



SYSTRAN
beyond language

Boosting Neural Machine Translation with Similar Translations

AMTA 2022

Specialization: highest quality translations for your content

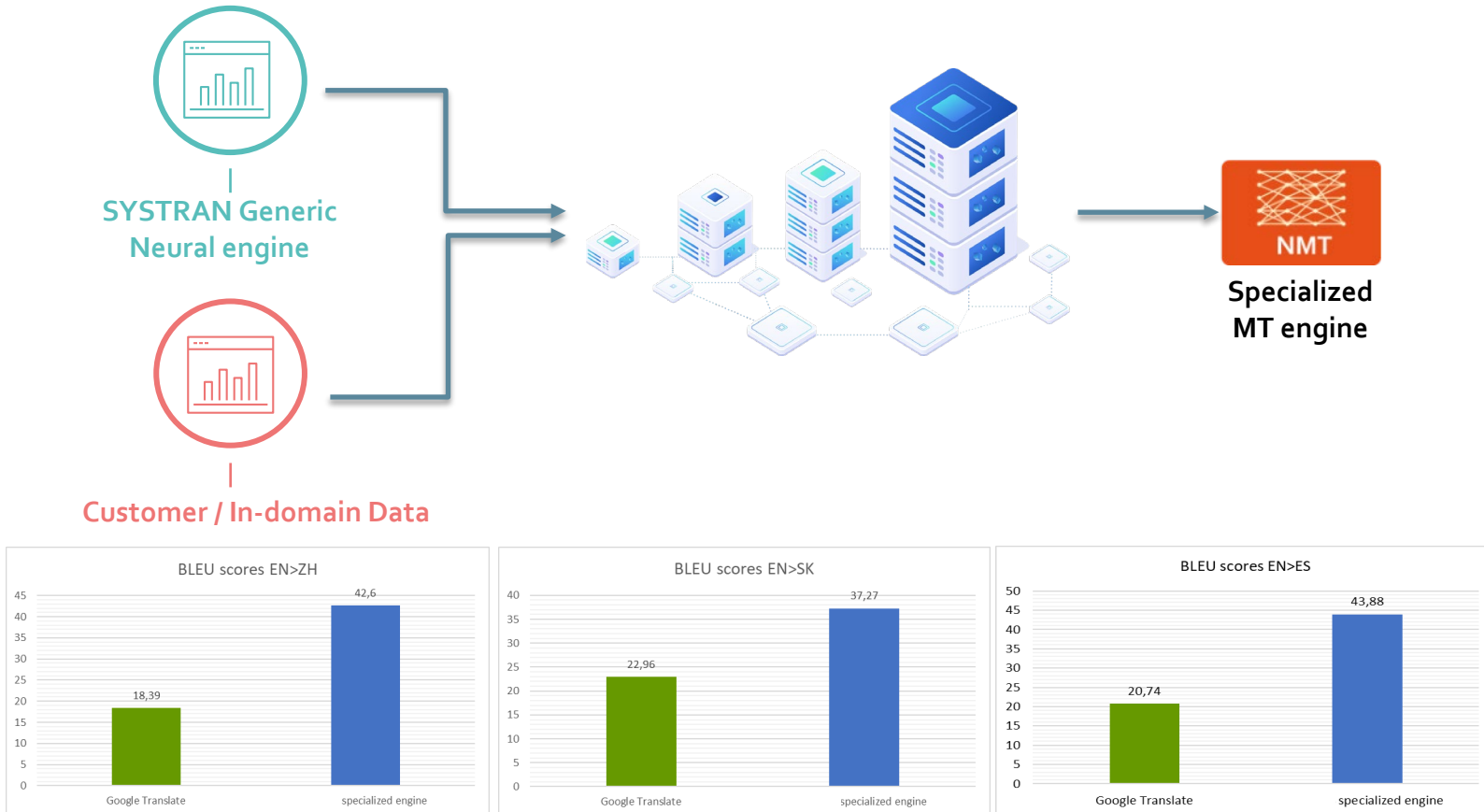
User Dictionaries & Translation memories

