

MODELING INTRA-WORD CODE-SWITCHING FOR KARELIAN ASR

Irina Kipyatkova, Kseniia Kiseleva, Mikhail Dolgushin, and Ildar Kagiroy

St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS)

{kipyatкова, kiseleva.k, dolgushin.m, kagiroy}@iias.spb.su



The aim of the research was to improve the accuracy of Karelian-Russian code-switching (CS) speech recognition, with special attention to intra-word CS phenomena.

The phenomenon of intra-word CS consists in morphological adaptation of Russian words to the Karelian language by the borrowing of a Russian word base and the addition of Karelian affixes. For example, a speaker can add Karelian affix 'an' to the Russian word 'училищ' ('college') to convert a word in an accusative case resulting in 'училищан' ('učiliščan' in Latin transcription).

Karelian Speech Data

Corpus features	Value	
	AnKaS	KarRusCoS
Type of Speech	Mostly formal (radio broadcasts)	Spontaneous
Speakers	17 (7 male, 10 female)	41 (16 male, 24 female)
Duration	4.5h	3 h
Utterances	4385	3012
Word occurrences	32037	22355
Unique words	9117	7091
Code-switching rate	1%	28%
Intra-word code-switching rate	<1%	6%
Training/development/test ratio	8:1:1	

Examples of the Developed Rules for Generating Words with the Intra-Word CS

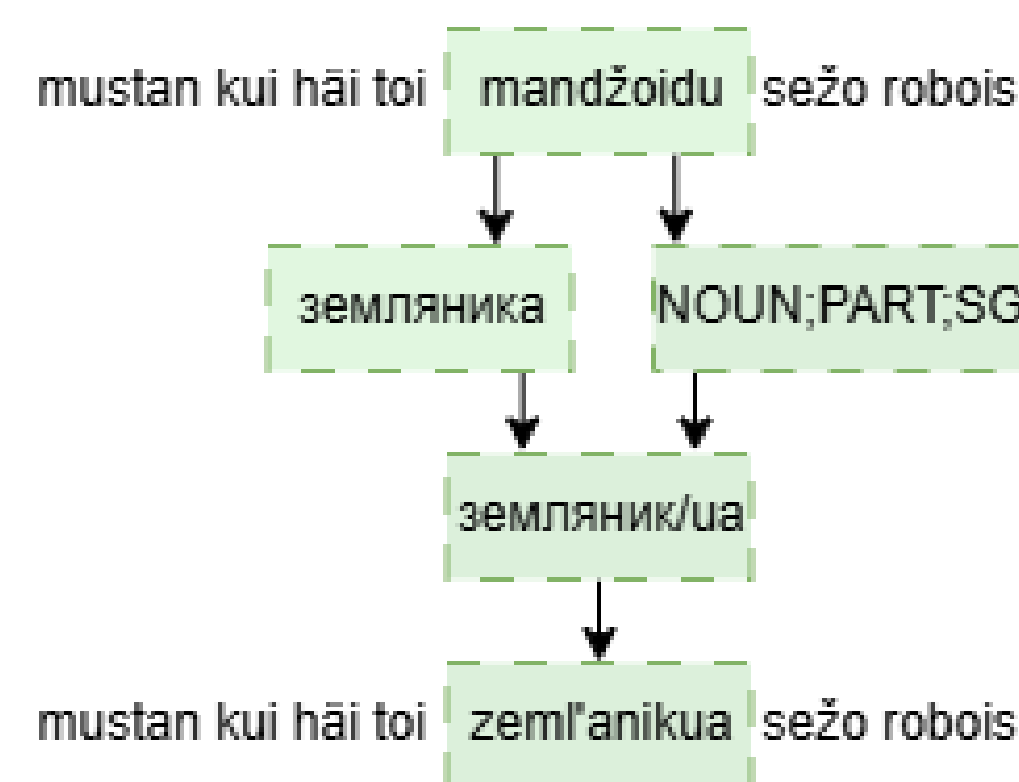
- Nouns:
 - formation of the inessive singular:
 - if the word is polysyllabic and ends in a diphthong, add *ies* to the stem (*карелия* → *каре^{lies}ies* (*kare^{lies}ies*));
 - in other cases, if the word ends in a soft consonant, add the ending *is* to the stem (*очередь* → *очеред^{is}is* (*očered^{is}is*));
 - in other cases, add the ending *as* to the stem (*город* → *город^{as}as* (*gorod^{as}as*)).
- Adjectives:
 - formation of the partitive singular:
 - add the ending *oidu* to the stem (*автобусный* → *автобусн^{oidu}oidu* (*avtobusn^{oidu}oidu*)).
- Verbs:
 - inflect the word into an imperative form;
 - if the imperative ends in a consonant, add *i* to the word in an imperative form;
 - for the reflexive verbs, drop the ending *сь* (*s'*) or *ся* (*s'a*);
 - formation of the present tense 1st person singular:
 - if the word in imperative form ends in *и* (*i*), add ending *mmo* (*береги* → *берег^{immo}mmo* (*bereg^{immo}mmo*));
 - in other case, if the word in imperative form ends in *й* (*j*) add ending *čemma* (*собирай* → *собирай^{čemma}emma* (*sobiraj^{čemma}emma*)).

