TALP: Xgram-based Spoken Language Translation System

Adrià de Gispert
José B. Mariño

Outline

- Overview
- Translation generation
- Training
- IWSLT'04 Chinese-English supplied task results
- Conclusion and further work

Overview

- TALP Statistical Machine Translation (SMT)
- Integrated speech-text approach
- Finite-State Transducer (FST) implementation
- Automatically learnt from parallel corpus
- Bilingual units called tuples

Translation generation

- Maximising joint probability
- Variable-length N-gram of bilingual units (tuples)

$$\hat{f} = \underset{f}{argmax}\, p(e, f)$$

X-gram

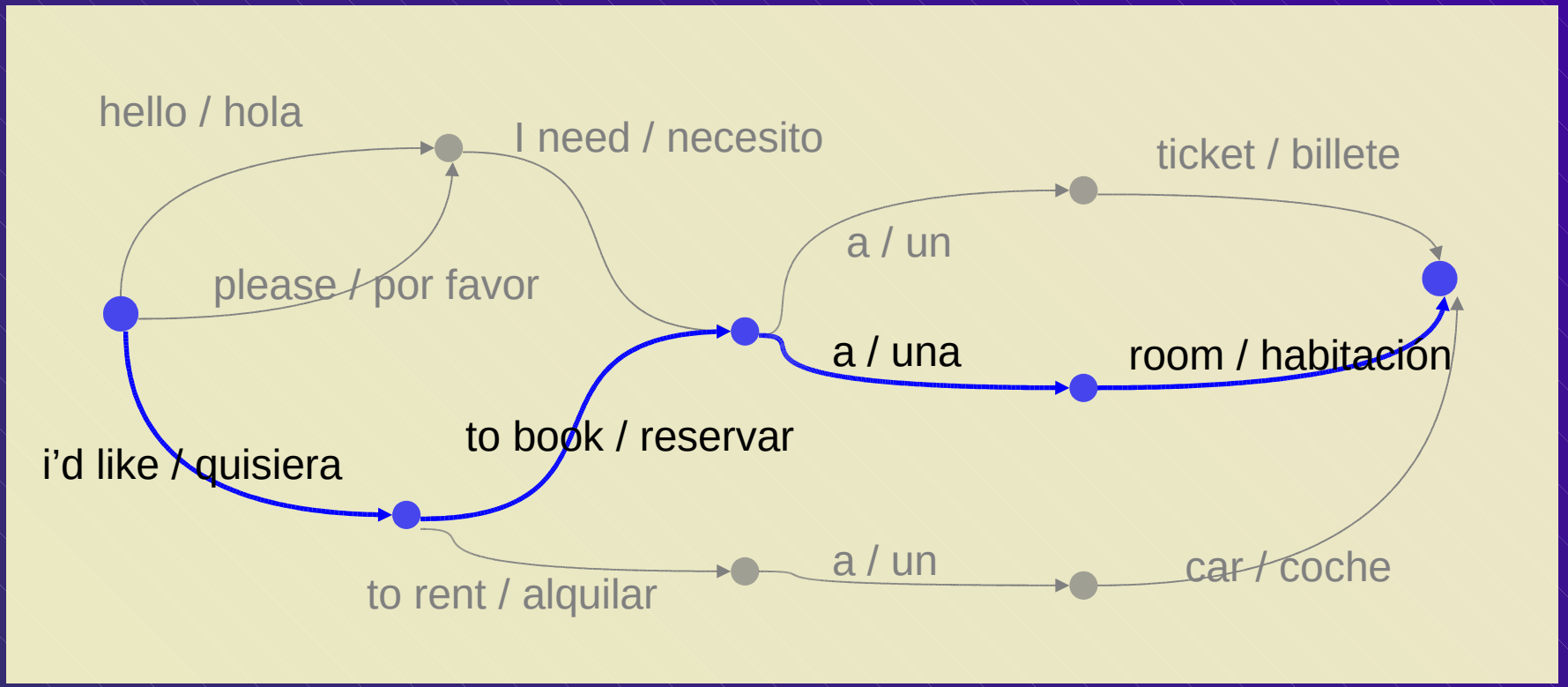
$$p(e, f) = \prod_{n=1}^{N} p((e, f)_n | (e, f)_{n-1} \ldots (e, f)_{n-X+1})$$