Linguistic Translation Profiles

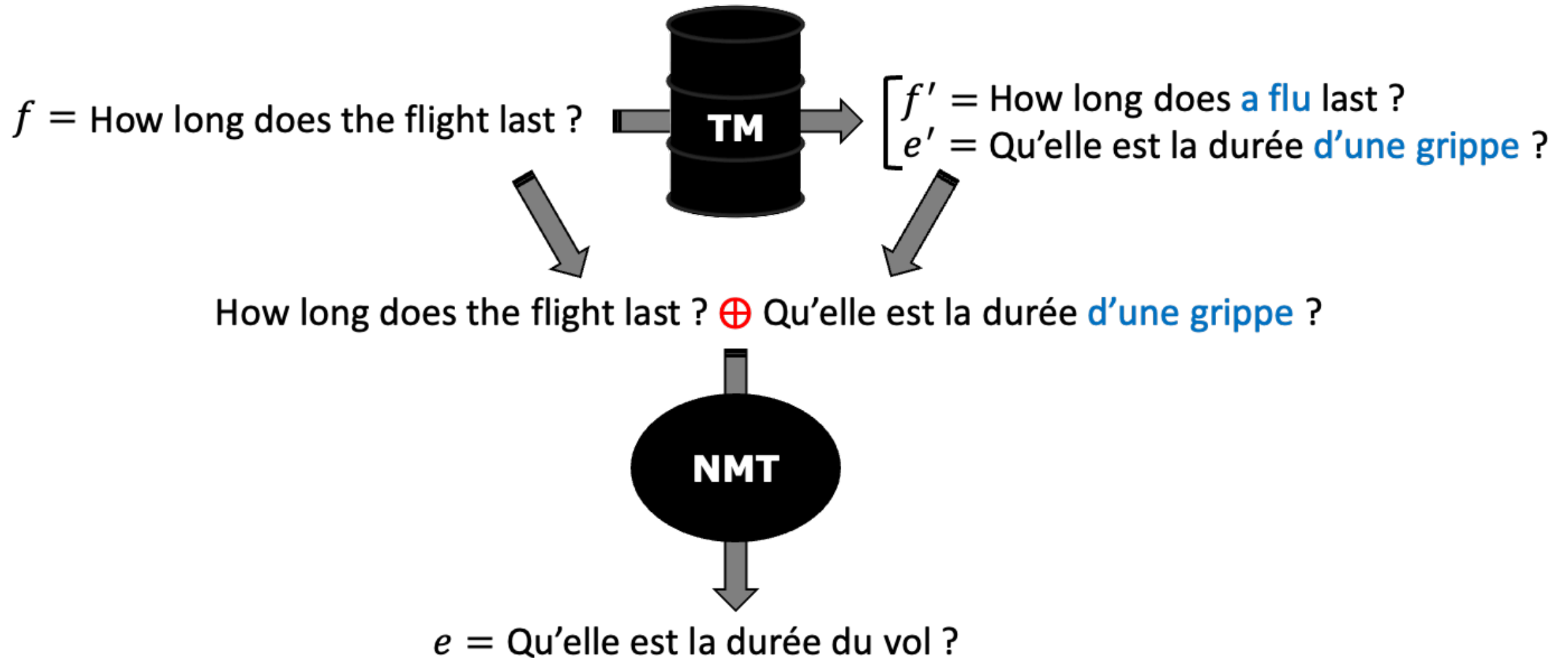
Feedback loop for continuous improvement

💡 As the pioneer in Machine Translation, we introduce new features on a regular basis. For example, in 2020 we introduced the ability to train multi-domain engines. This significantly improves the translation quality for low resource domains and allows you to address various business cases with the same neural engine.