In total, 33 rules for nouns, 6 rules for adjectives, and 17 rules for verbs were formulated.

Karelian Text Data

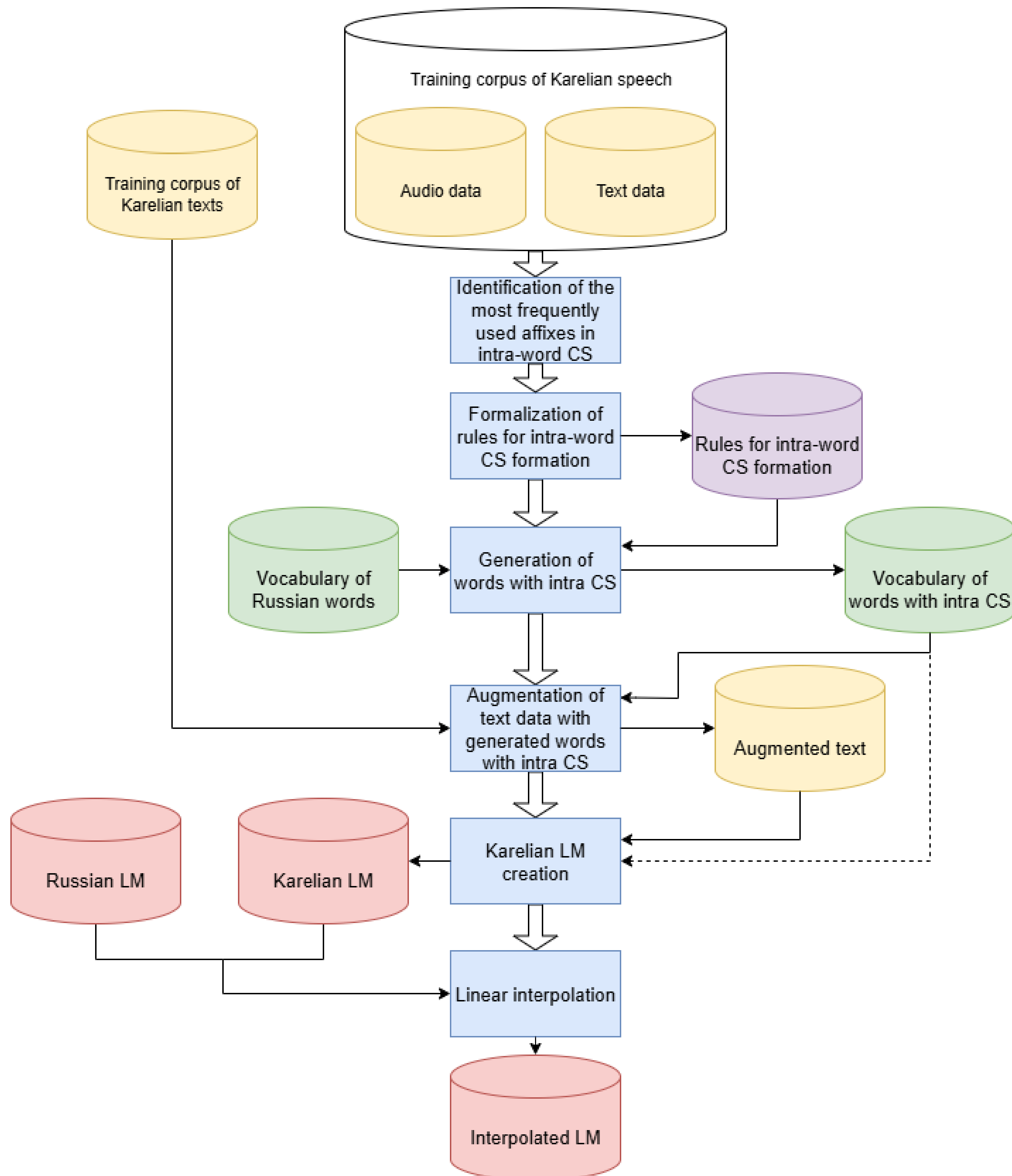
Parameter	Value
Volume	5M word occurrences
Content	Publications and periodicals in Livvi-Karelian, Livvi-Karelian part of VepKar
Vocabulary size (the number of words with frequency of appearance at least 2)	143,5 K words

