



The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus
Corpus processing
and training
Handling ASR input

The translator

Principles
Decoding details
Quantitative data
Translation's output

Post-decoding
experiments

Parameter tuning
Long sentence
handling
Improved target
language model

The 2006 LIMS Translation System for TC-STAR

Presentation for OpenLab 2006

Daniel Déchelotte Holger Schwenk Jean-Luc Gauvain

Group TLP (spoken language processing)
LIMS-CNRS
Université Paris-Sud, France

1st April 2006



Talk outline

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus

Corpus processing
and training

Handling ASR input

The translator

Principles

Decoding details

Quantitative data

Translation's output

Post-decoding
experiments

Parameter tuning

Long sentence
handling

Improved target
language model

- 1 The data
 - Corpus
 - Corpus processing and training
 - Handling ASR input
- 2 The translator
 - Principles
 - Decoding details
 - Quantitative data
 - Translation's output
- 3 Post-decoding experiments
 - Parameter tuning
 - Long sentence handling
 - Improved target language model



Data and condition

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus

Corpus processing
and training

Handling ASR input

The translator

Principles

Decoding details

Quantitative data

Translation's output

Post-decoding
experiments

Parameter tuning

Long sentence
handling

Improved target
language model

- EPPS data
- As of March 2006: from 1996 to May 2005
- Verbatim (manual transcription of speech) and ASR (automatic speech recognition) conditions
- True case, with punctuation

Sample verbatim and ASR source sentences, with references

Verbatim: conviene recordarlo , porque puede que se haya olvidado .

ASR: conviene recordar porque puede que se haya olvidar .

Reference #1: it is appropriate to remember this , because it may have been forgotten .

Reference #2: it is good to remember this , because maybe we forgot it .



Corpus processing and training

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus

Corpus processing
and training

Handling ASR input

The translator

Principles

Decoding details

Quantitative data

Translation's output

Post-decoding
experiments

Parameter tuning

Long sentence
handling

Improved target
language model

- Strip section titles and speaker names
- Semi-reversible conversion to latin1 (highlights numerous normalization issues)
- “verbatimization” of training data (which is in FTE condition): transform numbers from digits to letters, mainly
- Run Giza++ with all options at their default value (same training sequence, same “-p0 0.98” flag as in “trainGIZA++.sh” script, no word classes)



Handling ASR input

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus

Corpus processing
and training

Handling ASR input

The translator

Principles

Decoding details

Quantitative data

Translation's output

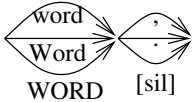
Post-decoding
experiments

Parameter tuning

Long sentence
handling

Improved target
language model

- ROVER input
 - case sensitive
 - little punctuation (no commas, few periods)
 - WER of 6.1% in case insensitive, no punctuation condition
 - WER of 16.1% in case sensitive, with punctuation
- Suppress filler words and partial words
- ROVER repunctuation and re-case-ification:

Replace every word by  , propose to insert
</s><s> at pauses and rescore.

Parameters adapted so as to reproduce the number of punctuation signs in the development data.



Decoder high-level principles

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus

Corpus processing
and training

Handling ASR input

The translator

Principles

Decoding details

Quantitative data

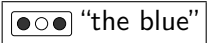
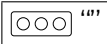
Translation's output

Post-decoding
experiments

Parameter tuning

Long sentence
handling

Improved target
language model

- A* decoder
- Manages partial hypotheses. Example:
 “the blue”, partial translation of “la maison bleue”.
- Starts from the empty hypothesis:  ""
- Considers the most promising hypothesis
 - if complete: it is the algorithm's output
 - if not, extend the partial hypothesis



Extension of a partial hypothesis (1/3)

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus

Corpus processing
and training

Handling ASR input

The translator

Principles

Decoding details

Quantitative data

Translation's output

Post-decoding
experiments

Parameter tuning

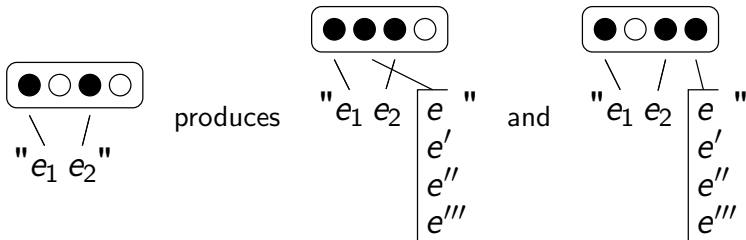
Long sentence
handling

Improved target
language model

Four extension operators.