Specialization of the neural MT engine using your own data

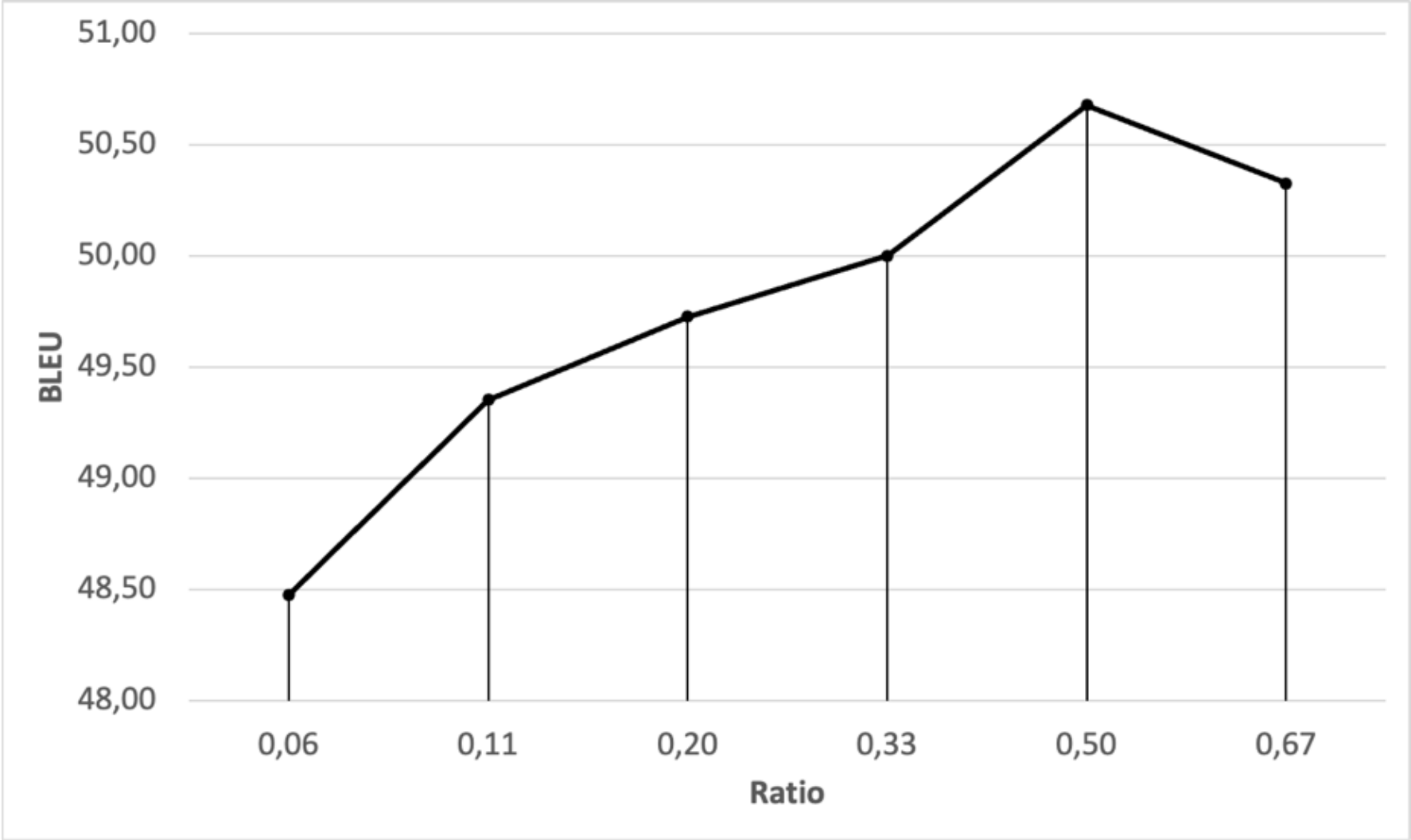