An Example of the Sentence Augmentation



(I remember how he also brought wild strawberries in a birch-bark basket)

The Proposed Language Modeling Approach



Trained Language Models (LMs)

- LM1 - A LM trained solely on the original Karelian texts;
- LM2 - A LM1 linearly interpolated with the Russian LM;
- LM3 - A LM trained on the original Karelian texts with the entire vocabulary interpolated with the Russian LM;
- LM4 - A LM trained on the augmented Karelian texts (comprising only words that occurred in the augmented text), linearly interpolated with the Russian LM;
- LM5 - A LM trained on augmented Karelian texts with the entire vocabulary linearly interpolated with the Russian LM.

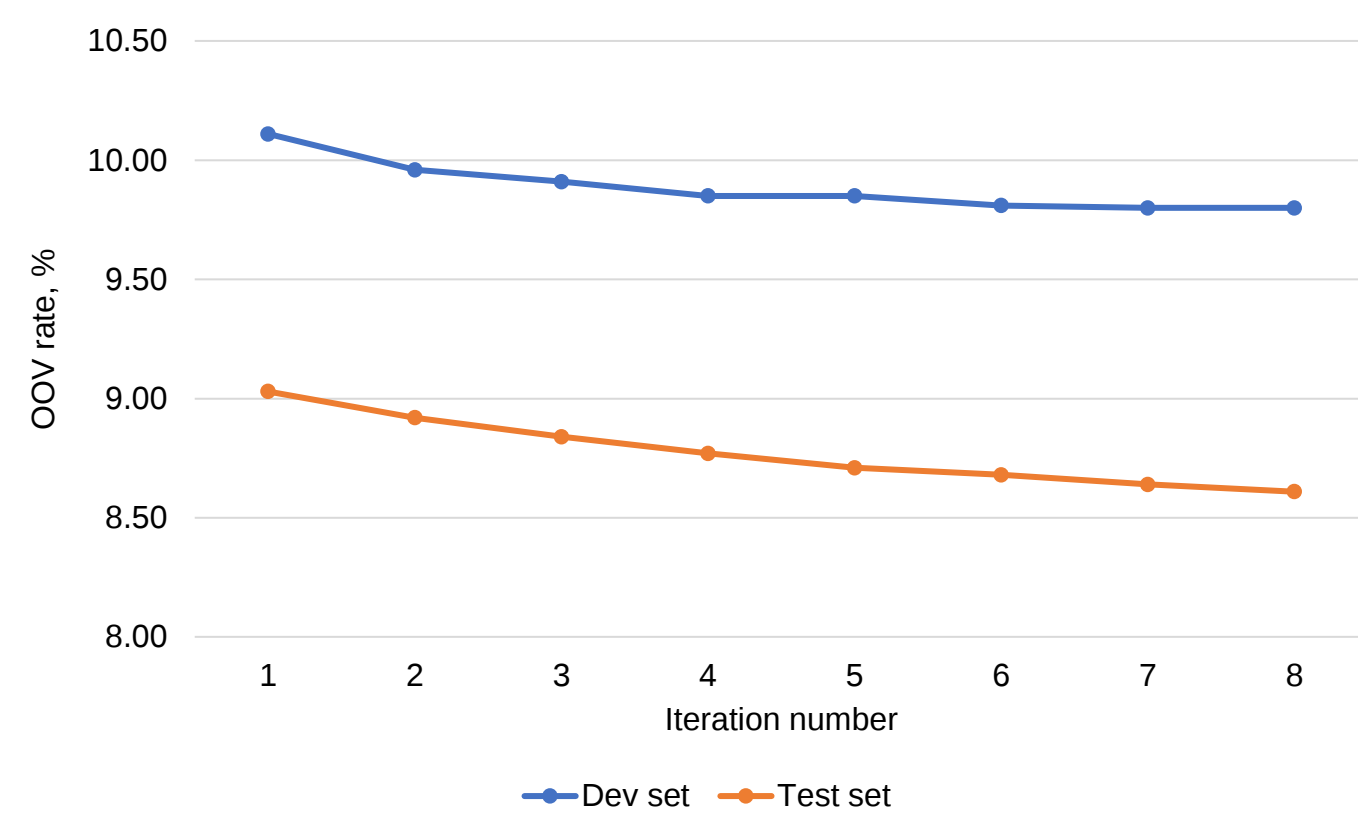
Characteristics of the Trained LMs

LM	Training text	Vocabulary	Interpolation with Russian LM
LM1	Original texts	Words from the original texts	No
LM2	Original texts	Words from the original texts	Yes
LM3	Original texts	Entire vocabulary	Yes
LM4	Augmented texts	Words from the augmented text	Yes
LM5	Augmented texts	Entire vocabulary	Yes

