



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

---

F. Bougares, **L. Besacier**, H. Blanchon

[Laurent.besacier@imag.fr](mailto: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. Hawaii. USA. October 2008.

	(bleu+meteor)/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**



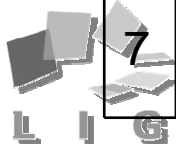
# Multiple morphological analyzers

	Seg.Buckwalter	Seg.ASVM
1	ما مقاسك What size do you were	ما مقاسك What your size
2	س ي ستغرق ذلك حوالي ثلث ون دقيقة It will take that thirty minutes	سيستغرق ذلك حوالي ثلاثون دقيقة It will take about thirty minutes
3	أيمكنني استعمال هاتفك Can i use your telephone	أيمكنني استعمال هاتفك Can i your phone
4	عندي حمى منذ ال بارحة 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	صدمت ني سيارة I was bit by a car	صدمت ني سيارة Was hit by a car
7	عندي ألم في ال أذن I have a pain in my ear	عندي 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

$$e^* = \arg \max_e P(e / f)$$

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

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

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

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

where

$P(e)$  is the target language model,

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

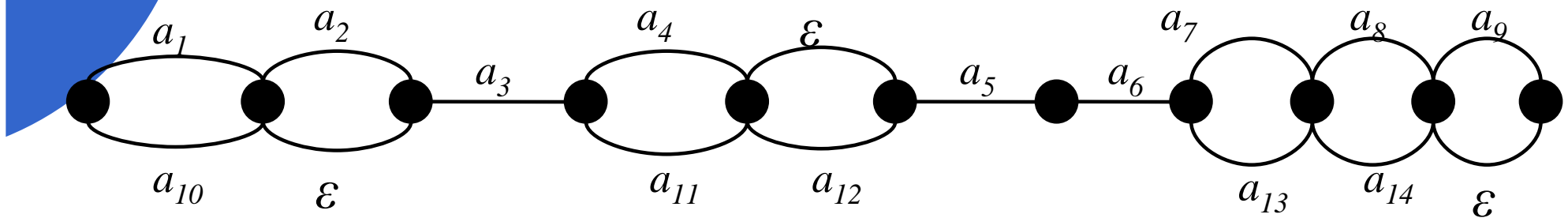
$P(a_1^k / e)$  is the translation model trained for a given segmentation



# Representing Segmentation Ambiguity using CN

<i>f1</i>	<i>f2</i>	<i>f3</i>	<i>f4</i>	<i>f5</i>			
<i>a</i> <sub>10</sub>	<i>a</i> <sub>3</sub>	<i>a</i> <sub>11</sub>	<i>a</i> <sub>12</sub>	<i>a</i> <sub>5</sub>	<i>a</i> <sub>6</sub>	<i>a</i> <sub>13</sub>	<i>a</i> <sub>14</sub>

