Use of linguistic information and reordering stratregies for Ngram-based SMT

Adrià de Gispert

Universitat Politècnica de Catalunya (UPC)

October 30th 2006
Seminar on SMT at University of Cambridge

UPC
Departament de Teoria del Senyal i Comunicacions
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Outline

1. **N-gram-based SMT**
   - Bilingual N-gram translation model
   - Additional feature functions
   - Training and decoding
   - Results and discussion
2. Introducing linguistic information
3. Word reordering strategies
4. Further research lines

Ngram-based SMT system

Log-linear combination of feature functions:

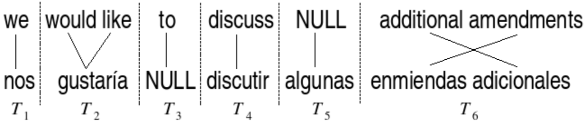
$$\hat{t}_1^I = \arg\max_{t_1^I} \left\{ \sum_{m=1}^{M} \lambda_m h_m(s_1^J, t_1^I) \right\}$$

- **Bilingual N-gram translation language model**
- target language model
- word bonus model
- source→target lexicon model (IBM1 probs.)
- target→source lexicon model (IBM1 probs.)
- target POS language model

Bilingual N-gram translation model

$$h_{BM}(s_1^J, t_1^I) = \log \prod_{i=1}^{K} p((s,t)_i | (s,t)_{i-N+1}, ..., (s,t)_{i-1})$$

- Given a word alignment, *tuples* $T_i = (s,t)_i$ are those bilingual units:
  - ▷ having minimal length
  - ▷ describing a monotonic segmentation of each sentence pair

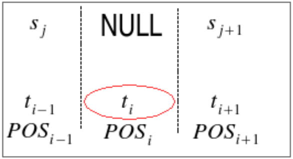


- Constraints define a unique possible segmentation

⇒ Except: NULL-source tuples

NULL-source tuples

- These units cannot be allowed (no word insertion model in decoding)
- Decision criteria:
  - ▷ always to one side (deterministically)
  - ▷ based on IBM model 1 word-to-word probabilities
  - ▷ based on **target POS entropy distributions**:



$$p^f_{POS} = \frac{N(t_{i-1}, t_i, POS_{i+1})}{\sum_{POS'} N(t_{i-1}, t_i, POS'_{i+1})}$$

$$p^b_{POS} = \frac{N(POS_{i-1}, t_i, t_{i+1})}{\sum_{POS'} N(POS'_{i-1}, t_i, t_{i+1})}$$

$\Rightarrow$Choose most entropic case

Feature functions

- target language model:

$$h_{TM}(s,t) = h_{TM}(t) = \log \prod_{k=1}^{K} p(w_k | w_{k-N+1}, \ldots, w_{k-1})$$

- word bonus model: $h_{WB}(s,t) = h_{WB}(t) = K$
- source→target lexicon model:

$$h_{LEX}(s,t) = \log \frac{1}{(I+1)^J} \prod_{j=1}^{J} \sum_{i=0}^{I} p_{IBM1}(t_j^n | s_i^n)$$

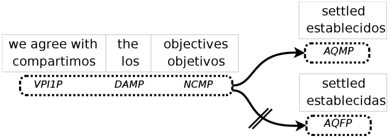
- target→source lexicon model (analogous)

Feature functions (2)

- target POS language model:

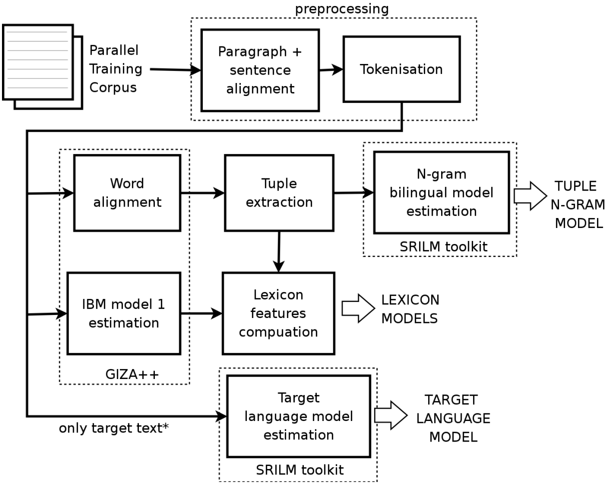
$$h_{TPM}(s,t) = \log \prod_{k=1}^{K} p(g_k | g_{k-N+1}, \ldots, g_{k-1})$$

