LIG approach for IWSLT09, Using Multiple Morphological Segmenters for Spoken Language Translation of Arabic

F. Bougares, L. Besacier, H. Blanchon

Laurent.besacier@imag.fr

University of Grenoble Laboratory of Informatics of Grenoble (LIG) GETALP team





Overview of LIG system

Third participation to IWSLT

- A/E task
- Phrase-based system using Moses+Giza+srilm

Use of IWSLT-provided data only except

- A 84k A/E bilingual dictionary taken from http://freedict.cvs.sourceforge.net/freedict/eng-ara/
- Different morphological analyzers
- LDC's Gigaword corpus (for english LM training)





Official IWSLT08 results

L. Besacier, A. Ben-Youcef, H. Blanchon « The LIG Arabic / English Speech Translation System à IWSLT08 » IWSLT08. Hawai. USA. October 2008.

| | (bleu+meteror)/2 | bleu | meteor |
|----------|------------------|--------|--------|
| verbatim | 0.5674 | 0.4595 | 0.6752 |
| ASR | 0.5022 | 0.3931 | 0.6113 |

- In 2008, the systems were ranked according to the (BLEU+METEOR)/2 score of the primary ASR output run submissions. The LIG was ranked 6th/10 based on this rule.
- The LIG ranking score (this year official evaluation metric) was 0.3756. The LIG was ranked 4th/10 based on this metric

Multiple morphological analyzers

| | Seg.Buckwalter | Seg.ASVM |
|---|-----------------------------------------------------------------------|---------------------------------------------------------------------|
| 1 | ما مقاس ك What size do you were | ما مقاسك What your size |
| 2 | س ي ستغرق ذلك حوالي ثلث ون دقيق ة It will take that thirty minutes | سيستغرق ذلك حوالي ثلاثون دقيقة It will take about thirty minutes |
| 3 | أيمكنني استعمال هاتف ك Can i use your telephone | أ يمكنني استعمال هاتف ك Can i your phone |
| 4 | عند ي حمى منذ ال بارح ة I have a fever since | عندي حمى منذ البارحة I have a fever since yesterday |
| 5 | هل ب إمكان ي التحدث إلى ال سيد كارتر Can i talk to mr | هل ب إمكان ي التحدث إلى السيد كارتر May i speak to mr carter |
| 6 | صدم ت ني سيارة I was bit by a car | صدمت ني سيارة Was hit by a car |
| 1 | عند ي ألم في ال أذن I have a pain in my ear | عند ي ألم في الأذن I have a pain in turn |

2 correct segmentations

2 incorrect segmentations

Correct seg.

Incorrect trans.

And vice-versa





Formalisation of Multiple Segmentation

$$e^* = \arg\max_{e} P(e/f)$$

$$e^* = \arg\max_{e} \sum_{a_1^k} P(e, a_1^k / f)$$

$$e^* = \arg\max_{e} \sum_{a_1^k} P(a_1^k / f) . P(e / f, a_1^k)$$

$$e^* \approx \arg\max_{e} \left\{ \max_{a_1^k} P(a_1^k / f) . P(e / a_1^k) \right\}$$

$$e^* \approx \arg\max_{e} \left\{ \max_{a_1^k} P(a_1^k / f) . P(a_1^k / e) . P(e) \right\} (5)$$

where

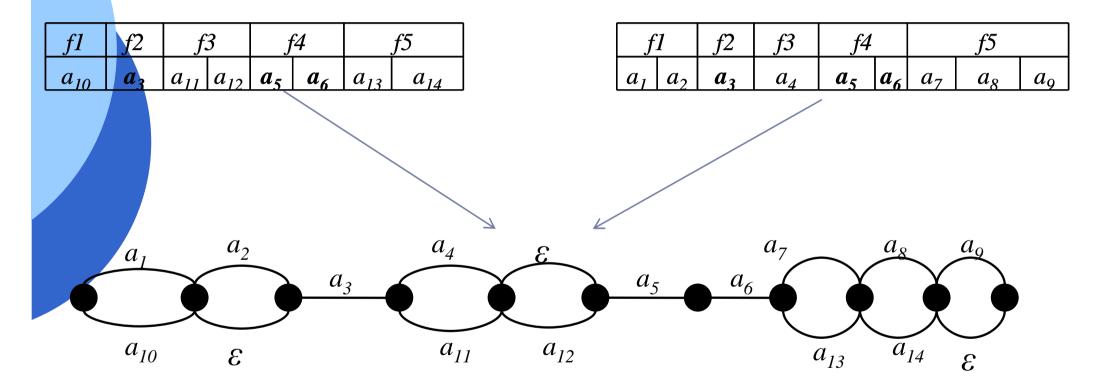
P(e) is the target language model,

 $P(a_1^k/f)$ is the "segmentation" model,

 $P(a_1^k/e)$ is the translation model trained for a given segmentation



Representing Segmentation Amibiguity using CN

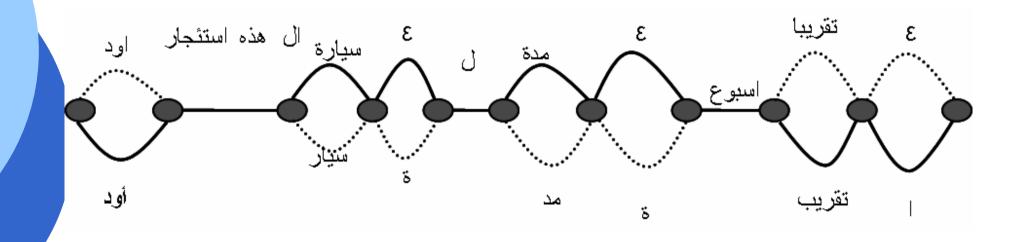


CN decoded using multiple translation tables (1 for each segmentation)





