

Purpose assess the quality of translation systems based on G-LexAR outputs.

G-Lexar

1. segments the text into Arabic morphological units (hyper-forms) and filters other strings inputs.
2. analyzes the hyper-forms independently of their contexts, it provides a tree with all of its segmentation, vocalization, lemmatization, and POS tags.
3. performs the grammatical tagging and pruning of these lexical trees.

```
<mot UM="بسرعة"
  Langue="A" Taille="5" Position="142" Etat="10">
  <AMG V="بسرعة"
    L="+ب- - - - - سرعة"
    HCG="ممنون بحروف جز"/>
  <AMG V="بسرعة"
    L="+ب- - - - - سرعة"
    HCG="مضاف بحروف جز"/>
  :
</mot>
```

BAMA

- ▶ concatenative lexicon-driven approach where morphotactics and orthographic rules are built directly into the lexicon;
- ▶ for each input string, provides a solution including the word's unique identifier, a breakdown of the constituent morphemes, and their POS values.

```
INPUT STRING: الغار
SOLUTION 1: >alogAz
  LEMMA_ID: lugoz_1
  POS: >alogAz/NOUN
  GLOSS: mysteries/enigmas
SOLUTION 2: >alogAzu
  LEMMA_ID: lugoz_1
  POS: >alogAz/NOUN+u/CASE_DEF_NOM
  GLOSS: mysteries/enigmas + [def.nom.]
SOLUTION 3: >alogAza
  LEMMA_ID: lugoz_1
  POS: >alogAz/NOUN+a/CASE_DEF_ACC
  GLOSS: mysteries/enigmas + [def.acc.]
  :
```

Statistical Machine Translation tools

Translations are performed by the phrase-based statistical machine translation toolkit Moses which makes use of the word aligner GIZA++.

Experiments	
Analyzer	Format
none	original
	vocalized
G-LexAr	lemmatized
	segmented
BAMA	lemmatized
	segmented

		no morphological analysis	
original	3	ما هي تكلفة السياحة ليوم كامل في حمام السباحة ؟	
G-LexAr			
vocalized	3	مَا هِيَ تَكْلِفَةُ السَّبَّاحَةِ لِيَوْمٍ كَامِلٍ فِي حَمَّامِ السَّبَّاحَةِ ؟	
lemmatized	3	مَا هِيَ تَكْلِفَةُ سَبَّاحَةِ يَوْمٍ كَامِلٍ فِي حَمَّامِ سَبَّاحَةِ ؟	
segmented	3	مَا هِيَ تَكْلِفَةُ أَلْ سَبَّاحَةِ لِ يَوْمٍ كَامِلٍ فِي حَمَّامِ أَلْ سَبَّاحَةِ ؟	
BAMA			
lemmatized	3	mA hiya takolif sab~AH yawom kAmil fiy HamAm sab~AH ?	
		mA hiya takolif ap	
		Al sab AH ap li	
segmented	3	yawom kAmil fiy HamAm Al sab AH ap ?	

Newsire: preliminary experiment

BTEC texts: IWSLT campaign

Corpus

Sample of 251,000 Arabic-English parallel sentences from the corpus released by the LDC (newsire).

Corpus

Arabic format	Size (Mb)	# tokens	tokens per line
untouched (original)	1	159,006	8.0
vocalized	2	159,006	8.0
G-LexAr			
lemmatized	1	159,006	8.0
segmented	1	203,338	10.2
BAMA			
lemmatized	1	159,000	8.0
segmented	1	255,948	12.8

Results

	mWER	BLEU	TER
original	0.4874 [0.4772, 0.4985]	0.2121 [0.1990, 0.2250]	0.8239 [0.8032, 0.8480]
vocalized	0.4962 [0.4855, 0.5071]	0.1978 [0.1847, 0.2113]	0.8394 [0.8175, 0.8634]
G-LexAr			
lemmatized	0.5000 [0.4896, 0.5106]	0.1973 [0.1850, 0.2092]	0.8451 [0.8237, 0.8699]
segmented	0.4823 [0.4722, 0.4929]	0.2066 [0.1850, 0.2092]	0.8165 [0.7955, 0.8400]
BAMA			
lemmatized	0.4869 [0.4774, 0.4972]	0.2091 [0.1963, 0.2214]	0.8111 [0.7905, 0.8332]
segmented	0.4822 [0.4721, 0.4924]	0.1957 [0.1835, 0.2091]	0.8430 [0.8208, 0.8689]
intersection	[0.4896, 0.4924]	[0.1990, 0.2091]	[0.8237, 0.8332]

Results

morph. analyzer	form (run)	BLEU	METEOR	f1	Prec.	Recl.	WER	PER	TER	GTM	NIST
none	original	0.408	0.693	0.742	0.776	0.711	0.414	0.371	35.67	0.703	6.855
	vocalized	0.206	0.488	0.604	0.733	0.514	0.573	0.537	47.64	0.555	2.433
	lemmatized	0.296	0.591	0.669	0.726	0.620	0.516	0.464	43.57	0.611	4.987
G-LexAr	segmented	0.287	0.607	0.673	0.706	0.643	0.525	0.460	44.76	0.628	5.555
	lemmatized	0.391	0.700	0.738	0.749	0.728	0.428	0.370	36.52	0.692	7.157
BAMA	(segmented)	0.386	0.717	0.736	0.717	0.757	0.438	0.376	38.50	0.716	7.434

For each given metric, all interval scores for G-LexAr, BAMA and original overlap. This shows that the small differences between median scores are not statistically significant.

These results show that the translation systems on an untouched corpus outperforms all G-LexAr-based systems. In addition, the translation systems based on BAMA also outperform G-LexAr based systems.

- ▶ BAMA and G-LexAr behave differently since the former does not really degrade translation system performance but the latter clearly does on BTEC texts.
- ▶ the clear differences observed in the results obtained by G-LexAr on the BTEC texts used for the evaluation campaign (short sentences, small corpus) and on newsire texts from the LDC (longer sentences, larger corpus) come from insufficient coverage of the vocabulary.