Word Alignment Without NULL Words

Philip Schulz, Wilker Aziz and Khalil Sima'an
{P.Schulz, W.Aziz, K.Simaan}@uva.nl
ILLC, University of Amsterdam

Problems with NULL Words

Several kinds of problems arise when using NULL words for word alignment:

- **Conceptual:** Untranslatable source words are aligned to NULL. Intuitively, they should depend on source side features as their presence is ultimately triggered by source side fluency requirements.
- **Distortion:** The NULL word has a position and thus induces implausibly long distortions that confuse the distortion model.
- **Sparsity:** There is one unique NULL word that occurs in every target sentence. Its distribution over source words tends to be extremely flat. This causes the "garbage collection problem".

Idea

- Use source side language model to generate untranslatable words
- Introduce latent indicator variable Z that models choice between alignment and language model component
- Also capture tendency of individual words to collocate with their neighbours

Our Model

Generative Story

- For each source position j
  - Choose an alignment link a_j according to $P(a_j)$
  - Dependent on the previous source word f_{j-1} decide whether to use the language model $(z_j = 1)$ or the alignment model $(z_j = 1)$ component
  - If $z_j = 1$, generate the source word f_j from its predecessor f_{j-1}
  - If $z_j = 0$, generate the source word f_j from its aligned target word e_{a_j}

We turn this into a Bayesian model by imposing Dirichlet/Beta priors on the Categorical/Bernoulli model parameters.

$$\begin{aligned}
&F_j|e, a_j, z_j = 0 \sim Cat(\theta_{e_{aj}}) && \Theta_{e_{aj}} \sim Dir(\alpha)\\
&F_j|f_{j-1}, z_j = 1 \sim Cat(\theta_{f_{j-1}}) && \Theta_{f_{j-1}} \sim Dir(\beta)\\
&Z_j|f_{j-1} \sim Bernoulli(q_{f_{j-1}}) && Q_{f_{j-1}} \sim Beta(s, r)\\
&A \sim Cat(\theta_a) && \Theta_a \sim Dir(\gamma)
\end{aligned}$$

Inference is done using a collapsed Gibbs sampler.

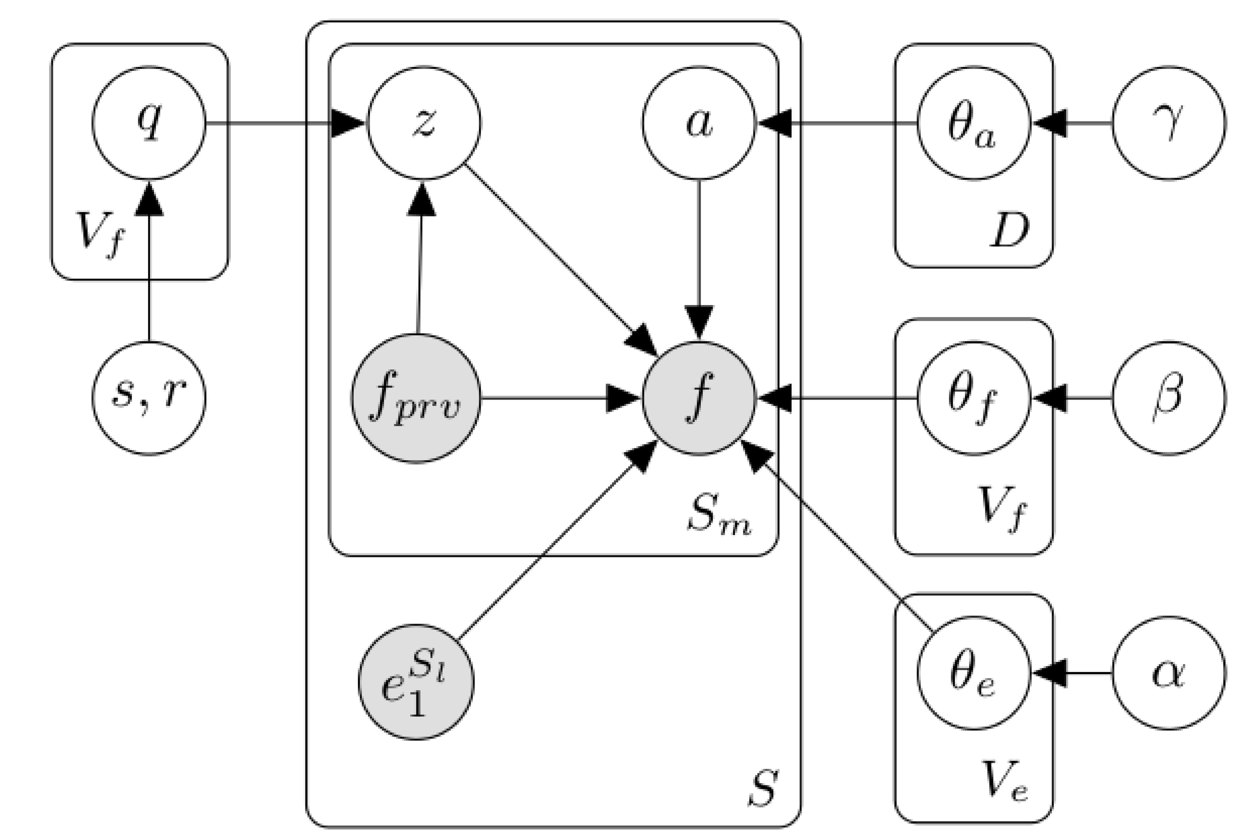

Figure : Graphical representation of our model.

Experiments

Experiments were performed on WMT 2014 news data.

Model	En-De	En-Fr	En-Cs	En-Ja	De-En	Fr-En	Cs-En	Ja-En
Brown et al. (model 1)	15.14	26.86	14.88	28.36	19.36	27.41	20.07	25.26
Brown et al. (model 2)	+0.84	+0.77	+1.14	+3.02	+1.80	+1.77	+1.15	+2.95
Mermer et al. (model 1)	+0.52	+0.80	+1.30	+3.19	+1.51	+1.60	+1.77	+2.44
Mermer et al. (model 2)	+0.63	+0.33	+1.94	+3.00	+2.02	+1.22	**+2.34**	+2.48
This work (model 1)	+0.39	+0.23	+1.31	**+3.33**	+1.61	+0.98	+1.87	+2.56
This work (model 2)	**+1.07**	**+1.47**	**+2.08**	+2.65	**+2.30**	**+2.19**	+2.13	**+3.21**
Giza	+1.59	+0.87	+1.70	+4.24	+2.54	+2.08	+2.36	+3.94
fastAlign	+1.39	+1.23	+1.87	+2.47	+2.44	+2.06	+2.21	+3.58

Table : Symmetrised: alignments obtained in both directions independently and heuristically symmetrised (grow-diag-final-and).

Future Work

- Extend the alignment component to HMM
- Symmetric alignments
- Develop variational algorithm
- Use hierarchical language model
- Experiment with hierarchical SMT