LIMSI @WMT12

Hai-Son Le, Thomas Lavergne, Alexandre Allauzen, Marianna Apidianaki,
Li Gong, Aurélien Max, Artem Sokolov, Guillaume Wisniewski, François Yvon
Univ. Paris Sud and LIMSI—CNRS Orsay, France

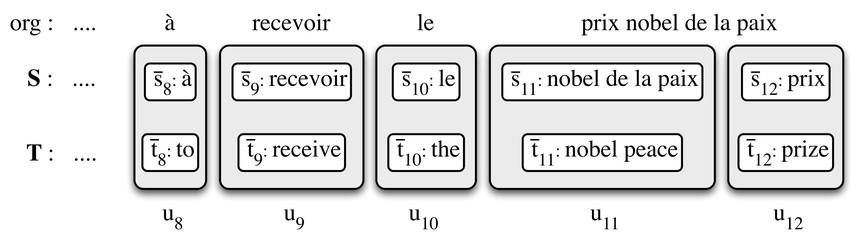
Highlights

French-English and German-English shared translation tasks

- *n*-code (`http://ncode.limsi.fr`):
  - Source reordering as pre-processing
  - Translation models based on bilingual *n*-grams
- **Translation models in a continuous space**
  - With a large vocabulary
  - A wide context ($n = 10$)
  - A reliable probability estimate with SOUL

n-code

Tuples are bilingual units



The translation model is a n-gram model of tuples

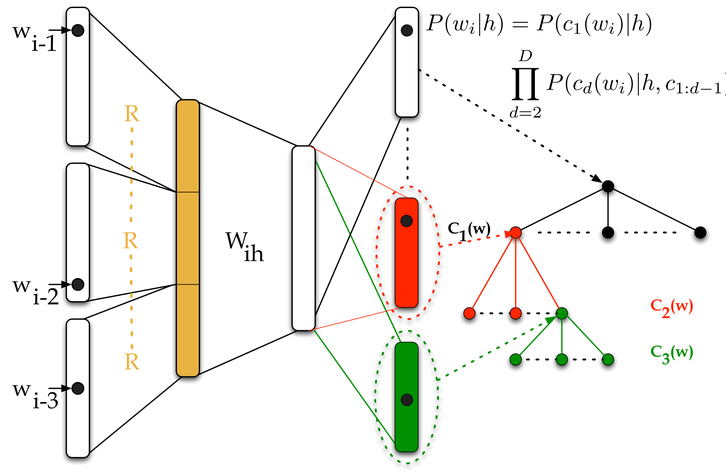
n-code models

- 3-gram tuple LM and 4-gram target word LM
- Four lexicon models (similar to the phrase table)
- Two lexicalized reordering models (predict orientation of next/previous translation unit)
- Weak distance-based distortion model
- Word-bonus and a tuple-bonus models

Data Pre-processing

- Better normalization tools provide better BLEU scores
- Specific pre-processing for German as source language
- Cleaning noisy data sets (*GigaWord*)
  - Discard sentences in other languages
  - Remove repeated sentences, or the ones included in the development sets
  - Normalize the character set
  - Select best half of the data set according to perplexity

SOUL *n*-gram Model Overview



Experimental setup

All the systems are constrained.

SOUL setup

- A wide context: $n = 10$.
- Two hidden layers (1000,500)
- k-best rescoring ($k = 300$)
- All the available corpora are used (resampling)

On-the-fly system

- Estimation of standard phrase-based models
- By sampling accordingly to the text to be translated

Factored *n*-gram translation models

$P(\mathbf{s}, \mathbf{t}) = \prod_{i=1}^{L} P(u_i|u_{i-1}, ..., u_{i-n+1})$, with

$$P(u_i|u_{i-1}, ..., u_{i-n+1}) = P(\bar{t}_i|\bar{s}_i, \bar{s}_{i-1}, \bar{t}_{i-1}, ..., \bar{s}_{i-n+1}, \bar{t}_{i-n+1}) \times P(\bar{s}_i|\bar{s}_{i-1}, \bar{t}_{i-1}, ..., \bar{s}_{i-n+1}, \bar{t}_{i-n+1}) \quad (1)$$

conditional translation model — distortion model

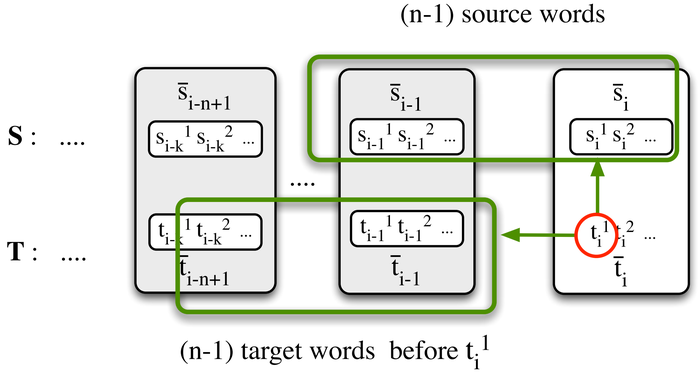
- An alternate decomposition is possible

→ 4 models to be estimated with SOUL

- A further decomposition in word streams
- Synchronized with the bilingual segmentation.

Word decomposition

In equation 1, to overcome data sparsity issues:



Experimental Results

Direction	*System*	BLEU *test2011*	BLEU *test2012**
en2fr	Baseline	32.0	28.9
	+ SOUL TM	**33.4**	**29.9**
	on-the-fly	31.7	28.6
fr2en	Baseline	30.2	30.4
	+ SOUL TM	**31.1**	**31.5**
en2de	Baseline	15.4	16.0
	+ SOUL TM	**16.6**	**17.0**
de2en	Baseline	21.8	22.9
	+ SOUL TM	**22.8**	**23.9**

* official results.