- each tuple carries a *single* target POS sequence, $T_i = (s,t,G)_i$
  - ▷ no need for tagging in decoding time
  - ▷ not used for bilingual model estimation
  - ▷ qualitative improvement on manual inspection



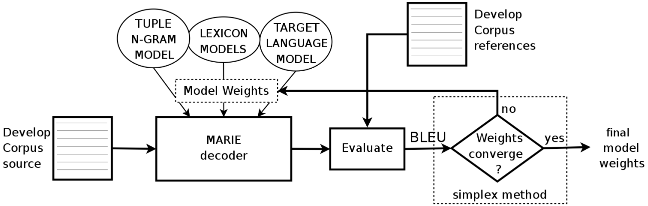
		BLEU	mWER
Sp→En	base	0.5412	34.98
	+POSLM	**0.5461**	**34.47**
En→Sp	base	0.4714	**40.22**
	+POSLM	**0.4750**	40.42

Global training scheme



Ngram-based Decoder

- MARIE version 1.3 (Crego et al., 2005)
  - ▷ beam-search strategy based on dynamic programming
  - ▷ threshold and histogram pruning for efficiency
  - ▷ admits both Ngram-based and phrase-based models
  - ▷ freely-available C++ source code



Some results

- TC-STAR project 1st evaluation (Fall 2004)
- Spanish↔English European Parliament data: (Ney et at., 2005)
  - train: ∼1.2M sents., ∼35M words per lang.
  - edited text, transcript (verbatim) and ASR output
  - dev/test: 1k sents., 2 human references

	Spanish→English			English→Spanish		
	site	BLEU	NIST	site	BLEU	NIST
ASR (1-best)	RWTH	41.5	9.12	RWTH	38.7	8.73
	IBM	39.7	8.81	IBM	34.3	8.13
	UPC	**37.7**	**8.56**	**UPC**	**33.8**	**8.00**
	ITC-irst	34.7	7.97	UKA	33.0	7.94
	UKA	32.3	7.85	UPV	19.1	5.46
	UPV	16.0	4.35			
Text	**UPC**	**53.3**	**10.55**	**UPC**	**46.2**	**9.65**
	IBM	53.1	10.38	IBM	45.2	9.44
	ITC-irst	47.5	9.60	RWTH'	38.9	8.72
	RWTH'	46.1	9.68	UKA	37.6	8.46
	UKA	40.5	8.96	UPV	34.1	7.51
	UPV	32.7	6.80			

Some results (2)

- MT shared task at ACL 2005 workshop on Building and using Parallel Texts
- European Parliament data: (Koehn and Monz, 2005)
  - train: ∼700k sents., ∼15M English words
  - dev/test: 2k sents., 1 human reference

Spanish→English		French→English		Finnish→English		German→English	
site	**BLEU**	site	**BLEU**	site	**BLEU**	site	**BLEU**
UW	30.95	UW	30.27	UW	22.01	UW	24.77
UPC	**30.07**	**UPC**	**30.20**	NRC	20.95	**UPC**	**24.26**
UPCm	29.84	NRC	29.53	**UPC**	**20.31**	NRC	23.21
NRC	29.08	RALI	28.89	RALI	18.87	RALI	22.91
RALI	28.49	CMUb	27.65	SAAR	16.76	SAAR	20.48
UPCj	28.13	CMUj	26.71	UJI	13.79	CMUj	18.93
SAAR	26.69	SAAR	26.29	CMUj	12.66	UJI	18.89
CMUj	26.14	GLGW	23.01				
UJI	21.65	UJI	21.25				

