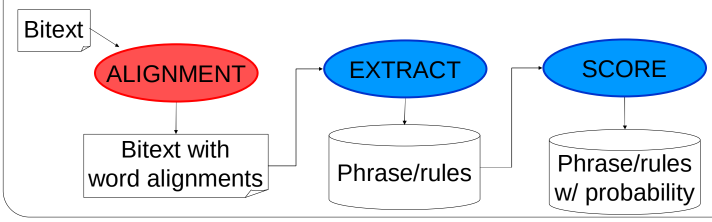
Analysis of Translation Model Adaptation

Kevin Duh, Katsuhito Sudoh, Hajime Tsukada
NTT Communication Science Laboratories, Kyoto, Japan
NTT

MOTIVATIONAL QUESTION:

Where and how does additional out-domain bitext help in the MT training pipeline?



FINDINGS:

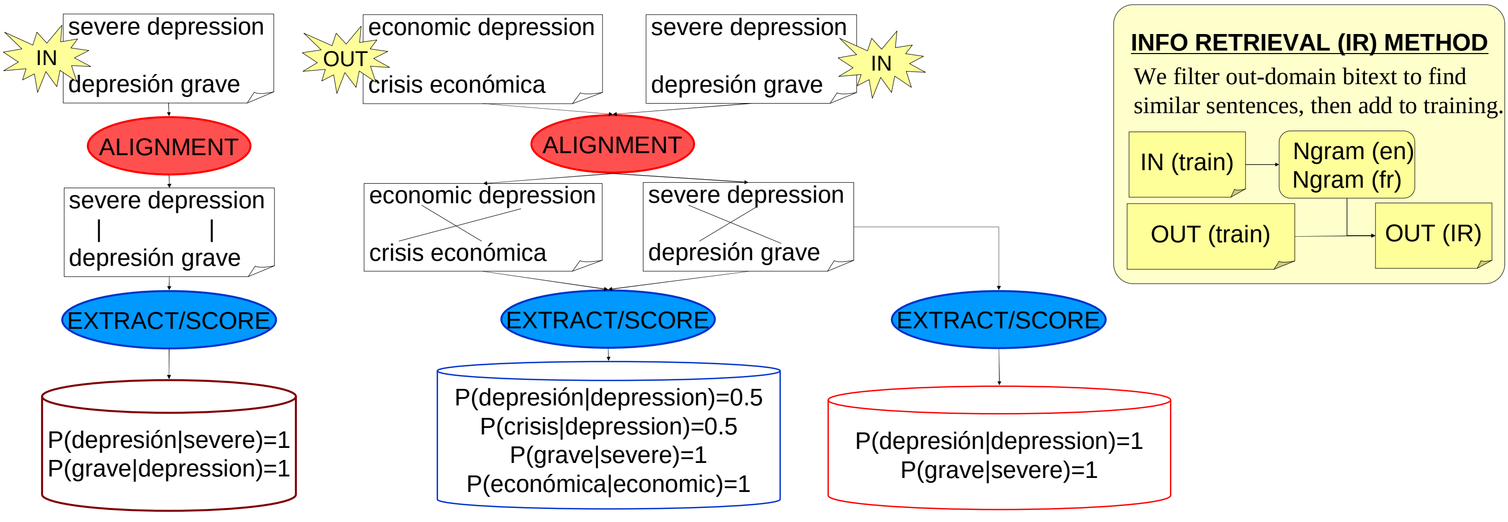
1. Out-domain bitext has different effects on word alignment (changes phrases units & probabilities) vs. phrase extraction (also decrease OOV, increase translation options)

2. Sometimes it's better to use out-domain data in only part of the training pipeline, e.g.:

Medicine: if you have <u>severe **depression**</u> // si padece una <u>**depresión** grave</u> IN
Parliament: <u>economic **depression**</u> in Europe //<u>**crisis** económica</u> en Europa OUT

ANALYSIS TECHNIQUE:

Compare systems where out-domain data is inserted to <u>partial or full</u> training pipeline



INFO RETRIEVAL (IR) METHOD

We filter out-domain bitext to find similar sentences, then add to training.

System1: IN **System2: IN+IR (Full Pipeline)** **System3: IN+IR (Align Only)**

EXPERIMENT 1: TED TALKS

Task: Improve TED translation (IN) using out-domain bitext (Europarl + News + UN corpora)

All systems use: Moses decoder, grow-diag-final-and, 4gram, MERT

Results:
- Using out-domain data for full pipeline improves. (22.04→22.66)
- Using it for Alignment Only improves even more! (22.04→23.28)

	IN	IN+IR Full Pipeline	IN+IR Align Only
BLEU (~700 test sentences)	**22.04**	**22.66**	**23.28**
Train Size for Alignment (#sent)	84k	307k	307k
Train Size for Extract (#sent)	84k	307k	84k
#Alignment Links per Sentence	11.46	21.61	11.19
Phrase Table Size (#entries)	1.8M	15.7M	1.9M
Out-of-vocabulary rate	2.5%	1.5%	2.3%

<u>Detailed BLEU Analysis:</u>

40% of correct ngrams unique to IN+IR(AlignOnly) are not present in IN phrase table → new in-domain phrases

68% of incorrect ngrams unique to IN+IR(FullPipeline) are not present in IN bitext → extraneous translation options

EXPERIMENT 2: TEN LANGUAGE PAIRS

Large-scale evaluation on 4 corpora and 10 language pairs: (da, de, el, es, fi, fr, it, nl, pt, sv) → en

All systems use: Moses decoder, grow-diag-final-and, 3gram, MERT

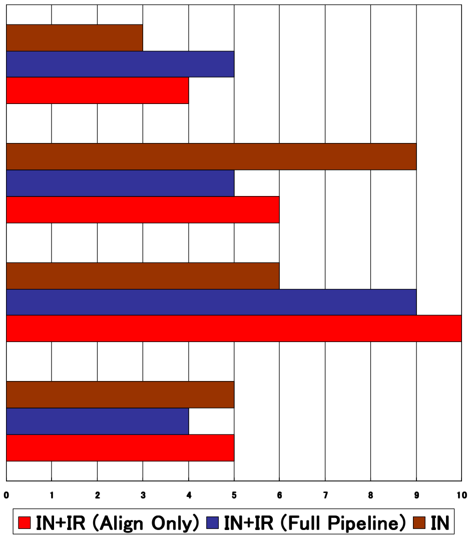
Mixed Results—Number of times a system is best or within 0.2 BLEU (out of 10 language pairs):

<u>KDE (computer)</u>
Out=Europarl
#sent: IN=83k/IR=37k
oov: 6.6→4.3%

<u>EMEA (medicine)</u>
Out=Europarl
#sent: IN=821k /IR=197k
oov: 2.8→2.3%

<u>EUROPARL (parliament)</u>
Out=EMEA
#sent: IN=1210k / IR=127k
oov: 0.6→0.5%

<u>OPENSUBTITLE (movie)</u>
Out=Europarl
#sent: IN=208k / IR=109k
oov: 7.4→3.4%


🟥 IN+IR (Align Only) 🟦 IN+IR (Full Pipeline) 🟫 IN