tuple

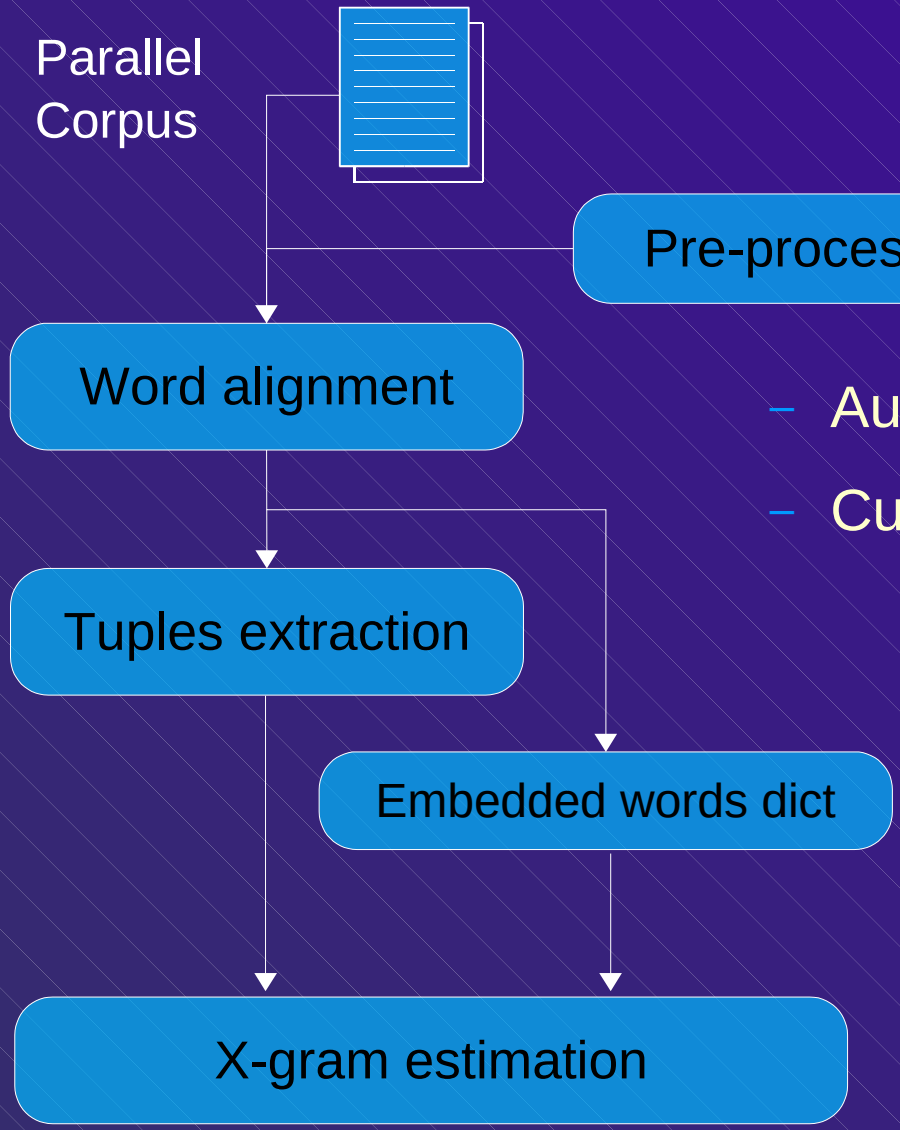
$$(e, f)_n = (e_{i(n)} \ldots e_{i(n)+I(n)}, f_{j(n)} \ldots f_{j(n)+J(n)})$$