Discussion

- captures effectively bilingual context, achieving state-of-the-art results
- yet Spa↔Eng error analysis reveals:
  - ▷ strong difficulty in verb form correct production
  - ▷ need for local word reorder strategies
  - ▷ *silly* morphology disagreement errors
- non-monotone language pairs?
  - ▷ word reordering strategies (later on)
- target length problem (tuples to NULL)
  - ▷ tends to produce less words than input
- optimisation criterion (**BLEU**, **NIST**, **QARLA**, combination?)
- output graph/N-best and re-ranking

1. N-gram-based SMT

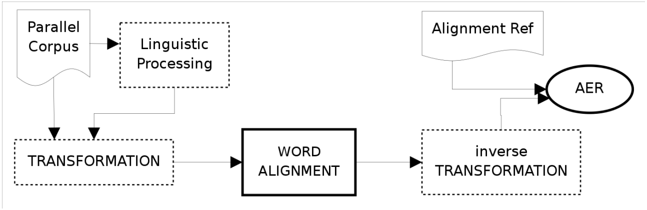
2. **Introducing linguistic information**
   - Word alignment
   - Verb form classification model
   - Study of morphology reduction

3. Word reordering strategies

4. Further research lines

Word alignment improvement

- evaluation on Precision, Recall and Alignment Error Rate (**AER**)
- cooccurrence-based alignment model (Cherry and Lin., 2003)
  - ▷ link probability iterative search algorithm
  - ▷ high-precision but one-to-one solution
  - ⇒ extended with verb form classification
- morphology transformations prior to GIZA-based alignment
  - ▷ base form, stems, full verb form, reduced Part-Of-Speech, etc.
  - ▷ correlation with MT results



Morphology transformations for word alignment

- Spanish–English Europarl corpus
  - train: ~1.2M sents. AND ~12k sents.
  - alignment reference: 400 sents., 3 human consensus

English	Asian countries have followed our example too .
base forms	asian country have follow our example too .
stems	asian countri have follow our exampl too .
full verbs	asian countries V[follow] our example too .

Spanish	Los países asiáticos han seguido también nuestro ejemplo .
base forms	el país asiático haber seguir también nuestro ejemplo .
stems	los país asiátic han segu también nuestr ejempl .
full verbs	los países asiáticos V[seguir] también nuestro ejemplo .

Word alignment results

union	1% train			full train		
	R_S	P_P	**AER**	R_S	P_P	**AER**
baseline*	69.33	67.65	31.56	73.98	84.41	20.92
baseline	73.37	69.43	28.77	78.42	86.43	17.56
base forms	73.93	75.01	**25.51**	76.73	87.90	17.82
stems	74.66	75.65	**24.82**	77.81	88.94	**16.74**
full verbs	73.96	71.36	27.45	78.60	87.37	**16.97**
verbs + stems	75.47	75.17	**24.69**	78.36	88.82	**16.42**

- significant AER reduction for small-data track
- stemming and verb form classification still useful for large-data track
- no Recall increase over baseline for large-data track

Correlation with translation results

AER	Eng→Spa BLEU	Eng→Spa NIST	Spa→Eng BLEU	Spa→Eng NIST
1% train				
31.6 baseline*	31.98	7.922	41.59	9.163
28.8 baseline	32.15	7.943	42.09	9.204
27.5 full verbs	32.51	8.015	42.18	9.249
24.8 stems	32.54	8.031	**42.83**	**9.319**
24.7 verbs+stems	**32.90**	**8.048**	41.90	9.229
full train				
20.9 baseline*	47.94	9.887	55.19	10.797
17.6 baseline	48.02	9.909	55.26	10.763
17.0 full verbs	47.90	9.922	55.14	10.779
16.7 stems	47.87	9.883	55.53	10.788
16.4 verbs+stems	47.85	9.889	55.25	10.765

- Small-data track: mostly positive correlation, yet small variation range
- Large-data track: unsignificant variation range
  - ▷ tendency to use more tuples to NULL (shorter output)
  - ▷ Kneser-Ney smoothing (?)

