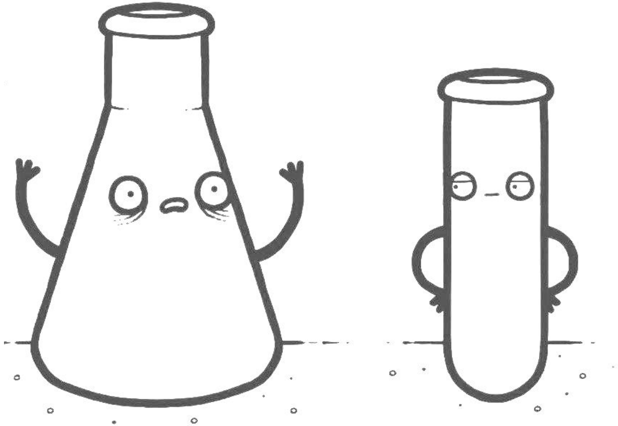
“The boating store has its best sale ever”: Pronunciation-attentive Contextualized Pun Recognition

Yichao Zhou, Jyun-yu Jiang, Jieyu Zhao, Kai-Wei Chang and Wei Wang
Department of Computer Science, University of California, Los Angeles

What is Pun?

I'd tell you a chemistry joke but I know I wouldn't get a reaction.



What is Pun?

I'd tell you a **chemistry** **joke** but I know I **wouldn't get a** **reaction**.

Global Context

Local Context

What is Pun?

I'd tell you a **chemistry joke** but I know I **wouldn't get a reaction**.

Global Context **Local Context**

- Both local and global contexts are consistent with the pun word **"reaction"**.
- **"Reaction"** both means **"chemical change"** and **"response"**.
- The **contrast** between two meanings create a humorous pun.

SCALABLE ANALYTICS INSTITUTE

UCLA Samueli Computer Science

Homographic Puns

I'd tell you a **chemistry joke** but I know I **wouldn't get a reaction**.

Homographic puns rely on multiple interpretations of the same expression.

SCALABLE ANALYTICS INSTITUTE

UCLA Samueli Computer Science

Heterographic Puns

The **boating** **store had its best** **sail** (**sale**) **ever**.

Global Context **Local Context**

- The local and global contexts are consistent with the pun word "sail" and "sale" separately.
- **"Sail"** links to **"boating"**, while **"sale"** relates to **"store had its best"** and **"ever"**.
- The **same or similar pronunciation** connects two words, while the **different meanings** create funniness.

Heterographic Puns

The **boating store had its best sail** (**sale**) **ever**.

Heterographic puns take advantage of phonologically same or similar words.

Puns

SCALABLE ANALYTICS INSTITUTE

UCLA Samueli Computer Science

Task and Previous Research

- In this paper, we tackle the pun detection and location tasks.
- Deploying **word sense disambiguation** methods or using **external knowledge base** **cannot** tackle heterographic puns (Pedersen, 2017; Oele and Evang, 2017).
- **Leveraging static word embedding techniques** that **could not** model pun very well because a word should have very different representations regarding of its context (Hurtado et al., 2017; Indurthi and Oota, 2017; Cai et al., 2018).

Contribution of our work

- In this paper, we propose *Pronunciation-attentive Contextualized Pun Recognition* (PCPR) to jointly model the **contextualized word embeddings** and **phonological word representations** for pun recognition.

- We prove the effectiveness of different embeddings and modules via extensive experiments.