NFA Workflow - Overview



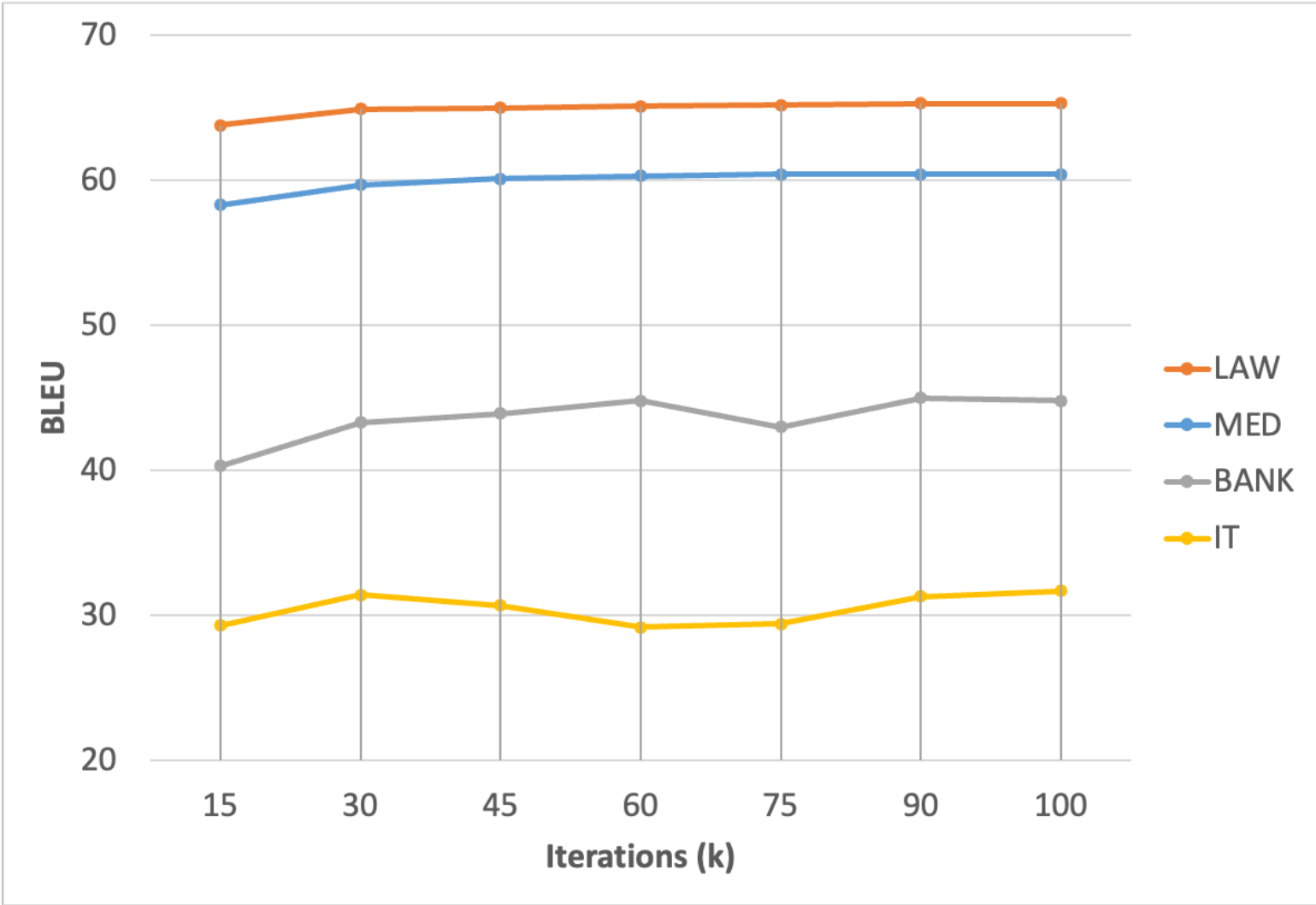


How many augmented examples must be used in training?



