model capable of assigning probabilities to sentences
  - Usually n-gram models, but also PCFGs
  - Exact same technology (and software) as in ASR
  - Train on a huge collection of monolingual corpora (documents in the target language)

Parallel Corpora

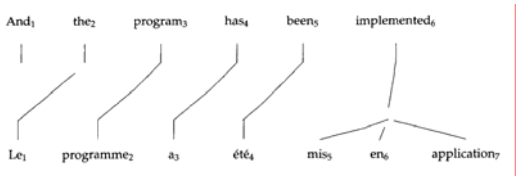
- Parallel corpora (or bitexts)
  - Collection of source-target translation pairs
  - Main resource for learning a translation model
  - Either naturally occurring (e.g. parliamentary proceedings, news translation services) or commissioned

Building a Translation Model

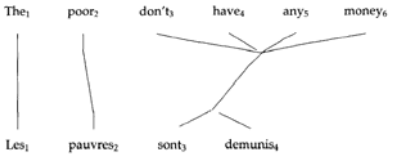
- Steps in building a simple statistical translation model
  - Match up words in training sentence pairs (word alignment)
  - Learn a lexicon from these alignments
  - Learn larger phrases



1-to-Many Alignments

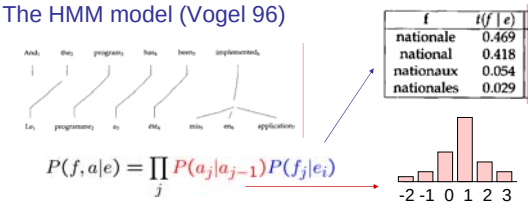


Many-to-Many Alignments



The HMM Alignment Model

- The HMM model (Vogel 96)



  - Re-estimate using the forward-backward algorithm
  - Handling nulls requires some care
- Note: alignments are not provided, but induced

Examples: Translation and Fertility

the

f	$t(f \mid e)$	ϕ	$n(\phi \mid e)$
le	0.497	1	0.746
la	0.207	0	0.254
les	0.155		
l'	0.086		
ce	0.018		
cette	0.011		

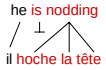
not

f	$t(f \mid e)$	ϕ	$n(\phi \mid e)$
ne	0.497	2	0.735
pas	0.442	0	0.154
non	0.029	1	0.107
rien	0.011		

farmers

f	$t(f \mid e)$	ϕ	$n(\phi \mid e)$
agriculteurs	0.442	2	0.731
les	0.418	1	0.228
cultivateurs	0.046	0	0.039
producteurs	0.021		

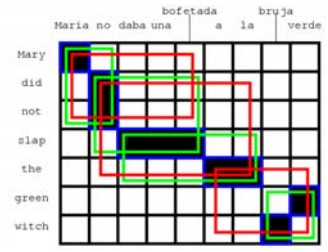
Phrases vs Word Models



nodding

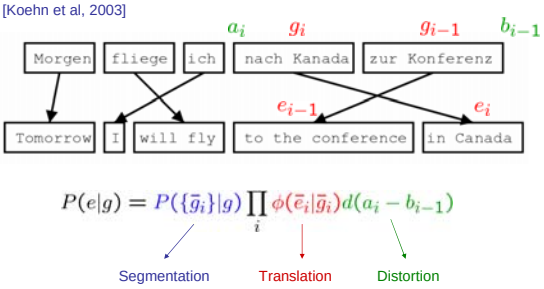
f	$t(f \mid e)$	ϕ	$n(\phi \mid e)$
signe	0.164	4	0.342
la	0.123	3	0.293
tête	0.097	2	0.167
oui	0.086	1	0.163
fait	0.073	0	0.023
que	0.073		
hoche	0.054		
hocher	0.048		
faire	0.030		
me	0.024		
approuve	0.019		
qui	0.019		
un	0.012		
faites	0.011		

Extracting Phrases



Basic Phrase-Based Model

[Koehn et al, 2003]



$$P(e|g) = P(\{\bar{g}_i\}|g) \prod_i \phi(\bar{e}_i|\bar{g}_i) d(a_i - b_{i-1})$$

Segmentation Translation Distortion

Decoding

- Now we have a phrase table:
  - A huge list of translation phrases (e.g. 1M phrases)
  - Each phrase has a probability P(f|e)
- When we see a new input sentence:
  - Grow a translation left to right
  - Extend translation using known phrases
  - Also multiply by language model score

The Pharaoh Decoder

- Probabilities at each step include LM and TM

Some Output

Madame la présidente, votre présidence de cette institution a été marquante.
Mrs Fontaine, your presidency of this institution has been outstanding.
Madam President, president of this house has been discoveries.
Madam President, your presidency of this institution has been impressive.

Je vais maintenant m'exprimer brièvement en irlandais.
I shall now speak briefly in Irish .
I will now speak briefly in Ireland .
I will now speak briefly in Irish .

Nous trouvons en vous un président tel que nous le souhaitions.
We think that you are the type of president that we want.
We are in you a president as the wanted.
We are in you a president as we the wanted.

Translations

- Even human translators aren't perfect:
  - In an Austrian ski hotel:
    Not to perambulate the corridors in the hours of repose in the boots of ascension.
  - In a Copenhagen airline ticket office:
    We take your bags and send them in all directions.
  - From a brochure of a car rental firm in Tokyo: When passenger of foot heave in sight, tootle the horn. Trumpet him melodiously at first, but if he still obstacles your passage then tootle him with vigor.

http://www.englishfirst.org/13166/funnytranslations.htm