

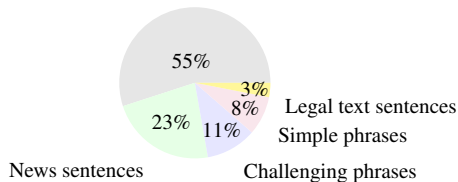


In-domain Data FTW

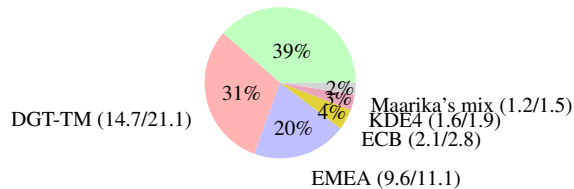
Mark Fishel



Simple sentences



JRC-Acquis (18.4/25.0)



General domain-tuned, general domain-tested:

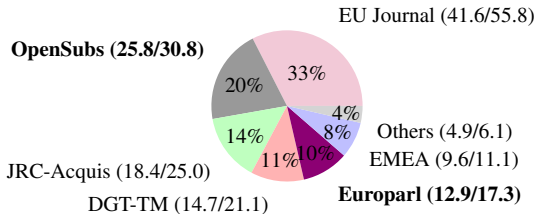
	BLEU	METEOR	TER	Length (%)
GIZA++ <i>diag</i>	56.1 / 57.7	43.3 / 44.0	36.0 / 36.0	97.8 / 100.3
GIZA++ <i>inter</i>	55.7 / 62.9	42.1 / 45.9	34.5 / 33.6	86.0 / 101.2
Berkeley	60.2 / 63.4	45.1 / 46.5	31.1 / 30.9	93.7 / 99.7

General domain-tuned, **in-domain**-tested:

	BLEU	METEOR	TER	Length (%)
GIZA++ <i>diag</i>	14.1 / 13.7	24.1 / 24.3	67.3 / 71.0	88.3 / 94.0
GIZA++ <i>inter</i>	11.8 / 10.2	22.0 / 22.4	67.6 / 85.0	77.9 / 107.9
Berkeley	14.1 / 13.8	24.2 / 24.6	65.2 / 70.2	82.2 / 93.3

In-domain tuned and tested (cross-validation):

	BLEU	METEOR	TER	Length (%)
GIZA++ <i>diag</i>	14.1 / 16.0	24.1 / 25.7	67.3 / 68.0	88.3 / 96.8
GIZA++ <i>inter</i>	11.8 / 14.6	22.0 / 24.3	67.6 / 71.5	77.9 / 98.4
Berkeley	14.1 / 16.4	24.2 / 25.9	65.2 / 69.1	82.2 / 99.0



TMCombine:

Without in-domain-like data:

	BLEU	METEOR	TER
Uniform	16.0	25.7	68.0
Optimized	17.2	26.5	66.7

With in-domain-like data:

	BLEU	METEOR	TER
Uniform	27.5	33.6	54.7
Optimized	27.4	33.6	55.7

	BLEU	METEOR	TER
masintolge.ut.ee	27.5	33.6	54.7
Google Translate	29.3	34.4	51.1
Bing Translator	24.9	31.2	57.7

masintolge/Google:

src: teleka pult
ref: tv remote control
masintolge: the tv remote
Google: tv remote control
Bing: teleka remote control

	BLEU	METEOR	TER
Simple sentences	31.3 / 31.7	36.9 / 36.5	48.4 / 48.0
Simple phrases	27.4 / 31.5	34.0 / 34.7	54.6 / 54.3
News, Legalese	13.9 / 21.1	26.8 / 28.9	72.5 / 62.1
Proverbs, Swearing, etc.	17.7 / 18.0	26.8 / 27.5	61.0 / 63.1

src: mul on paha
ref: i feel sick
masintolge: i have bad
Google: i have a bad
Bing: i have a bad

src: kas sa šokolaadi tahad ?
ref: do you want a chocolate ?
masintolge: do you want the chocolate ?
Google: do you want chocolate ?
Bing: do you want a chocolate ?

src: mine metsa kuradi tõlge
ref: go to hell , bloody translation
masintolge: go to the forest fucking translation
Google: translation of the devil in the forest
Bing: go to the translation of forest fucking