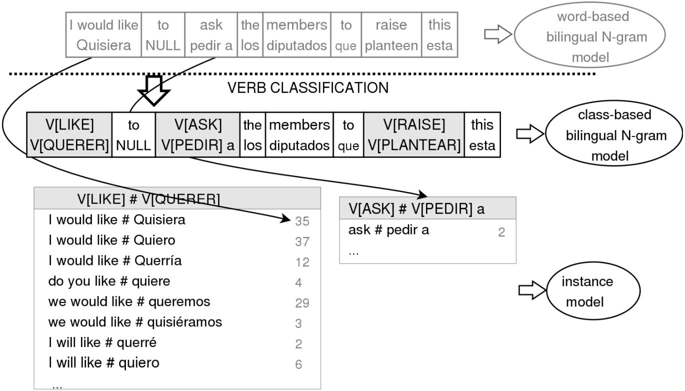
Verb form classification model

Given observance of extensive correct verb form generation errors:

$$h_{BM}(s,t) \approx h_{BM}(\tilde{s},\tilde{t}) \cdot h_{IM}(t|s,\tilde{s},\tilde{t})$$

- deterministic detection of full verb forms (including pronoun, if any)
- standard bilingual model estimated on classified text $(\tilde{s},\tilde{t})$
- novel **instance model** into log-linear combination to select t
  - ▷ relative frequency solution: $h_{IM}(t|s,\tilde{s},\tilde{t}) = \dfrac{N(\tilde{s},\tilde{t},s,t)}{N(\tilde{s},\tilde{t},s,*)}$

Verb form classification



- generalisation power using Part-Of-Speech information
  - person-number generalisation implemented
  - only for unseen forms or for all classified tuples (extended)

Verb form classification. Results

- English→Spanish LC-STAR project data:
  - ▷ spontaneous speech transcriptions (tourist, travel-planning domain)
  - ▷ train: 30k sents., ∼400k words per lang.
  - ▷ dev/test: 500 sents., 3 human references
  - ▷ lexicon models **discarded**

	BLEU	WER
baseline	0.671	23.16
verb class	0.686	22.22
verb class +gen	**0.692**	**21.65**
verb class +genEX	**0.689**	**21.62**

⇒ significant improvement in translation quality

⇒ slight contribution of generalisation (∼5% unseen forms)

Verb form classification. Results (2)

- English→Spanish European Parliament data:
  - ▷ train: ∼1.2M sents., ∼35M words per lang.
  - ▷ dev/test: 1k sents., 2 human references
  - ▷ lexicon models **included** (defined over classes)

	BLEU	WER
baseline	**0.4789**	**40.22**
verb class	0.4696	40.88
verb class + gen	0.4707	40.77
verb class + genEX	0.4696	41.11

- no gain in translation quality

Verb form classification. Complementary exps.

- discarding lexicon models (comparable experiment):

	BLEU	WER
baseline noLEX	**0.4422**	**44.34**
verb class noLEX	0.4376	44.80

  - ▷ no gain in translation quality either

- evaluation against verb-classified references (ignore instances):

	BLEU	WER
baseline	**0.5055**	**38.42**
verb class	0.5013	38.49

  - ▷ no gain with classified translation modelling
  - ▷ seemingly simple relative frequency instance model may suffice

⇒ too big classes (including excessive amount of verb form variety) seem to be weakening bilingual context

Difficulty of morphology production

English→Spanish example:

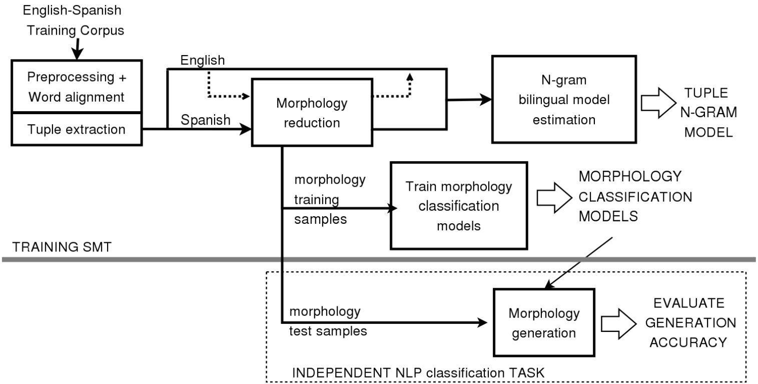
I ask you and your party to give support for the release			
Les pido a usted y a su partido que	respalden	la	liberación
Part-Of-Speech information →	POS:VM M:sjve T:pres P:3rd N:pl	POS:DA G:fem N:sing	POS:NC G:fem N:sing
base form →	[respaldar]	[el]	[liberación]

