

In-Domain



Fixed
Vocabulary

LM_{IN}

LM_{GEN}

CE_{IN}

CE_{GEN}

$$\text{DiffCE} = \text{CE}_{\text{IN}} - \text{CE}_{\text{GEN}}$$

General-Domain

$\text{Train}_{\text{GEN}}$

Test_{GEN}

Bin 1

Bin 2

Bin 3

Test_{GEN}

Bin N - 2

Bin N - 1

Bin N

Test_{GEN}

$\max(\text{diffCE})$
Test set closest to
in-domain data

Extraction of:
Grammar
Phrase Table
Paraphrase Table

Domain-Adapted Paraphrase
Table