




# Pre-training with monolingual data

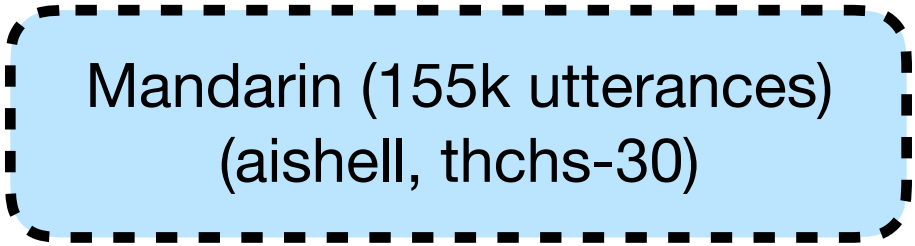


20

**Pre-train network with additional monolingual text to improve upon the monolingual stretches in the code-switched text**



English (250k utterances)  
(switchboard)



Mandarin (155k utterances)  
(aishell, thchs-30)

Perplexity	Data	RNNLM	DLM
With pre-training	Dev	74.06	72.41
	Test	61.66	60.73

# Pre-training with monolingual data

Pre-train network with additional monolingual text to improve upon the monolingual stretches in the code-switched text

English (250k utterances)  
(switchboard)

Mandarin (155k utterances)  
(aishell, thchs-30)

Perplexity	Data	RNNLM	DLM
With pre-training	Dev	74.06	72.41
	Test	61.66	60.73



# Code-switched Text Generation

---