⇒ Limitation of the bilingual n-gram formulation

T_1	T_2	T_3	T_4	T_5	T_6
your su	party partido	to que	give support for respalden	the la	release liberación

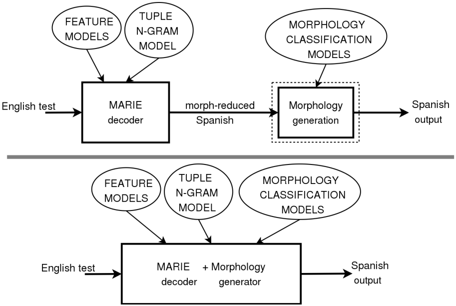
Morphology reduction framework

- how much is morphology affecting the potential of bilingual n-gram formulation?
- evaluation framework based on morphology reduction



Post-processing oracles vs modelling oracles

- morphology reduction of human references provides *oracles*
  - modelling oracles (morph-reduced bilingual model)
  - post-processing oracles (morph-reduced baseline output)
- oracle comparison reveals how much is being *lost in morphology* under Ngram-based SMT framework
- final morphology generation as NLP classification problem



Morphology reduction configurations

English	she	has	a	strong	interest	in
Eng POS	PRP	VBZ	DT	JJ	NN	IN
Spanish	ella	tiene	el	máximo	interés	en
Spa POS	PP3SF	VMIP3S	DAMS	AQMS	NCMS	SPS

S:D	ella tiene DAgn[el] máximo interés en
S:A	ella tiene el AQgn[máximo] interés en
S:N	ella tiene el máximo NCMn[interés] en
S:P	PPpng[] tiene el máximo interés en
S:V	ella VMIPpn[tener] el máximo interés en
S:V_{MT}	ella VMmtpn[tener] el máximo interés en
S:DAV_{MT}	ella VMmtpn[tener] DAgn[el] AQgn[máximo] interés en
S:$full$	PPpng[] VMmtpn[tener] DAgn[el] AQgn[máximo] NCMn[interés] en
E:$full$	PRPpng[] VBPp[have] a very strong Nn[interest] in

Oracle results. Post-processing

reduction	post-processing BLEU	post-processing NIST	morph-reduced model BLEU	morph-reduced model NIST
baseline	0.4785	9.889		
S:D	**0.4853**	9.925		
S:A	0.4823	9.947		
S:N	0.4789	9.880		
S:P	0.4793	9.893		
S:V	0.4833	9.941		
S:V_{MT}	**0.4916**	**10.036**		
S:DAV_{MT}	0.5042	10.148		
S:$full$	**0.5093**	**10.203**		
S:$full$+E:$full$	–	–		

- post-processing oracles:
  - ▷ verbs (∼9%), determiners (∼13%) and adjectives (∼8%)
  - ▷ nouns (∼22%) and pronouns (∼1%) should be left unchanged
  - ✓ significant maximum oracle of around 0.024 absolute BLEU
  - × whole oracle achievement is unrealistic

Study of oracles

determiners

Adjacent disagreement error	9%	nombrar a un Comisión (REF:una Comisión)
Far disagreement error	39%	las atroces situación (REF:la atroz situación)
Wrong Translated noun	29%	las cautiverio (REF:los cautivos ← the captives)
Reference noun mismatch	23%	las Naciones Unidas (REF:la ONU)

verbs

Verb morph. error	69%	la Unión Europea , que legalizaron (REF:legalizó)
3rd person confusion	14%	como sabe (REF:como saben ← as you know)
Reference mismatch	17%	el pueblo prefiere (REF:los ciudadanos prefieren)