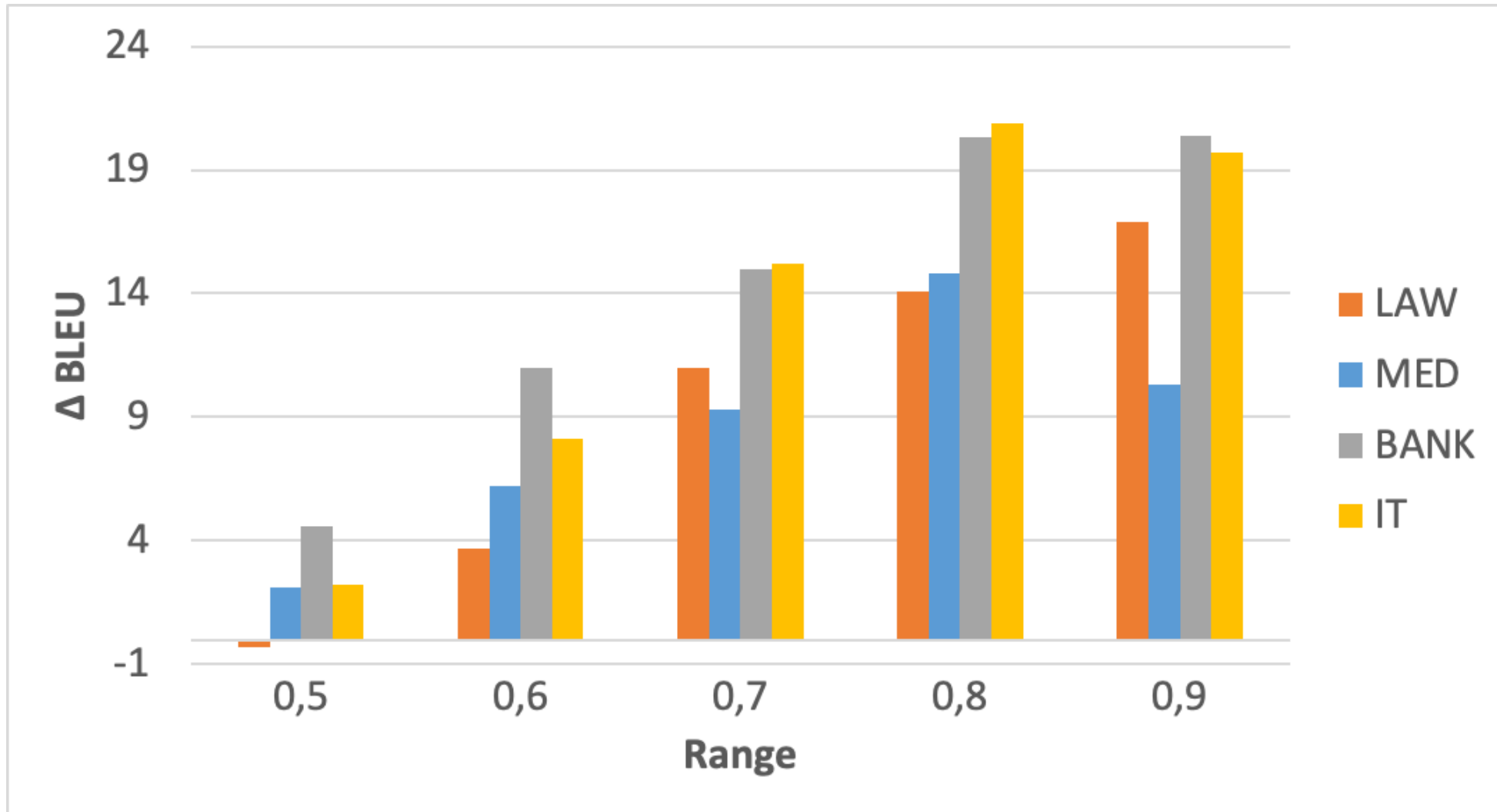


How many training iterations are needed to effectively teach a regular model to use similar translations?