FST implementation

- Search for best-scoring path
- Speech translation: include acoustic models



Training



- Automatic from parallel corpus
- Currently based on Giza++ alignments

Preprocessing

- Particular for each pair of languages
- Categorisation *personal names, dates, times, numbers, ...*

- **Chinese-English** IWSLT'04 supplied track:
  - Clearing out punctuation *no gain*
  - Segmentation of longer sentences

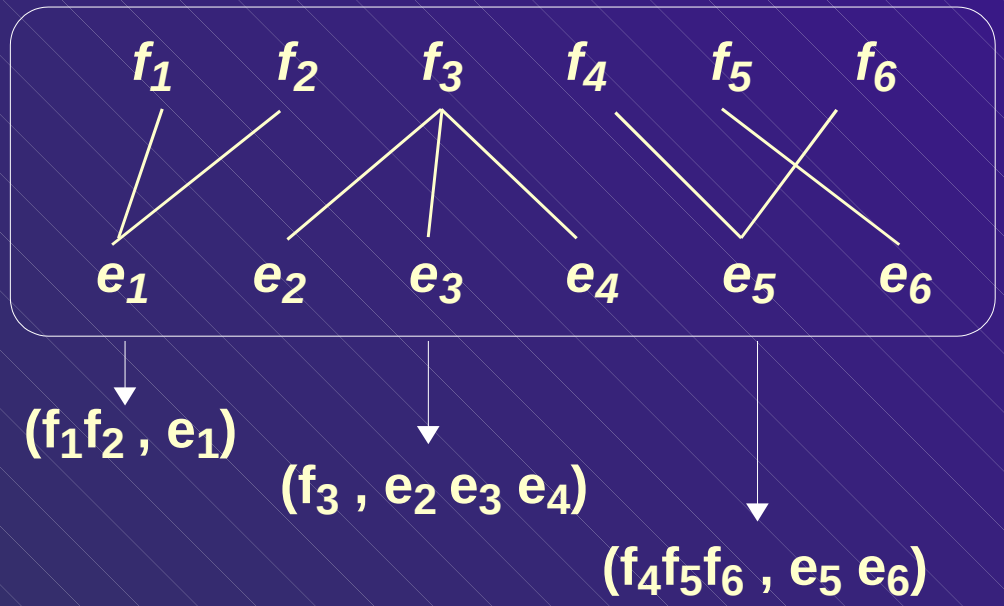
	# sentences	Lavg
Chinese	20 K (22.2 K)	9.1 (8.2)
English		9.4 (8.5)

Word alignment

- Standard GIZA++ alignments $1^5H^5 3^3 4^3$
  - Source-to-Target (s2t)
  - Target-to-Source (t2s)
  - Union
  - Intersection
- Tuples can be extracted from any alignment
- Usually **union** and **s2t** are used

Tuples extraction

- Tuples are bilingual units containing
  - one or more source words
  - zero, one or more target words
- Subset of phrases, unique under following conditions
- Example

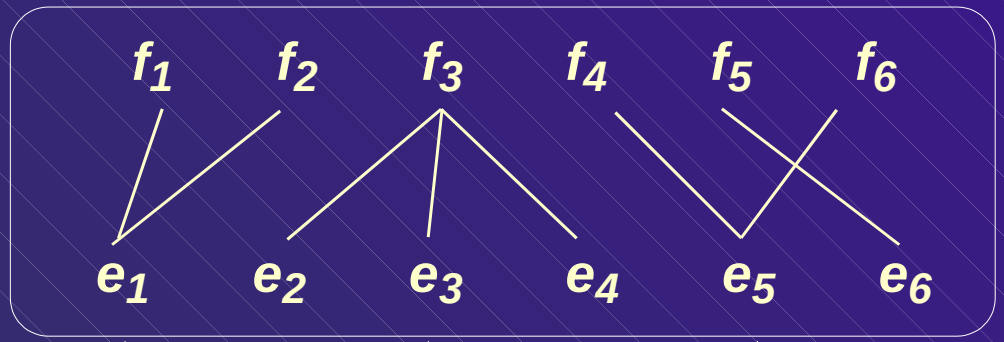


1. Monotonous segmentation of the pair
2. Words are consecutive along both source and target
3. No word in the tuple is aligned to a word outside the tuple
4. The tuple cannot be decomposed without violating 1-3