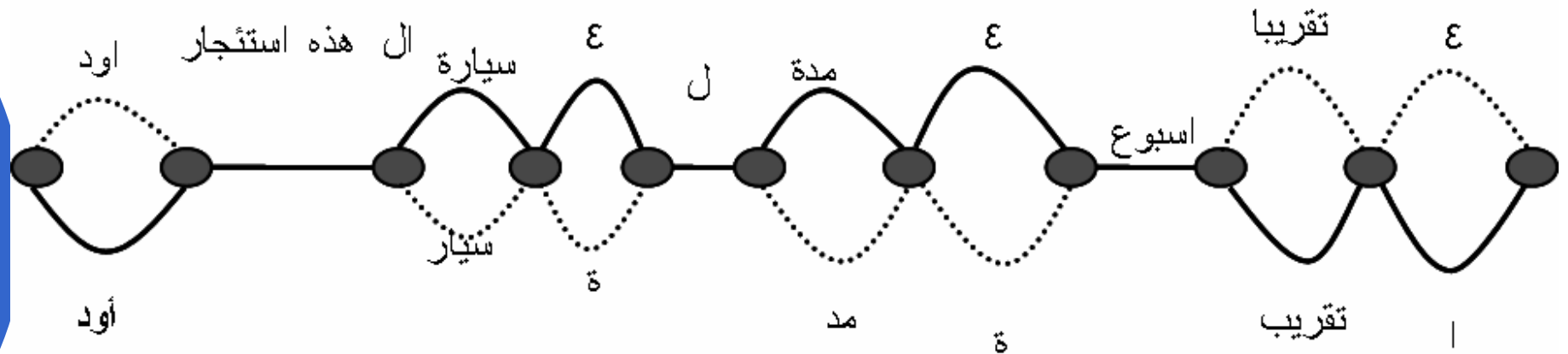
<i>f1</i>	<i>f2</i>	<i>f3</i>	<i>f4</i>	<i>f5</i>				
<i>a</i> <sub>1</sub>	<i>a</i> <sub>2</sub>	<i>a</i> <sub>3</sub>	<i>a</i> <sub>4</sub>	<i>a</i> <sub>5</sub>	<i>a</i> <sub>6</sub>	<i>a</i> <sub>7</sub>	<i>a</i> <sub>8</sub>	<i>a</i> <sub>9</sub>



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



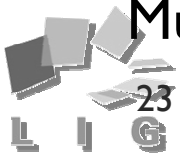
# Examples



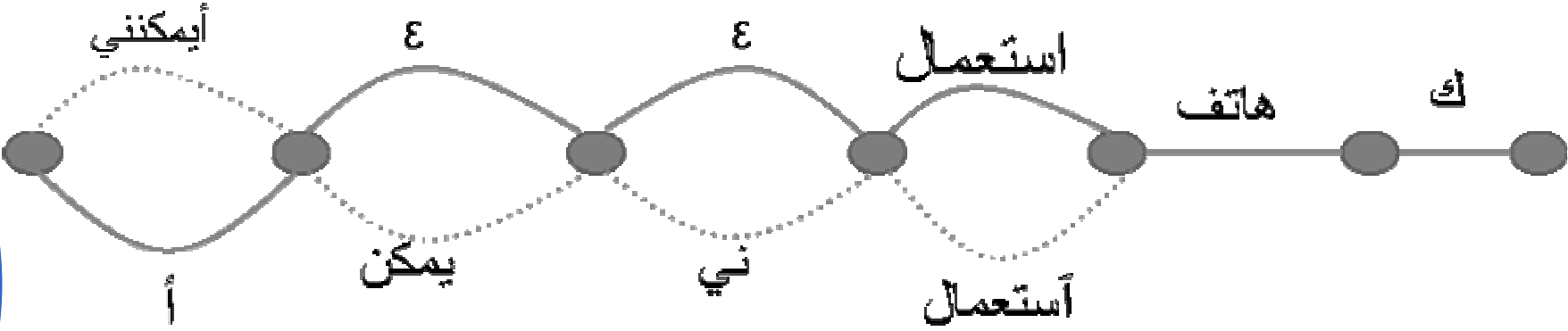
Buckwalter : I'd like this car for about a week