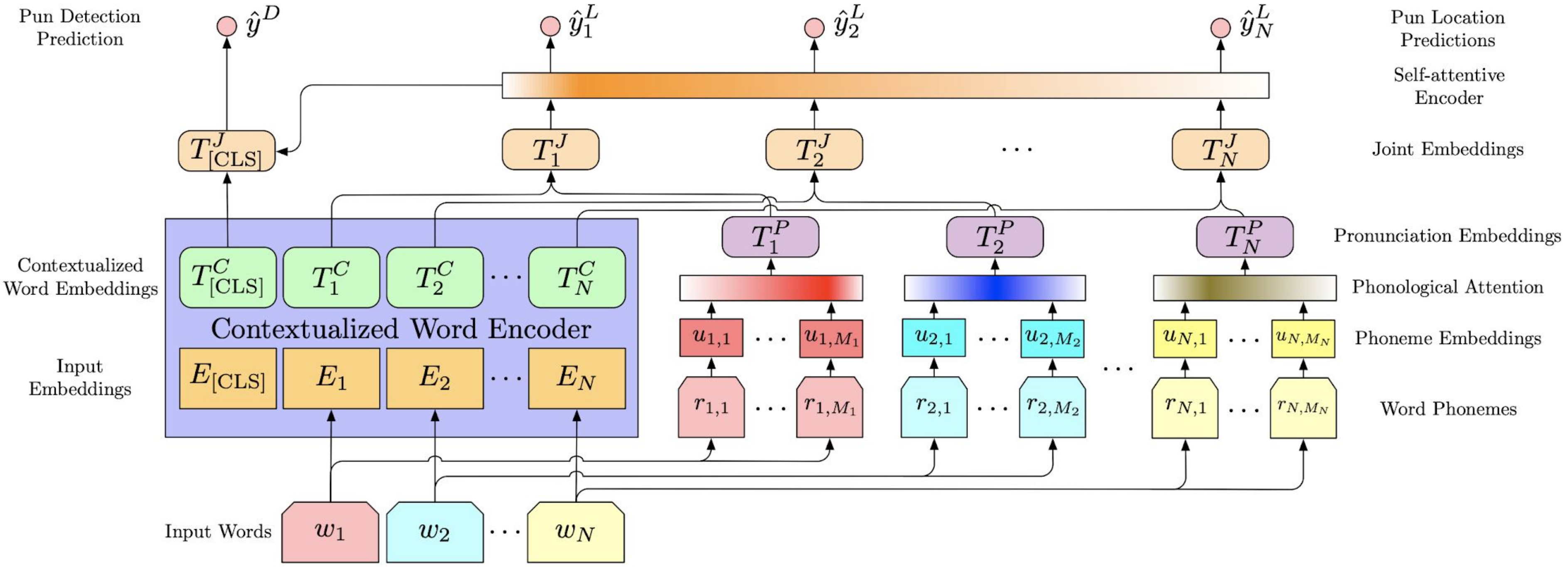
Task Formulation

Suppose the input text consists of a sequence of ***N*** words. For each word with ***M*** phonemes in its pronunciation.

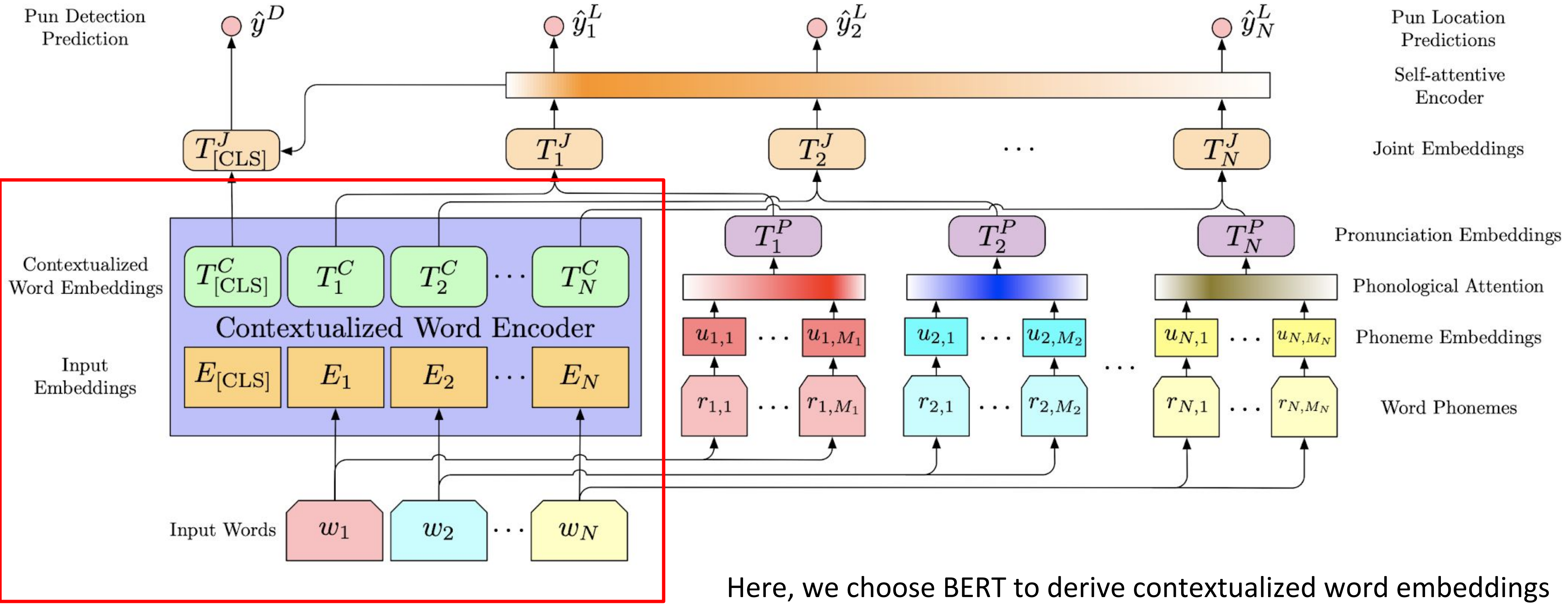
For instance, the phonemes of the word “pun” are {P, AH, N}.