Exemples



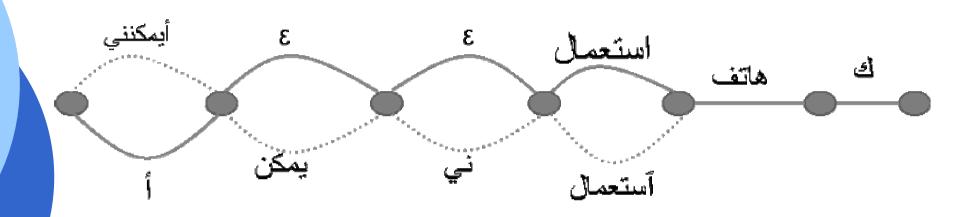
Buckwalter: I'd like this car for about a week

ASVM: to rent this car for long old almost

Multi-segmentation: I'd like to rent this car for about a week



Exemples



Buckwalter: Can I use your phone

ASVM: Can I use your phone

multi-segmentation: I use your phone





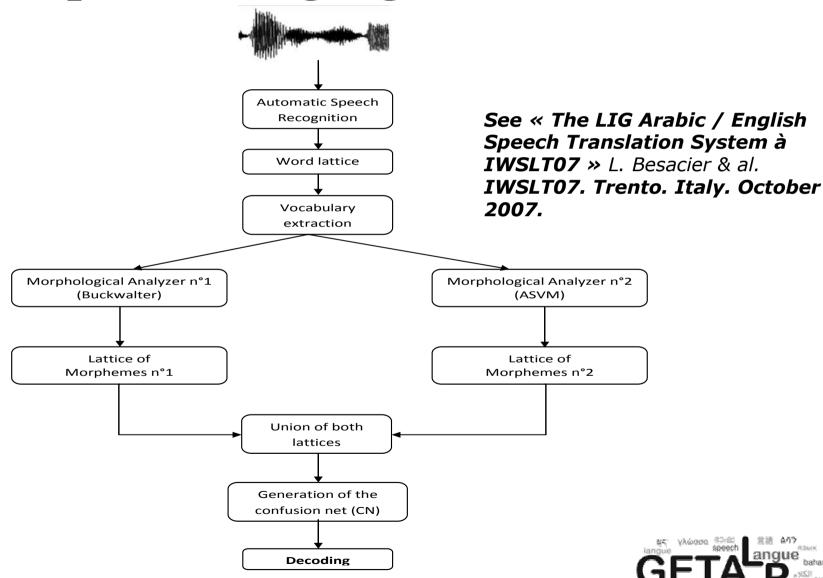
Text Translation Results on IWST data

| | Dev06 | | | Tst06 | | | Tst07 | | |
|------------|-------|------|--------|-------|------|--------|-------|------|--------|
| Score | BLEU | NIST | METEOR | BLEU | NIST | METEOR | BLEU | NIST | METEOR |
| ASVM | 32.34 | 7.07 | 67.46 | 25.47 | 6.37 | 64.87 | 50.52 | 8.18 | 77.93 |
| Buckwalter | 35.11 | 7.47 | 68.72 | 28.17 | 6.84 | 65.61 | 50.79 | 8.27 | 77.56 |
| Multiple | 34.39 | 7.28 | 64.24 | 29.19 | 6.72 | 67.31 | 48.47 | 7.99 | 77.32 |





Multi-segmentation for Spoken Language Translation





ASR Translation Results on IWST data

| | Dev06 | | | Tst06 | | | Tst07 | | |
|------------|-------|------|--------|-------|------|--------|-------|------|--------|
| Score | BLEU | NIST | METEOR | BLEU | NIST | METEOR | BLEU | NIST | METEOR |
| ASVM | 28.27 | 6.47 | 62.92 | 22.78 | 5.91 | 60.99 | 41.04 | 7.13 | 70.92 |
| Buckwalter | 27.69 | 6.40 | 63.04 | 24.19 | 6.09 | 62.14 | 42.54 | 7.29 | 71.44 |
| Multiple | 29.27 | 5.57 | 66.73 | 24.01 | 5.21 | 66.09 | 42.72 | 7.41 | 72.60 |





Exemple of translated sentences

| Buckwalter | How will you pay in cash or card | | | | |
|------------------|---------------------------------------------|--|--|--|--|
| ASVM | How do you pay in cash credit card | | | | |
| Mul tiple | How will you pay in cash or credit card | | | | |
| Buckwalter | I'd like this car for about a week | | | | |
| ASVM | To rent this car for long old almost | | | | |
| M ultiple | I'd like to rent this car for about a week | | | | |
| Buckwalter | I'm sorry sir non-smoking a on the train | | | | |
| ASVM | Sorry sir non-smoking a on the train | | | | |
| Multiple | I'm sorry sir non-smoking seat on the train | | | | |





LIG progress at IWSLT