- "Append"

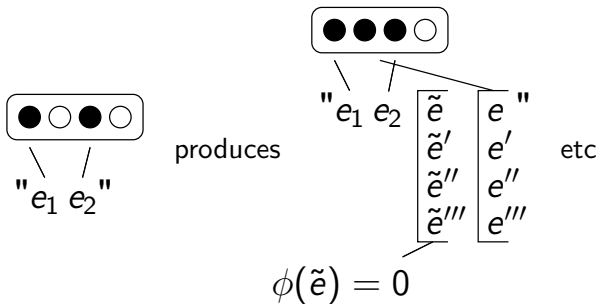
- Produces one target word
- Aligns one uncovered source word to a new target word
- Produces several hypotheses





Extension of a partial hypothesis (2/3)

- “Insert n infertile words and Append”
 - Produces several target words
 - Aligns one uncovered source word to the last one
 - Produces several hypotheses (in practice, $n = 1$)





Extension of a partial hypothesis (3/3)

- “Extend”

- Produces no target word
- Aligns one uncovered source word to the last produced target word (increasing its fertility)
- When applicable, produces one partial hypothesis



- “Complete with e_0 ”

- Produces no target word
- Aligns all uncovered source words to e_0
- When applicable, produces one complete hypothesis





Managing partial hypotheses

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus
Corpus processing
and training
Handling ASR input

The translator

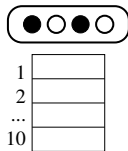
Principles
Decoding details
Quantitative data
Translation's output

Post-decoding
experiments

Parameter tuning
Long sentence
handling
Improved target
language model

- Hypotheses that cover (i.e. translate) the same words:
direct comparison

- One queue per subset of source positions:
total of 2^J queues (!)
- Small queues: 10 or 20 hypotheses



- Inter-queue comparison: admissible heuristics

Maximum probability to translate source word f_j :

$$H^{TFD}(j) = \max \left\{ t(f_j|e_0), \max_{e \neq e_0, \phi} t(f_j|e) \sqrt[n(\phi|e)]{h^D} \right\} \quad (1)$$

T: Translation F: Fertility D: Distorsion (h^D constant)



Limitation, pruning, and decoding time

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchellotte

The data

Corpus
Corpus processing
and training
Handling ASR input

The translator

Principles
Decoding details
Quantitative data
Translation's output

Post-decoding
experiments

Parameter tuning
Long sentence
handling
Improved target
language model

- List of alternative translations: up to 40 per source word, pruned at 10^{-2}
- Queue size of 10 or 20
- “Berger-like” reordering limitation: max. of 4 untranslated words on the left of the rightmost translated word
- Drastic source sentence length limitation: up to 16 words (workaround in a few slides)
- Translation time: 1 second for a 10 word long sentence, doubling every extra source word



Translation's output

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus
Corpus processing
and training
Handling ASR input

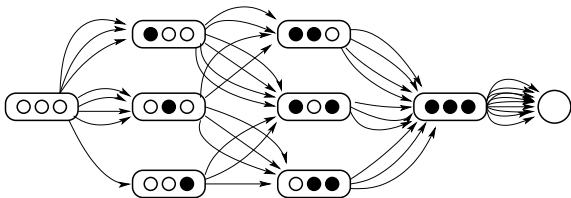
The translator

Principles
Decoding details
Quantitative data
Translation's output

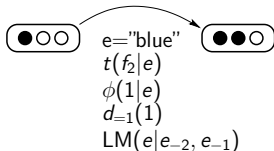
Post-decoding
experiments

Parameter tuning
Long sentence
handling
Improved target
language model

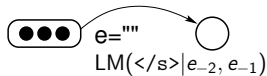
- (Extremely) Schematic word lattice



- Zoom on normal edge



- Zoom on final edge





Translation's output

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus
Corpus processing
and training
Handling ASR input

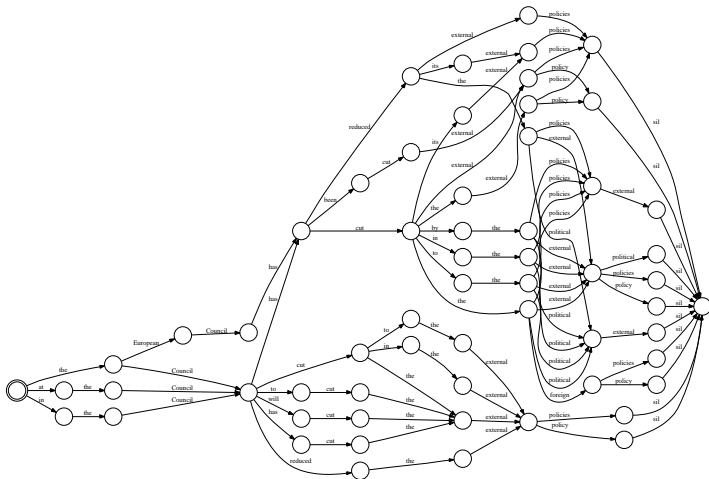
The translator

Principles
Decoding details
Quantitative data

Translation's output

Post-decoding
experiments

Parameter tuning
Long sentence
handling
Improved target
language model





Parameter tuning

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus
Corpus processing
and training
Handling ASR input