- morphology post-processing oracles are not impressive
- potential gain margin basically owing to verb forms (correlated with error analysis)

Oracle results. Morph-reduced models

reduction	post-processing BLEU	post-processing NIST	morph-reduced model BLEU	morph-reduced model NIST
baseline	0.4785	9.889	0.4785	9.889
S:D	**0.4853**	9.925	0.4840	9.936
S:A	0.4823	9.947	0.4819	9.984
S:N	0.4789	9.880	–	–
S:P	0.4793	9.893	–	–
S:V	0.4833	9.941	0.4888	9.916
S:V_{MT}	**0.4916**	**10.036**	**0.4972**	**10.085**
S:DAV_{MT}	0.5042	10.148	0.5060	10.227
S:$full$	**0.5093**	**10.203**	**0.5168**	**10.325**
S:$full$+E:$full$	–	–	0.5143	10.250

- modelling oracles:
  - ▷ neither determiners (∼14%) nor adjectives (∼8%) contribute to a better bilingual model estimation
  - ✓ verb form reduction yields improvement in translation oracles
  - × improvement is slight

Discussion

- Morphology reduction is benefitial (especially for verbs) but not sufficient
- Complex lexical and morphological dependencies are not captured by the translation model
- Ngram-based formulation needs to be reviewed
  - ▷ word reordering strategies
  - ▷ (bilingual and monolingual) syntax-aware features
- AdaBoost for morphology generation (feasible)

1. N-gram-based SMT

2. Introducing linguistic information

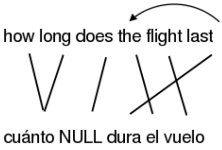
3. **Word reordering strategies**
   - Tuple unfolding and reordered search
   - Linguistic reordering patterns
   - Comparison and discussion

4. Further research lines

Tuple unfolding

In order to apply *constrained* reordered search:

- bilingual n-gram translation model has to be estimated in non-monotone order
- tuple definition:
  - bilingual monotonicity $\rightarrow$ **monolingual monotonicity**
  - tuple sequence defined according to **target** language order
- tuple unfolding:



REGULAR TUPLES:

how long#cuánto
does#NULL
the flight last#dura el vuelo

UNFOLDED TUPLES:

how long#cuánto
does#NULL
last#dura
the#el
flight#vuelo

Reordered search

MARIE decoder with reordering capabilities:

- allowance to cover source positions in any order (in principle)
- target language order is unchanged (translation model, target LM)
- high CPU cost, two constraining parameters:
  - distortion limit (m): we allow source word order change unless it does not exceed a distortion limit, measured in words (first word in each tuple)
  - reordering limit (j): we only allow a maximum number of jumps (word order change) in any complete translation path
- distortion feature model:

$$h_{DIST} = \sum_{k=1}^{K} d_k$$

where d_k: distance between the first source word of k^{th} tuple, and the last source word of $k-1^{th}$ tuple plus 1.

Results

IWSLT'06	Chi→Eng (dev123)			Chi→Eng (official)		
	BLEU	**NIST**	**WER**	**BLEU**	**NIST**	**WER**
baseline	0.3972	8.584	47.02	–	–	–
unf+reord.srch.	**0.4626**	**8.846**	**40.85**	**0.1863**	**5.571**	**68.04**

IWSLT'06	Ara→Eng (dev123)			Ara→Eng (official)		
	BLEU	**NIST**	**WER**	**BLEU**	**NIST**	**WER**
baseline	0.5242	10.179	33.95	–	–	–
unf+reord.srch.	**0.5511**	**10.445**	**31.67**	**0.2323**	**6.238**	**62.71**

EuroParl	Eng→Spa			Spa→Eng		
	BLEU	**NIST**	**WER**	**BLEU**	**NIST**	**WER**
baseline	0.4775	9.73	41.89	0.5532	10.70	34.28
unf+reord.srch.	0.4780	9.81	41.94	0.5447	10.63	34.95

train size: Chinese: 40k, Arabic: 20k

parameters: m=5, j=3 dev123: dev'04 + test'04 + test'05 (16 refs)