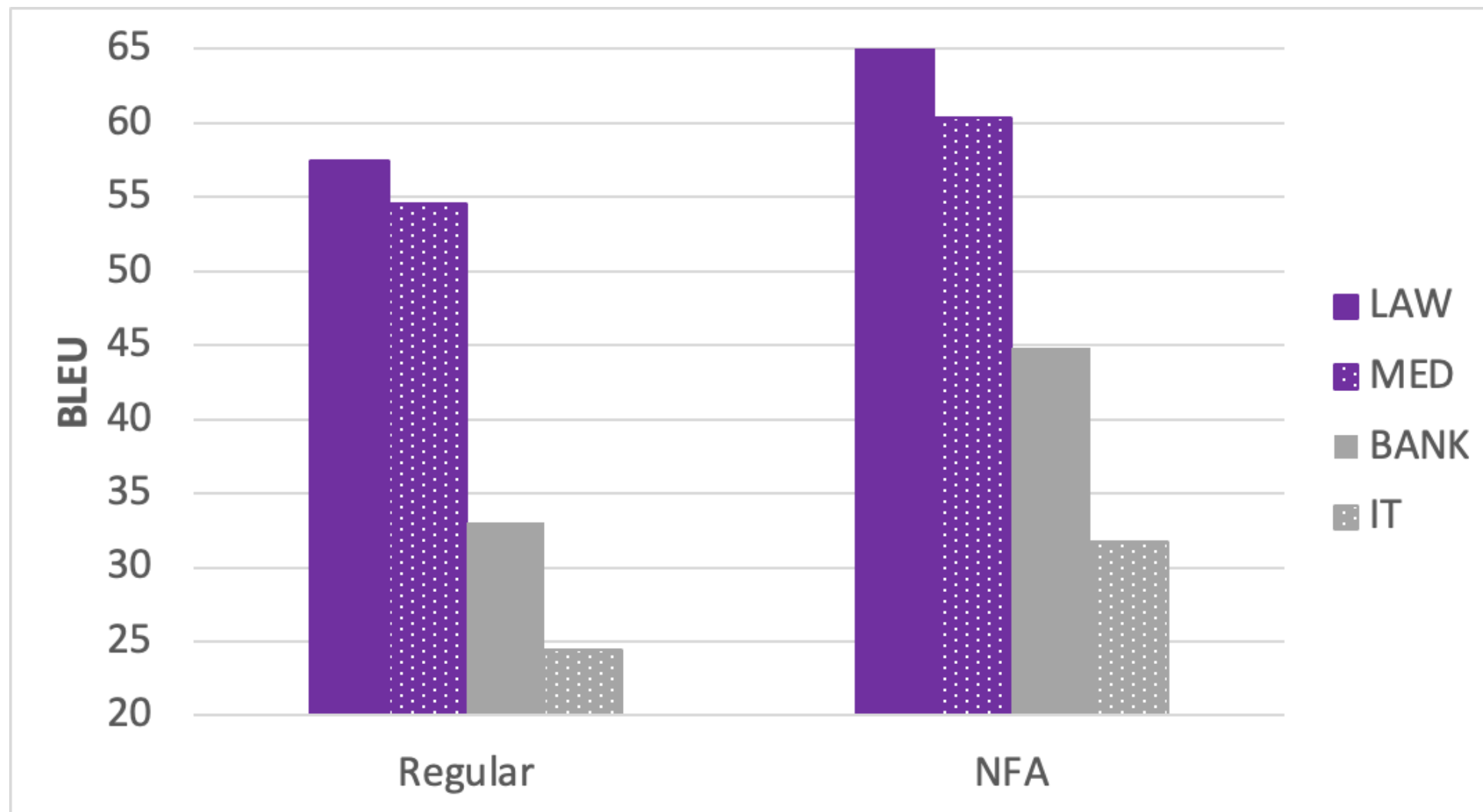


How “similar” must similar sentences be?





Translating with out-domain translation memories





Qualitative Analysis

Input sentence augmented with a fuzzy match and match/non-match information.

f = How long does the flight last ?

f' = How long does a flu last ?

e' = Qu'elle est la durée d'une grippe ?

How_{|S} long_{|S} does_{|S} the_{|S} flight_{|S} last_{|S} ?_{|S} \oplus _{|B} Qu'elle_{|M} est_{|M} la_{|M} durée_{|M} d'une_{|NM} grippe_{|NM} ?_{|M}



Automated quality assessment




Automatic measure evaluation.

	BLEU ↑	METEOR ↑	TER ↓	chrF ↑	BERTScore ↑	COMET ↑	DTED ↓
TM	44.49	56.10	60.26	59.67	84.86	14.46	63.09
Baseline NMT	58.87	75.53	29.66	75.52	92.56	78.33	29.76
NFR	66.07	79.45	25.25	79.33	93.49	81.79	26.33



Automated quality assessment

BLEU scores for the TM, baseline NMT and NFA system in different FM ranges.

