Context-Aware Lexical Stress Prediction and Phonemization for Ukrainian TTS Systems

Presenter:
Anastasiia Senyk

Authors:
Anastasiia Senyk[1],
Mykhailo Lukianchuk[2],
Valentyna Robeiko[2],
Yurii Paniv[1]

Motivation

High-quality Text-to-Speech (TTS) synthesis for low-resource languages like

Ukrainian depends critically on robust text preprocessing.

- **Phonemization**: accurate mapping of graphemes to phonemes.

- **Lexical Stress Prediction**: identifying the correct syllable for stress.

2

Motivation

Ukrainian language

- Limited Resources

- Rich Morphology

- Phonological Complexity

- Non-Deterministic Stress

- **дз**еркало [**дз**е́ркало]
- ві**дз**нака [в'і**дз**на́ка]

- ші**стд**е**с**я́т [ш'і**зд**еис`**а**́т]

"mirror; honors; sixty"

Ми за́вжди́ мріяли побачити старовинний за́мок, але щойно підійшли, як зламався замо́к на його дверях.

"We had always dreamed of seeing an ancient castle, but as soon as we arrived, the lock on its door broke."

3

Related Work - Phonemization

Methods used with other languages.

- Rule-based systems.

- Statistical models (HMMs, joint models).

- Neural models (LSTMs, CNNs, Transformers).

Ukrainian G2P is mostly rule-based.

Common limitations of current solutions.

- Overgeneralized rules.

- Manual input fixes.

- Poor handling of assimilation and phonology.

- Lack of public/open tools.

Phonemization

- IPA phonemes, 52 symbols.

left context + grapheme seq. + right context → phoneme seq.

Phonemization

Step	Rule Type	Examples
1	Grapheme and digraph mapping (я, ю, є, ї, ь, й, щ, дз, дж)	щука → шчука (pike) яблуко → jаблуко (apple) синю → синjу → син´у (blue)
2	Special handling of -ться	робиться → робицца (is being done)
3	Consonant cluster reduction	студентс´киj → студенс´киj (student) невістчин → невісчин (daughter-in-law)
4	Voiced/voiceless consonant assimilation	борот´ба → бород´ба (fight) зсипати → ссипати (pour)
5	Sibilant assimilation	л´отчик → л´оччик (pilot) погодишс´а → погодисс´а (agree) дочц´і → доцц´і (daughter)
6	Palatalized consonant assimilation	с´огодн´і → с´огод´н´і (today)
7	Allophonic variation	вовк → воўк (wolf) гілка → г'ілка (branch)

6

Related Work – Lexical Stress Prediction

Model	Context-Aware	Supports OOV Words	System Design
Ukrainian Word Stress (Syvokon, 2022)	Partially	No	Dictionary
Ukrainian Accentor (Smoliakov, 2022)	No	Yes	Model trained on dictionary data
Accentor Transformer (Mykhailenko, 2023)	Yes	Partially	Transformer trained on synthetic dictionary-labeled corpus

Lexical Stress. Benchmark.

The First Lexical Stress Benchmark for the Ukrainian Language with Sentence-Level Context

Data Collection

Wikipedia Corpus

Style: encyclopedic

Pluperfect GRAC Corpus

Styles: fiction, poetic, folklore, and 7 more

Custom addition

Total sentences: 288

- **Total sentences: 1026**
- Word Forms with at least two stress forms in the dataset: **296**

8

Lexical Stress Model

Task: Train a context-aware Lexical Stress Prediction Model

Synthetic Dataset with Stress-Augmented ASR

Training

- **Dataset**: Common Voice (~60h)
- **Labels**: existing tools
- **Model**: Wav2Vec2

Inference

- **Dataset**: cleaned Voice of America corpus (~300h)

Lexical Stress Prediction Model

- **Model:** ByT5 G2P

9

Results

Results - Phonemizer

Dataset	WER	Notes
Manually constructed dataset	1.23%	Incorrect cases
Automatically generated dataset	3.07%	Incorrect cases
Automatically generated dataset	6.15%	Incorrect + controversial cases
Baseline system	48.75%	Incorrect cases

11

Results – Lexical Stress Model

Model	Sentence-Level Accuracy	Word-Level Accuracy	Ambiguous Word Accuracy	Unambiguous Word Accuracy	Mean-Macro F1 (Ambiguous Word Pairs)
ByT5 G2P	35.3%	87.7%	58.1%	94.8%	37.2%
Uk Accentor	16.6%	73.2%	41.6%	78.7%	28.7%
Uk Accentor Transformer	26.9%	83.4%	43.7%	96.3%	32.4%
Uk Word Stress (First)	41.5%	88.7%	**64.3%**	98.6%	**47.3%**
Uk Word Stress (Skip)	32.5%	86.0%	42.3%	98.6%	35.7%
ByT5 G2P + Uk Word Stress (Skip)	**52.0%**	**92.5%**	61.0%	**98.7%**	46.7%
Uk Accentor + Uk Word Stress (Skip)	48.8%	91.9%	59.1%	**98.7%**	46.3%

Conclusion

- Introduced a two-stage system.

  a. **Context-aware stress predictor**: state-of-the-art performance, outperforming neural baselines and approaching dictionary-level accuracy.

  b. **Rule-based phonemizer**: achieves 1.23% word error rate.

- Released the first public benchmark for Ukrainian sentence-level lexical stress prediction.

13

Future Work

Lexical Stress Model

○ Enhance heteronym coverage in training data.

○ Improve labeling accuracy via manual annotation.

Phonemization

○ Extend to sentence-level processing.

○ Add support for abbreviations and numerical expressions.

○ Adapt to dialects and informal language.

14

Thank you!