Discussion

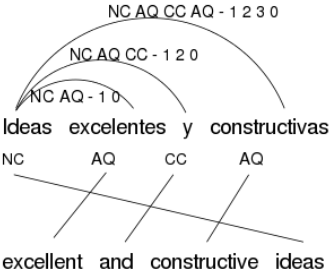
- Successfully incorporates word reordering for Chinese and Arabic to English tasks in Ngram-based SMT formulation $\rightarrow$ significant improvement in scores
- Even with constraining parameters, computational costs are high
- Despite the need for local reordering, no positive impact in Spanish$\leftrightarrow$English task

Discussion

- Successfully incorporates word reordering for Chinese and Arabic to English tasks in Ngram-based SMT formulation → significant improvement in scores
- Even with constraining parameters, computational costs are high
- Despite the need for local reordering, no positive impact in Spanish↔English task
- Alternative approach: (Crego and Mariño, 2006)
  - ▷ Only allow word reordering in certain situations (if test sentence has similar properties than reordered training sents.)
  - ▷ Extend monotone search graph with reordered paths according to automatically-learnt source Part-Of-Speech sequences

Reordering pattern extraction

- Using word alignments and source POS tags (first 2 chars)
- Extract crossed link patterns produced in training
- Filter unfrequent patters:
  - min occurrences (1000), min reorder/monotone prob (0.2)
  - max src–trg diffsize (3 tokens), max monolingual size (5)



Spanish→English patterns

17 patterns extracted

Pattern	Insts.	Example
NC AQ 1 0	877,580	preguntas serias
NC RG 1 0	54,968	actividades aparentemente
AQ AQ 1 0	46,509	medioambientales europeas
RN VM 1 0	45,777	no promuevan
NC AQ AQ 2 1 0	35,661	decisiones políticas delicadas
NC RG AQ 1 2 0	32,887	ideas muy sencillas
NC AQ CC AQ 1 2 3 0	27,119	programa ambicioso y realista
RG VA 1 0	9,824	ahora habíamos
AQ RG 1 0	8,701	suficiente todavía
RG VS 1 0	5,043	supuestamente somos
VM PP 1 0	4,769	estar ustedes
NC CC NC AQ 3 0 1 2	3,355	mezquitas y centros islámicos
AQ RG AQ 1 2 0	2,777	europea más sólida
NC RG AQ CC 1 2 3 0	2,226	ideas muy sencillas y
NC AQ RG AQ 2 3 1 0	1,971	control fronterizo más estricto
NC RG RG 1 2 0	1,473	texto mucho más
NC RG AQ CC AQ 1 2 3 4 0	1,406	ideas muy sencillas y elementales

English→Spanish patterns

29 patterns extracted

Pattern	Insts.	Example
JJ NN 1 0	784,572	Italian parliamentarians
NN NN 1 0	472,809	food scandals
MD RB 1 0	55,226	will actively
JJ JJ 1 0	40,825	liberal European
JJ NN NN 2 1 0	31,395	Belgian Supreme Court
CC JJ NN 2 0 1	30,287	and pro-European forces
JJ JJ NN 2 1 0	29,834	American occupying forces
RB JJ NN 2 0 1	29,379	absolutely rigid control
JJ CC JJ NN 3 0 1 2	27,795	political and symbolic issues
NN PO 1 0	19,216	Barroso 's
NN PO NN 2 0 1	16,493	children 's questions
PO NN 1 0	13,875	's problems
NN JJ 1 0	13,359	EU military
CC NN NN 2 0 1	12,642	and Mrs Zimmer
NN CC NN NN 3 0 1 2	10,559	Lambert and Mrs Zimmer
NN JJ NN 2 1 0	6,351	EU military operation
NN NN PO 2 0 1	3,860	President Bush 's
NN PO JJ 2 0 1	3,576	Bush 's foreign
NN NN PO NN 3 0 1 2	2,684	European Union 's appreciation
JJ CC NN NN 3 0 1 2	2,656	political and policy complexion
NN PO JJ NN 3 2 0 1	2,013	Union 's targeted sanctions
...	...	...