Embedded words dictionary

- Embedded words
  - translation appears always inside a tuple (never isolated)
  - an 'accurate' dictionary is built *intersection*
- New unigrams (history independent)

DICTIONARY ENTRIES:



$(f_1 f_2 , e_1)$

$(f_3 , e_2\ e_3\ e_4)$

$(f_4 f_5 f_6 , e_5\ e_6)$

Given a source word, look for the most freq. aligned words

1. Target words are consecutive
2. Target words are aligned *only* to the source word

→ (f_5 , e_6)

X-gram estimation

$ ($f_1 f_2$, e_1) (f_3 , e_2 e_3 e_4) ($f_4 f_5 f_6$, e_5 e_6) $

$$p(e, f) = \prod_{n=1}^{N} p(T_n | T_{n-1} \ldots T_{n-X+1})$$

- Usually, maximum memory is 3
- Pruning strategies
  - Min. number of times a certain history must occur *k*
  - threshold of divergence between output prob. distributions for two nodes sharing recent history *f*

k is untouched (= 1)
f used for slight pruning (f = 0.2)

Chinese-English supplied track

- **Results on development** set
  - Union (aU) vs. s2t alignment (a2)
  - Normal vs. segmented corpus (seg)
  - Normal vs. FST pruning (f)

run	BLEU	NIST	WER	PER	GTM	E
aU	0.244	5.169	0.615	0.529	0.591	7
aU,seg	0.251	5.187	0.607	0.521	**0.595**	**7**
aU,seg,f	0.255	**5.210**	**0.603**	**0.518**	0.594	7
a2	**0.319**	3.789	0.614	0.552	0.573	16
a2,seg	0.318	3.871	0.606	0.546	0.573	18
a2,seg,f	0.314	3.678	0.607	0.548	0.570	19

→ ***run A***

→ ***run B***

no embed. dictionary

aU,seg,-D	0.264	4.741	0.606	0.524	0.592	7
a2,seg,-D	0.315	3.706	0.607	0.547	0.571	19

Automatic evaluation results

- Statistics of submitted runs

run	tuples	vcb	length	embed
aU,seg,f	97 K	27 K	3.9	4.7 K
a2,seg,f	140 K	29 K	2.9	1.5 K

  - Longer tuples with Union (more embedded)
  - Many tuples to NULL with s2t (28% over total, 7.5% union)

- **Results**

run	BLEU	NIST	WER	PER	GTM	E
aU,seg,f	0.279	**6.778**	**0.556**	**0.465**	**0.647**	**5**
a2,seg,f	**0.331**	5.391	**0.550**	0.490	0.620	11

  - Contrast BLEU vs. NIST (related to length)

Manual evaluation results

- Results

run	fluency	adequacy
aU,seg,f	2.792	3.022

- Expected fluency deficiency *no explicit long reordering*
- 'Much of the information' is transmitted

- Examples:

Translation: *that what time start*
Reference1: *what time does it start*

Translation: *stomach very hurts*
Reference1: *i have a severe pain in my stomach*

Conclusion and further research

- Tuple-based FST translation system presented
- Adequate for pairs of languages *similar in word-order*

- Further research
  - Embedded N-grams
  - Generalization of tuples
  - Explicit reordering techniques

Thanks for attention

Centre de Tecnologies i Aplicacions del Llenguatge i la Parla
TALP Research Center
Universitat Politècnica de Catalunya (UPC)
Barcelona