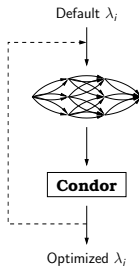
The translator

Principles
Decoding details
Quantitative data
Translation's output

Post-decoding
experiments

Parameter tuning
Long sentence
handling
Improved target
language model

- Five features to tune: lexical, fertility, distortion, and spontaneous insertion models, target language model
- Produce lattices (for the whole development set) with default weights
- Write a script that, given the weights, rescores the lattices and outputs the corresponding BLEU score
- Use Condor [1] to find the best weights
- Iterate, if the new tuning weights are very different
- Gain: from 37.36 to 42.35, in a few hours



[1] <http://iridia.ulb.ac.be/~fvandenb/>



Chopping long sentences into smaller chunks (1/3)

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus

Corpus processing
and training

Handling ASR input

The translator

Principles

Decoding details

Quantitative data

Translation's output

Post-decoding
experiments

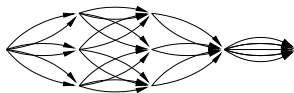
Parameter tuning

Long sentence
handling

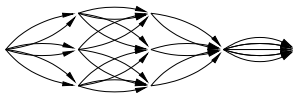
Improved target
language model

- Arbitrary input sentence length (dev06's maximum: 222)
- Current decoder doesn't scale to long sentences
- Split into smaller chunks:
 - Where? At punctuation marks, and then uniformly
 - How to influence decoding? Extra marker in TLM
 - Produce lattices, then merge them and rescore

- Translating chunks with special LM



$LM(e_1|<s>)$

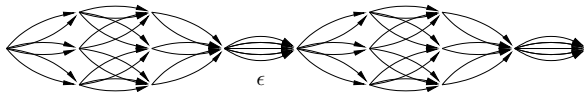


$LM(|e_{-2}, e_{-1})$

$LM(e'_1|)$

$LM(</s>|e'_{-2}, e'_{-1})$

- Joining lattices and rescoreing with normal LM



$LM(e_1|<s>)$

$LM(e'_1|e_{-2}, e_{-1})$

$LM(</s>|e'_{-2}, e'_{-1})$



Chopping long sentences into smaller chunks (3/3)

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus

Corpus processing
and training

Handling ASR input

The translator

Principles

Decoding details

Quantitative data

Translation's output

Post-decoding
experiments

Parameter tuning

Long sentence
handling

Improved target
language model

BLEU (%) on verbatim dev06

- Before parameter tuning

	1-best concatenation	Lattice concatenation
Normal 3g	31.14	33.57
3g with 	36.54	37.36

- After parameter tuning

	1-best concatenation	Lattice concatenation
Normal 3g	40.20	41.63
3g with 	41.45	42.35



Improved target language model

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus
Corpus processing
and training
Handling ASR input

The translator

Principles
Decoding details
Quantitative data
Translation's output

Post-decoding
experiments

Parameter tuning
Long sentence
handling
Improved target
language model

Type	Data set (# of words)	Perplexity	BLEU
3g	EPPS (33.8 M)	85.5	42.35
4g + WP	EPPS + audio trs (740 k)	79.6	43.26
4g + WP	EPPS + audio trs + BN (352 M) + CNN (232 M)	74.1	no gain
neural 4g + WP	EPPS + audio trs	65.0	44.30

- Note: new feature added: word-penalty
- Submitted to ACL (*Continuous Space Language Models for Statistical Machine Translation*)



Conclusions and perspectives

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus

Corpus processing
and training

Handling ASR input

The translator

Principles

Decoding details

Quantitative data

Translation's output

Post-decoding
experiments

Parameter tuning

Long sentence
handling

Improved target
language model

- Work on verbatim and ASR conditions
- A* decoder for translation
- Output: word lattice with individual scores, allows efficient rescoring
- Tuning with Condor
- Neural 4g target language model
- Future: context dependent models, and phrase-based model



Questions and discussion

Thank you for your attention

The 2006
LIMS
Translation
System for
TC-STAR

D. Déchelotte

The data

Corpus

Corpus processing
and training

Handling ASR input

The translator

Principles

Decoding details

Quantitative data

Translation's output

Post-decoding
experiments

Parameter tuning

Long sentence
handling

Improved target
language model

Questions and discussion