Extending Search Graph

- Whenever a reordering pattern is found in the test file, the (reordered) path is added.
- Reordering paths are not included in the input graph when the translation unit (tuple) exists

DT NC AQ RG AQ
la institución europea más vanguardista

NC AQ -> AQ NC
NC AQ RG AQ -> RG AQ NC AQ

la # the
institución europea # european institution
institución # institution
europea # european
más # most
vanguardista # avant-garde

the | most | avant-garde | european | institution

the | european institution | most | avant-garde

Results (cont.)

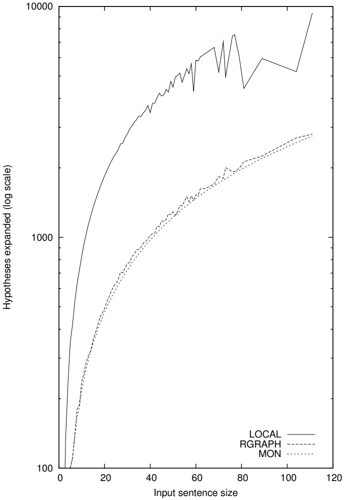
IWSLT'06	Chi→Eng (dev123)			Chi→Eng (official)		
	BLEU	**NIST**	**WER**	**BLEU**	**NIST**	**WER**
baseline	0.3972	8.584	47.02	–	–	–
unf+reord.srch.	**0.4626**	**8.846**	**40.85**	**0.1863**	**5.571**	**68.04**
reord.patterns	0.4463	8.993	43.55	0.1834	5.740	69.76

IWSLT'06	Ara→Eng (dev123)			Ara→Eng (official)		
	BLEU	**NIST**	**WER**	**BLEU**	**NIST**	**WER**
baseline	0.5242	10.179	33.95	–	–	–
unf+reord.srch.	**0.5511**	**10.445**	**31.67**	**0.2323**	**6.238**	**62.71**
reord.patterns	0.5471	10.412	31.79	0.2267	6.135	63.10

EuroParl	Eng→Spa			Spa→Eng		
	BLEU	**NIST**	**WER**	**BLEU**	**NIST**	**WER**
baseline	0.4775	9.73	41.89	0.5532	10.70	34.28
unf+reord.srch.	0.4780	9.81	41.94	0.5447	10.63	34.95
reord.patterns	**0.5006**	**10.00**	**39.73**	**0.5611**	**10.76**	**33.59**

parameters: m=5, j=3 dev123: dev'04 + test'04 + test'05 (16 refs)

Efficiency comparison



- extended search graph with reordering patterns expands a similar number of nodes than monotone search
- distortion model is not needed anymore
- faster optimisation loop

Discussion

- significant improvement in Spa↔Eng tasks
- close to monotone-search efficiency
- insufficient contribution in Chi→Eng and Ara→Eng
  - ▷ long-distance word reordering needs
  - ▷ Part-Of-Speech sequence may not be generalising
  - ▷ *syntax-aware* reordering patterns should be extracted

NIST'06	Chi→Eng (test'05)			Chi→Eng (official)	
	BLEU	**NIST**	**WER**	**BLEU**	**NIST**
unf+reord.srch.	0.2097	7.500	73.69	0.2071	7.217
reord.patterns	0.2120	7.601	72.85	0.2075	7.231

Further research

- inducing syntactic structures from data
  - ▷ syntax-aware feature functions, aiming at:
    - word order estimation (lexicalised vs. general trade-off)
    - lexical and morphology dependencies estimation
  - ▷ translation unit redefinition
    hierarchical phrases (Chiang, 2005)
  - ▷ in combination with morphology generalisation
    factored models (Kirchhoff, 2005)
- powerful analysis tools: online decoding, error classification measures (Popovic, 2006)
- re-ranking and system combination (RWTH Aachen)
- integration with speech recognition

Thanks for your attention

Questions, comments?