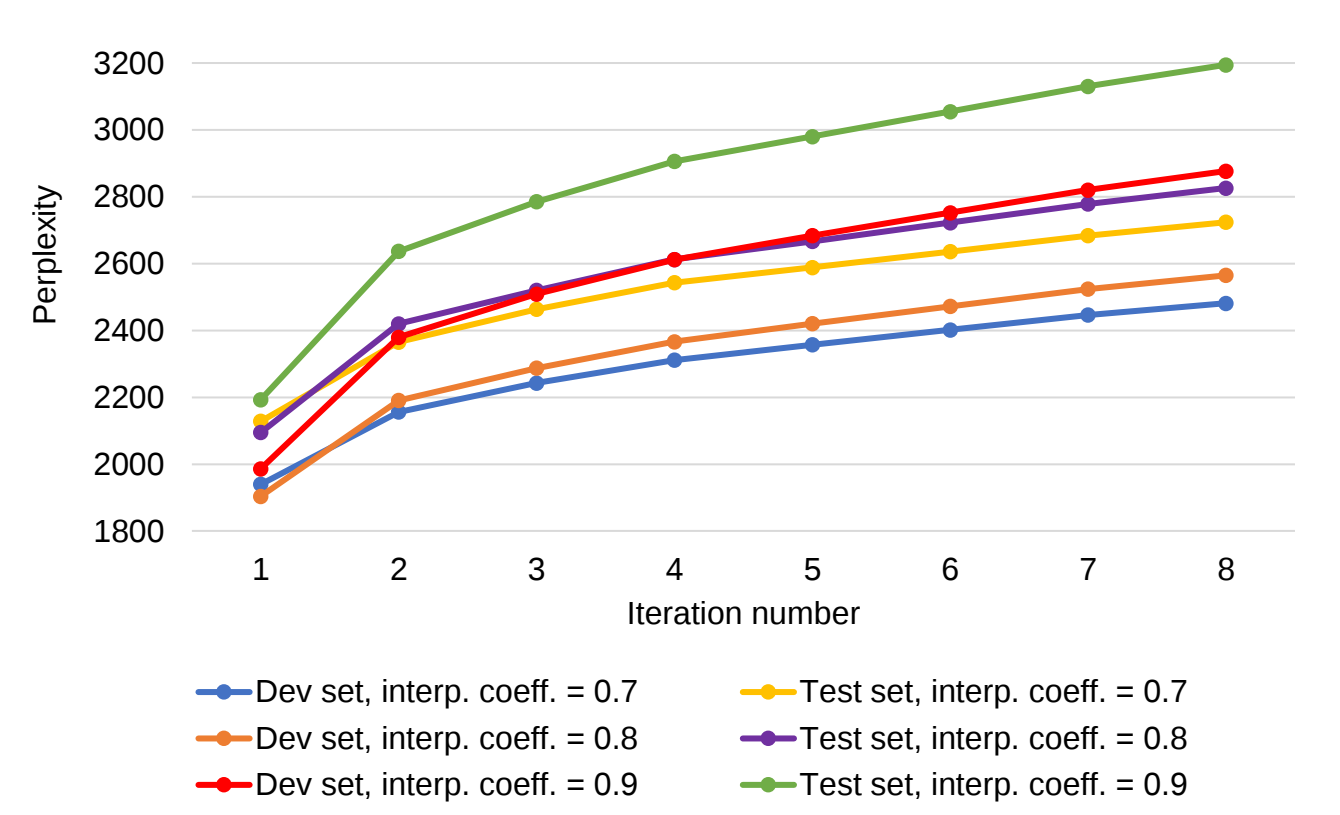
Perplexities and OOV Rates of Created LMs

LM	Vocabulary size, K	Dev set			Test set				
		OOV rate	Perplexity		OOV rate	Perplexity			
LM1	143	21.53	1476.18		20.88	1530.97			
		Interpolation coefficients			Interpolation coefficients				
			0.7	0.8	0.9		0.7	0.8	0.9
LM2	287	11.28	1753.6	1728.0	1812.7	9.93	1972.7	1947.5	2047.4
LM3	851	7.67	2518.3	2480.6	2596.1	7.04	2618.3	2584.2	2712.3
LM4	309	10.11	1939.8	1904.0	1986.4	9.03	2128.8	2095.2	2193.3
LM5	851	7.67	2465.9	2418.8	2518.9	7.04	2572.5	2529.8	2643.6

