

# Language Model Interpolation

September 14, 2013

# Linear Interpolation

$$p_{\text{interp}}(\text{one} \mid \text{of}) = \lambda p_1(\text{one} \mid \text{of}) + (1 - \lambda) p_2(\text{one} \mid \text{of})$$

## One Implementation

- Store both models in RAM
- Query both, mix probabilities

# Linear Interpolation

$$p_{\text{interp}}(\text{one} \mid \text{of}) = \lambda p_1(\text{one} \mid \text{of}) + (1 - \lambda) p_2(\text{one} \mid \text{of})$$

## One Implementation

- Store both models in RAM
- Query both, mix probabilities

Can we use less RAM?

## SRILM claims to:

- 1  $p_{\text{interp}}(w | h) = \lambda p_1(w | h) + (1 - \lambda)p_2(w | h)$

## SRILM claims to:

- 1  $p_{\text{interp}}(w | h) = \lambda p_1(w | h) + (1 - \lambda)p_2(w | h)$
- 2 Produce an ARPA file to be queried in the normal way

## SRILM claims to:

- 1  $p_{\text{interp}}(w | h) = \lambda p_1(w | h) + (1 - \lambda)p_2(w | h)$
- 2 Produce an ARPA file to be queried in the normal way
- 3 Store the union of  $n$ -grams from  $p_1$  and  $p_2$

## SRILM claims to:

- 1  $p_{\text{interp}}(w | h) = \lambda p_1(w | h) + (1 - \lambda) p_2(w | h)$
- 2 Produce an ARPA file to be queried in the normal way
- 3 Store the union of  $n$ -grams from  $p_1$  and  $p_2$

Impossible

## SRILM **actually implements:**

- 1  $p_{\text{interp}}(w | h) \neq \lambda p_1(w | h) + (1 - \lambda)p_2(w | h)$
- 2 Produce an ARPA file to be queried in the normal way
- 3 Store the union of  $n$ -grams from  $p_1$  and  $p_2$



# Impossible

If “is one” is in any model, store  $p_{\text{interp}}(\text{one} \mid \text{is})$ .

$\implies$  All probabilities are fixed.

# Impossible

If “is one” is in any model, store  $p_{\text{interp}}(\text{one} \mid \text{is})$ .

⇒ All probabilities are fixed.

Model sums to one

⇒ Backoffs are determined from probabilities and fixed.

# Impossible

If “is one” is in any model, store  $p_{\text{interp}}(\text{one} \mid \text{is})$ .

$\implies$  All probabilities are fixed.

Model sums to one

$\implies$  Backoffs are determined from probabilities and fixed.

Query “is periwinkle”

$p_{\text{interp}}(\text{periwinkle} \mid \text{is})$

$= p_{\text{interp}}(\text{periwinkle}) b_{\text{interp}}(\text{is})$

$\neq \lambda p_1(\text{periwinkle}) b_1(\text{is}) + (1 - \lambda) p_2(\text{periwinkle}) b_2(\text{is})$

# Perplexity

Interpolate Europarl and Multi-UN.

1000 tuning and 1000 evaluation sentences held out from Europarl.

SRI Fake Interpolate	41.9301
SRI Dynamic Interpolate	42.0268
Manually Interpolate	41.1426

# What Now?

- Can we find a non-ARPA but efficient way?
- Implement the broken version anyway?
- Convince Philipp Koehn to drop SRI from Moses.