	All	50–59%	60–69%	70–79%	80–89%	90–99%
TM	44.49	5.18	10.77	23.52	43.99	75.25 
Baseline NMT	58.87	34.64	42.94	50.24	59.09	74.64
NFR 	66.07	35.97	43.63	53.93	67.29	85.24 

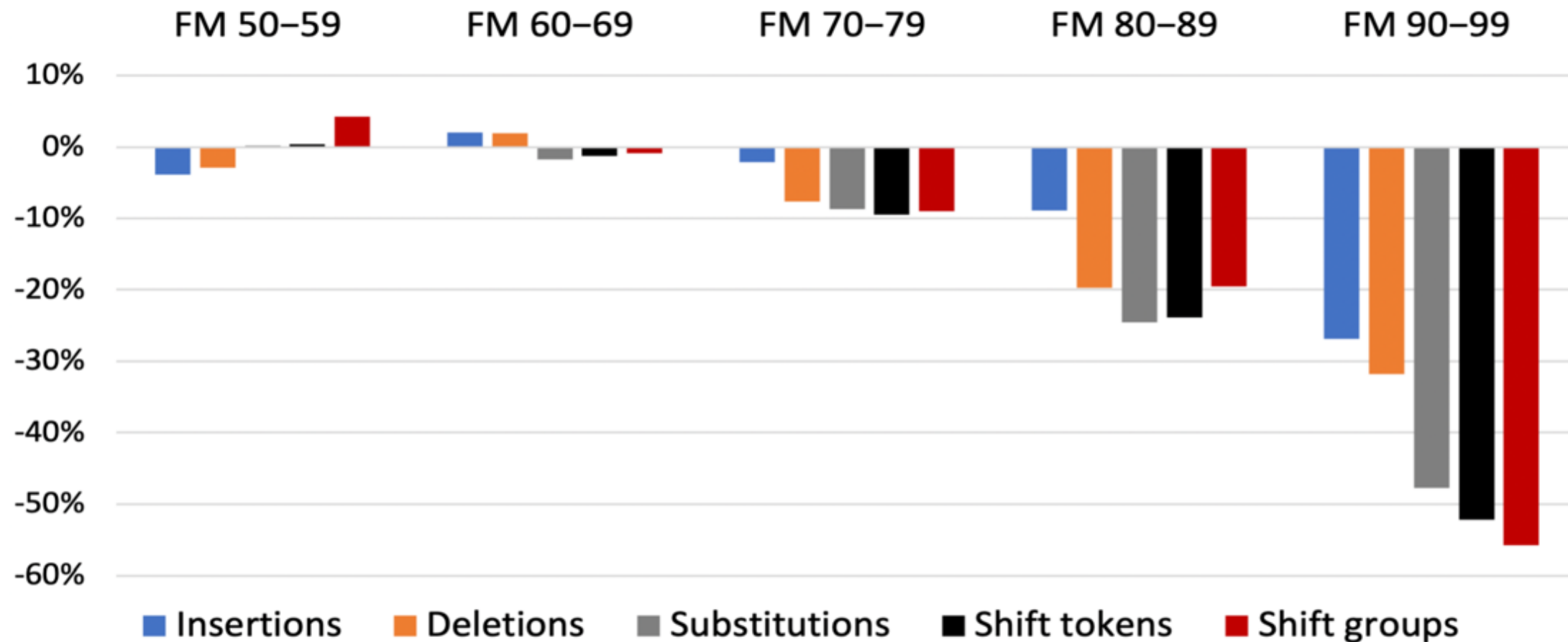
TER: Edit operations

		Baseline NMT		NFR		NMT vs. NFR
		Total	/Sent	Total	/Sent	% Difference
Insertions	Content words	4747	0.76	4239	0.68	-10.70
	Function words	5527	0.89	4966	0.80	-10.15
	Other	1114	0.18	1089	0.18	-2.24
	<i>Total</i>	<i>11,388</i>	<i>1.83</i>	<i>10,294</i>	<i>1.66</i>	<i>-9.61</i>
Deletions	Content words	3063	0.49	2677	0.43	-12.60
	Function words	3609	0.58	3102	0.50	-14.05
	Other	475	0.08	431	0.07	-9.26
	<i>Total</i>	<i>7147</i>	<i>1.15</i>	<i>6210</i>	<i>1.00</i>	<i>-13.11</i>
Substitutions	Content words	12,258	1.97	10,146	1.63	-17.23
	Function words	8511	1.37	7030	1.13	-17.40
	Other	919	0.15	832	0.13	-9.47
	<i>Total</i>	<i>21,688</i>	<i>3.49</i>	<i>18,008</i>	<i>2.90</i>	<i>-16.97</i>
Token shifts	Content words	3253	0.52	2651	0.43	-18.51
	Function words	4552	0.73	3610	0.58	-20.69
	Other	415	0.07	352	0.06	-15.18
	<i>Total</i>	<i>8220</i>	<i>1.32</i>	<i>6613</i>	<i>1.07</i>	<i>-19.55</i>
Total token-level edits		48,443	7.80	41,125	6.62	-15.10
Group shifts		2255	0.36	1832	0.30	-18.76



NFA produced translations that required fewer edit operations overall

Percentage of difference in total number of edits, per edit type, and different FM ranges between NFA and baseline NMT output.



Fine-Grained Error Analysis

Number of errors per category of the SCATE error taxonomy, for the baseline NMT and NFA systems.

	BASE	NFR		BASE	NFR
FLUENCY	89	71	ACCURACY	110	128
• Coherence	20	11	• Mistranslation	58	57
◦ Logical problem	18	9	◦ Multiword expression	10	5