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

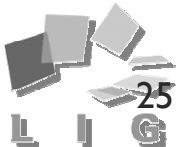
multi-segmentation : I use your phone



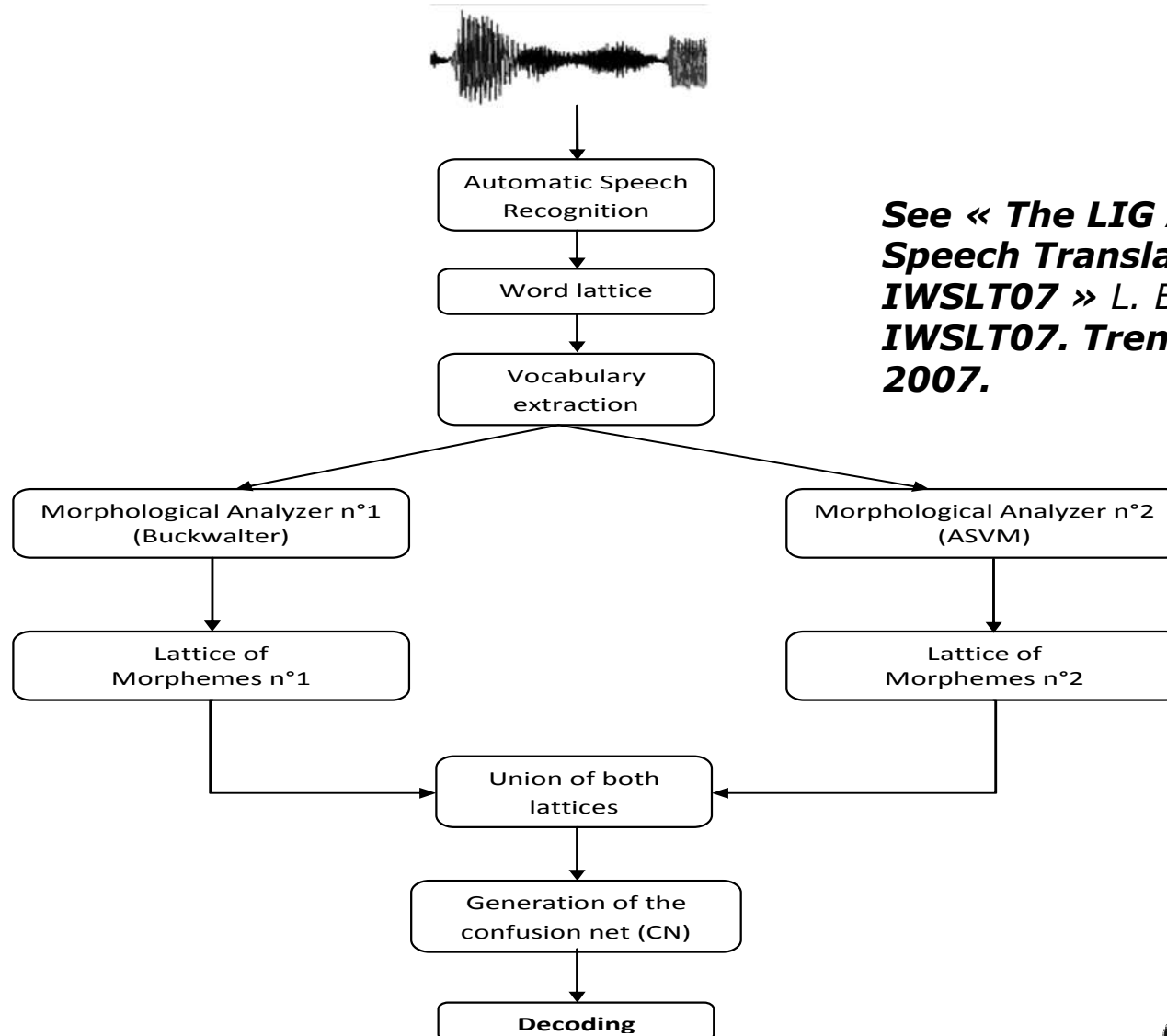


# Text Translation Results on IWST data

Score	Dev06			Tst06			Tst07		
	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	<b>77.93</b>
Buckwalter	<b>35.11</b>	<b>7.47</b>	<b>68.72</b>	28.17	<b>6.84</b>	65.61	<b>50.79</b>	<b>8.27</b>	77.56
Multiple	34.39	7.28	64.24	<b>29.19</b>	6.72	<b>67.31</b>	48.47	7.99	77.32



# Multi-segmentation for Spoken Language Translation



**See « The LIG Arabic / English  
Speech Translation System à  
IWSLT07 » L. Besacier & al.  
IWSLT07. Trento. Italy. October  
2007.**



# ASR Translation Results on IWST data

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



# Exemple of translated sentences

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



# LIG progress at IWSLT

