$$\begin{aligned} \mathrm{e}_{\mathrm{best}} &= \arg\max_{\mathrm{e}} p(\mathrm{e} \mid \mathrm{f}) \\ &= \arg\max_{\mathrm{e}} p(\mathrm{f} \mid \mathrm{e}) p_{\mathrm{LM}}(\mathrm{e}) \omega^{\mathrm{length(e)}} \end{aligned} \quad (1)$$

Where the translation model $p(\mathrm{f} \mid \mathrm{e})$ can be decomposed into

$$\begin{aligned} & p(\overline{f}_1^I \mid \overline{e}_1^I) \\ &= \prod_{i=1}^{I} \phi(\overline{f}_i \mid \overline{e}_i) d(a_i - b_{i-1}) p_{\mathrm{w}}(\overline{f}_i \mid \overline{e}_i, a)^{\lambda} \end{aligned} \quad (2)$$

Where $\phi(\overline{f}_i \mid \overline{e}_i)$ is the phrase translation probability. a_i denotes the start position of the source phrase that was translated into the *i*th target phrase, and b_{i-1} denotes the end position of the source phrase translated into the (*i*-1)th target phrase. $d(a_i - b_{i-1})$ is the distortion probability. $p_{\mathrm{w}}(\overline{f}_i \mid \overline{e}_i, a)$ is the lexical weight, and λ is the strength of the lexical weight.

3.2 Interpolated Models

We train synthetic models with the synthetic bilingual corpus produced by the RBMT systems. We can also train a translation model, namely standard model, if a real bilingual corpus is available. In order to make full use of these two kinds of corpora, we conduct linear interpolation between them.

In this paper, the distortion probability in equation (2) is estimated during decoding, using the same method as described in Pharaoh (Koehn, 2004). For the phrase translation probability and lexical weight, we interpolate them as shown in (3) and (4).

$$\phi(\overline{f} \mid \overline{e}) = \sum_{i=0}^{n} \alpha_i \phi_i (\overline{f} \mid \overline{e}) \quad (3)$$

$$p_{\mathrm{w}}(\overline{f} \mid \overline{e}, a) = \sum_{i=0}^{n} \beta_i p_{\mathrm{w},i}(\overline{f} \mid \overline{e}, a) \quad (4)$$

Where $\phi_0(\overline{f} \mid \overline{e})$ and $p_{\mathrm{w},0}(\overline{f} \mid \overline{e}, a)$ denote the phrase translation probability and lexical weight trained with the real bilingual corpus, respectively. $\phi_i(\overline{f} \mid \overline{e})$ and $p_{\mathrm{w},i}(\overline{f} \mid \overline{e}, a)$ ($i = 1,...,n$) are the phrase translation probability and lexical weight estimated by *n* synthetic corpora produced by the RBMT systems. α_i and β_i are interpolation coefficients, ensuring $\sum_{i=0}^{n} \alpha_i = 1$ and $\sum_{i=0}^{n} \beta_i = 1$.

4 Resources Used in Experiments

4.1 Data

In the experiments, we take English-Chinese translation as a case study. The real bilingual corpus includes 494,149 English-Chinese bilingual sentence pairs. The monolingual English corpus is selected from the English Gigaword Second Edition, which is provided by Linguistic Data Consortium (LDC) (catalog number LDC2005T12). The selected monolingual corpus includes 1,087,651 sentences.

For language model training, we use part of the Chinese Gigaword Second Edition provided by LDC (catalog number LDC2005T14). We use 41,418 documents selected from the ZaoBao Newspaper and 992,261 documents from the XinHua News Agency to train the Chinese language model, amounting to 5,398,616 sentences.

The test set and the development set are from the corpora distributed for the 2005 HTRDP[2] evaluation of machine translation. It can be obtained from Chinese Linguistic Data Consortium (catalog number 2005-863-001). We use the same 494 sentences in the test set and 278 sentences in the development set. Each source sentence in the test set and the development set has 4 different references.

4.2 Tools

In this paper, we use two off-the-shelf commercial English to Chinese RBMT systems to produce the synthetic bilingual corpus.

We also need a trainer and a decoder to perform phrase-based SMT. We use Koehn's training scripts[3] to train the translation model, and the SRILM toolkit (Stolcke, 2002) to train language model. For the decoder, we use Pharaoh (Koehn, 2004). We run the decoder with its default settings (maximum phrase length 7) and then use Koehn's implementation of minimum error rate training (Och, 2003) to tune the feature weights on the de-

[2] The full name of HTRDP is National High Technology Research and Development Program of China, also named as 863 Program.

[3] It is located at http://www.statmt.org/wmt06/shared-task/baseline.html.