◦ Co-reference	—	—	◦ Semantically unrelated	12	27
◦ Cultural reference	—	—	◦ Word sense	14	8
◦ Verb tense	2	2	◦ Part-of-speech	2	1
◦ Inconsistency	—	—	◦ Partial	—	—
• Grammar and Syntax	8	15	◦ Other	20	16
◦ Extra word	2	1	• Omission	36	40
◦ Missing word	2	6	• Addition	8	25
◦ Multi-word syntax	—	—	• Do-not-translate	—	—
◦ Word order	2	2	• Untranslated	2	—
◦ Word form	2	6	• Mechanical	4	5
◦ Other	—	—	◦ Capitalization	—	—
• Lexicon	29	12	◦ Punctuation	4	5
◦ Lexical choice	17	9	◦ Other	—	—
◦ Non-existing/Foreign	7	1	• Other	2	1
◦ Wrong preposition	5	2			
• Orthography	2	4			
◦ Punctuation	2	1			
◦ Capitalization	—	—			
◦ Spelling	—	3			
◦ Other	—	—			
• Style and Register	30	29			
◦ Disfluent	26	26			
◦ Register	1	1			
◦ Repetition	3	2			
◦ Untranslated	—	—			



Impact of data augmentation

Analysis of match (m) and no-match (nm) words in FMs and whether they appeared in the MT output (pass) as well as in the reference translation (REF).

	NFR	Full Test Set Baseline NMT	50–59% NFR	90–99% NFR
Fuzzy match 1				
% m/FM1	62.2	—	26.6	82.7
% m-pass	88.4	82.6	74.5	95.8
% m-pass-REF	69.5	66.6	59.5	77.4
% nm/FM1	37.8	—	73.4	17.3
% nm-pass	36.5	31.9	23.5	45.8
% nm-pass-REF	29.6	27.3	17.4	38.7



NFA > NMT > TM

Improving translation quality with Neural Fuzzy Adaptation



Results showed that all automated metrics, which compared the MT output to a reference translation, showed higher scores for the NFA system than for an NMT baseline and confirmed the significant improvements it achieved.



A detailed TER analysis showed that the strengths of the NFA system are to produce translations with more similar lexical choices and word order than the reference translations.