Dependency of the OOV Rate on the Number of Iterations



Dependency of the Perplexity on the Number of Iterations

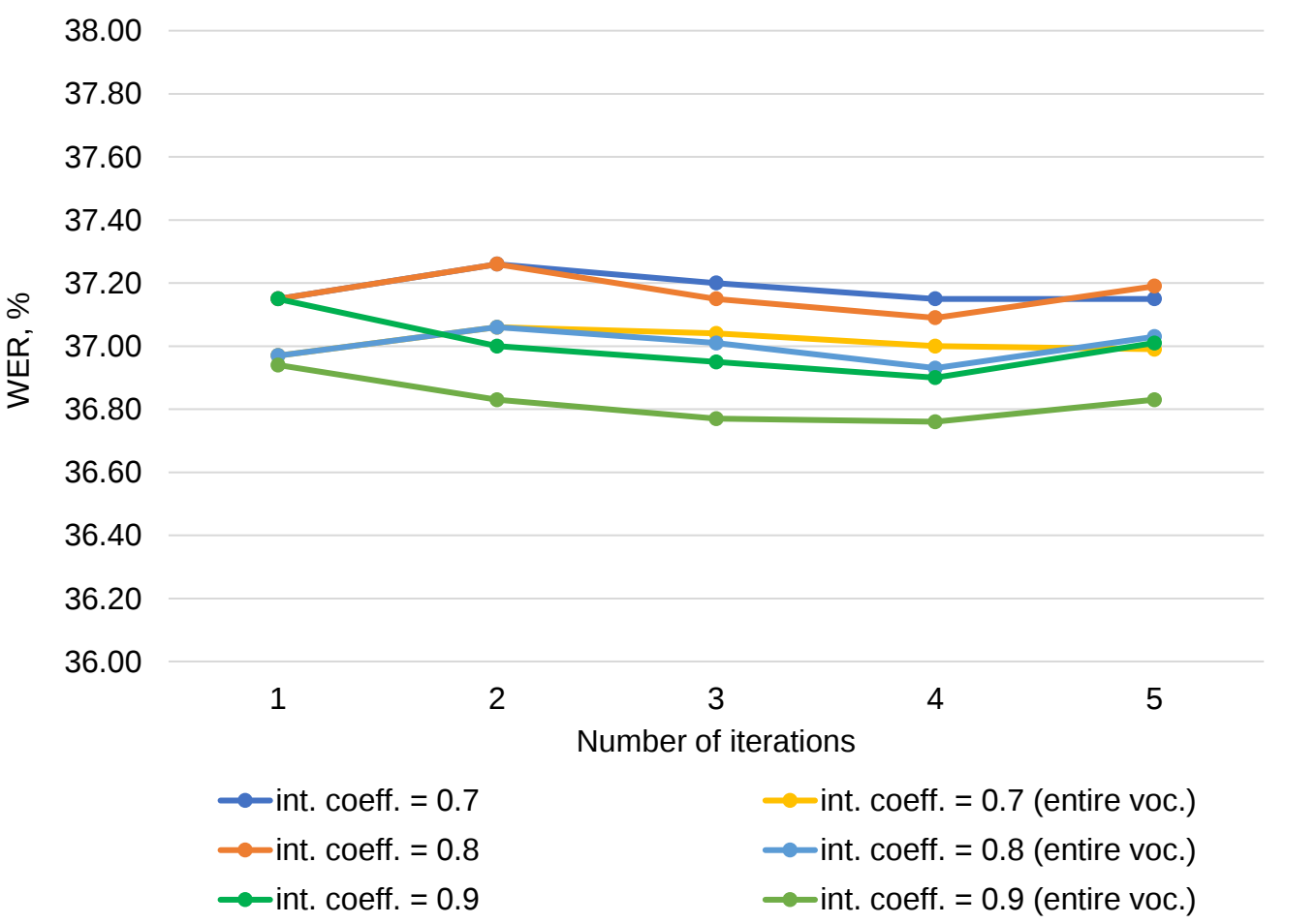


Experimental Results of Karelian Speech Recognition

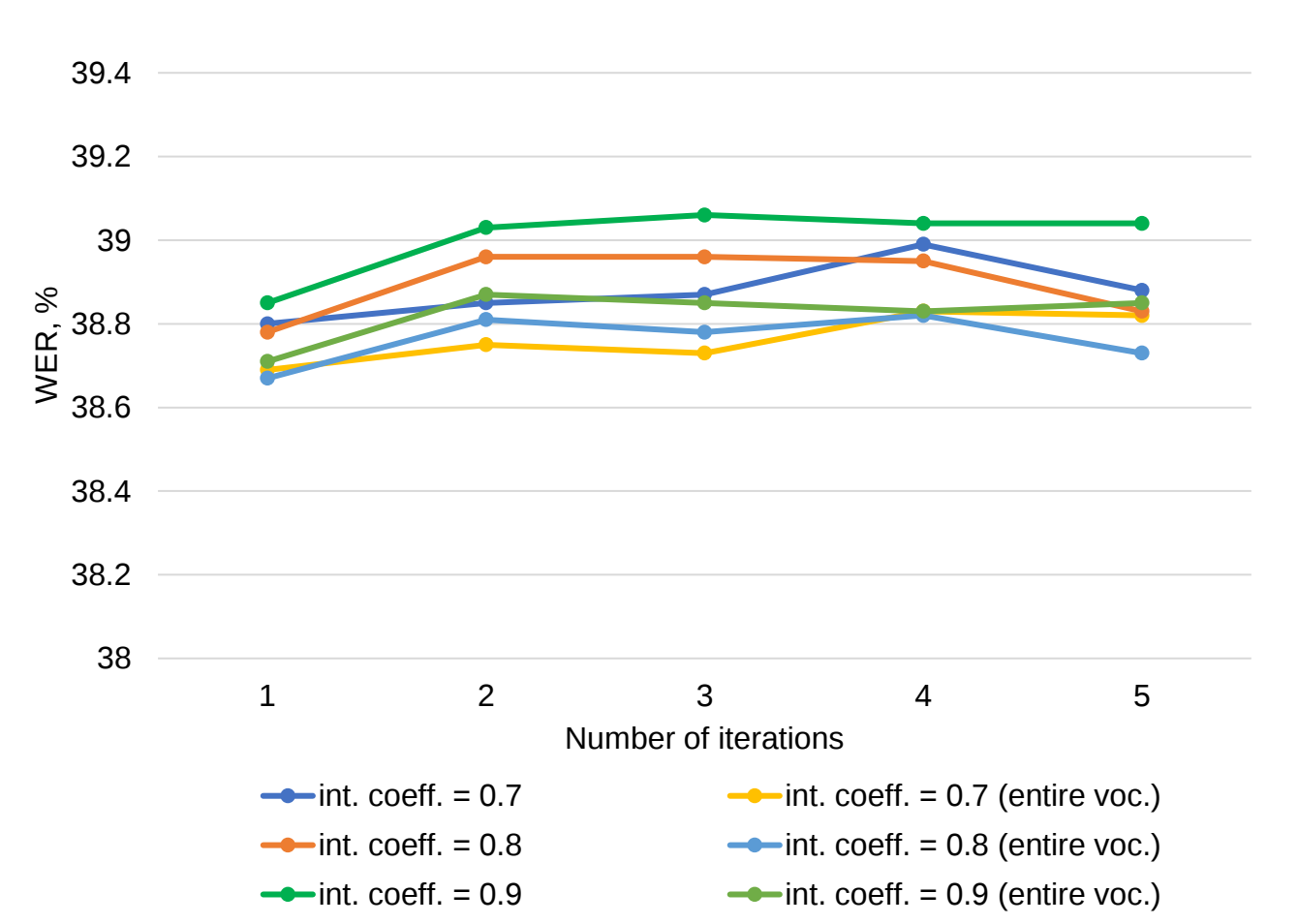
Wav2Vec 2.0 Large Uralic VoxPopuli v2 was fine-tuned for 10K steps, with a batch size of 8 and 4 gradient accumulation steps.

LM	WER, %					
	Dev set			Test set		
Without LM	41.47			46.38		
LM1	41.19			44.00		
		Interpolation coefficients			Interpolation coefficients	
		0.7 0.8 0.9			0.7 0.8 0.9	
LM2	37.64	37.59	37.55	39.01	39.05	39.13
LM3	37.19	37.09	37.14	38.73	38.75	38.78
LM4	37.15	37.15	37.15	38.80	38.78	38.85
LM5	36.97	36.97	36.94	38.69	38.67	38.71
LM4 (4 iterations)	37.15	37.09	36.90	38.99	38.95	39.04
LM5 (4 iterations)	37.00	36.93	36.76	38.83	38.82	38.83

Dependency of WER on Number of Interactions (Dev set)



Dependency of WER on Number of Interactions (Test set)



Acknowledgments. This research was funded by the Russian Science Foundation, grant number 24-21-00276, <https://rscf.ru/en/project/24-21-00276/>