that the targeting is not too aggressive and the original meaning is retained in the paraphrase for most of the segments.

3.1.2.1.9 Context Language Model Used in Chinese P3.5 Evaluation

(Contributor: Libin Shen)

Previously we used context language models for Arabic to English MT and obtained a small gain in performance. In this period we applied the same technique to Chinese MT and achieved similar improvement.

For the sake of completeness, we re-introduce the definition of context LM here. HierDec uses a CKY decoder that builds partial translation theories in a bottom up manner, together with beam search which prunes out unpromising partial theories along the way to control decoding complexity. A drawback for such a pruning strategy is that partial theories are selected solely based on the context of the current input span. Please note the ultimate goal of the decoder is to find the best theories for the whole sentence rather than the best partial theories for each input span. The motivation of context language model is to exploit the sentence context to select partial theories that are likely to lead to better translations for the whole sentence. Since modeling the whole sentence is too complex, we used the immediate words before and after the current input span to improve the selection of translation theories. Suppose we have source phrase "f1 f2 f3 f4 f5", and a translation hypothesis for "f2 f3 f5" is "e7 e8 e9", as shown below

```
   | e7 e8 e9 |
   |          |
f1 | f2 f3 f4 | f5
```

We employ the context language model by calculating the probabilities of the following two events.

- $P_{left}(f1|e7,e8)$
- $P_{right}(f5|e9,e8)$

P_{left} and P_{right} are incorporated in the score estimate for theory selection. In this way, when a hypothesis is generated, we have a context LM feature representing the likelihood of the source context given the inside hypothesis. We estimate the probability of the context LM over the bi-lingual training data from which the translation rules are extracted.

Expt.	Set	TER		BLEU	
		Lower	Mixed	Lower	Mixed
30409_l1g	SysCombTune	54.26	56.21	38.63	36.77
30409_m4	SysCombTune	54.01	55.89	38.85	37.02
30409_l1g	Test	58.89	60.94	26.11	24.46
30409_m4	Test	58.88	60.85	26.32	24.76
30409_l1g	Tune	55.66	57.70	37.47	35.60
30409_m4	Tune	55.14	57.13	37.81	35.99

Table 3-29: Effect of context LM on Chinese newswire translation. 30409_l1g: baseline. 30409_m4: with context LM.

Expt.	Set	TER		BLEU	
		Lower	Mixed	Lower	Mixed
31550e	SysCombTune	566.39	68.47	17.41	15.93
31550f	SysCombTune	66.12	68.28	18.04	16.50
31550e	Test	64.93	66.58	15.24	14.27
31550f	Test	64.44	66.05	15.91	14.89
31550e	Tune	64.10	66.18	18.25	16.86
31550f	Tune	64.05	66.11	18.76	17.37

Table 3-30: Effect of context LM on Chinese web translation. 31550e: baseline, 31550f: with context LM.

We tuned the weights on IBM BLEU for newswire, and on TER-BLEU for web. With context LM, we obtained consistent improvement in both BLEU and TER on all sets. The improvement is larger on web than on newswire.

3.1.2.1.10 Modeling Length Distributions of Non-terminals in HeirDec

(Contributor: Libin Shen)

We introduced a new decoding feature based on the length distribution of non-terminals in translation rules. We obtained consistent improvement in Chinese-English MT with this new feature.