- Pun detection is a **sentence binary classification problem**.
- Pun location can be modeled as a **sequential tagging task**, assigning a binary label to each word.

Framework Architecture



Framework Architecture

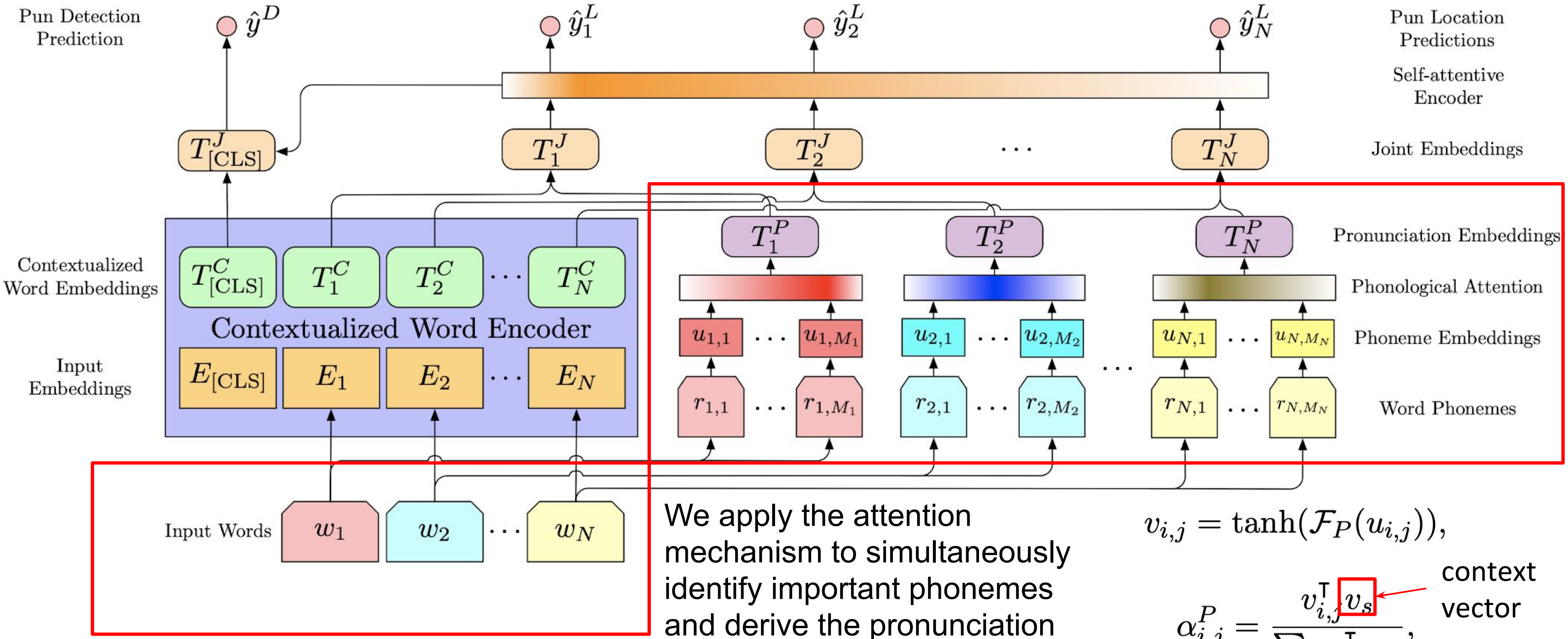


Here, we choose BERT to derive contextualized word embeddings without loss of generality.

SCALABLE ANALYTICS INSTITUTE

UCLA Samueli Computer Science

Framework Architecture



We apply the attention mechanism to simultaneously identify important phonemes and derive the pronunciation embedding for each word.

$$v_{i,j} = \tanh(\mathcal{F}_P(u_{i,j})),$$

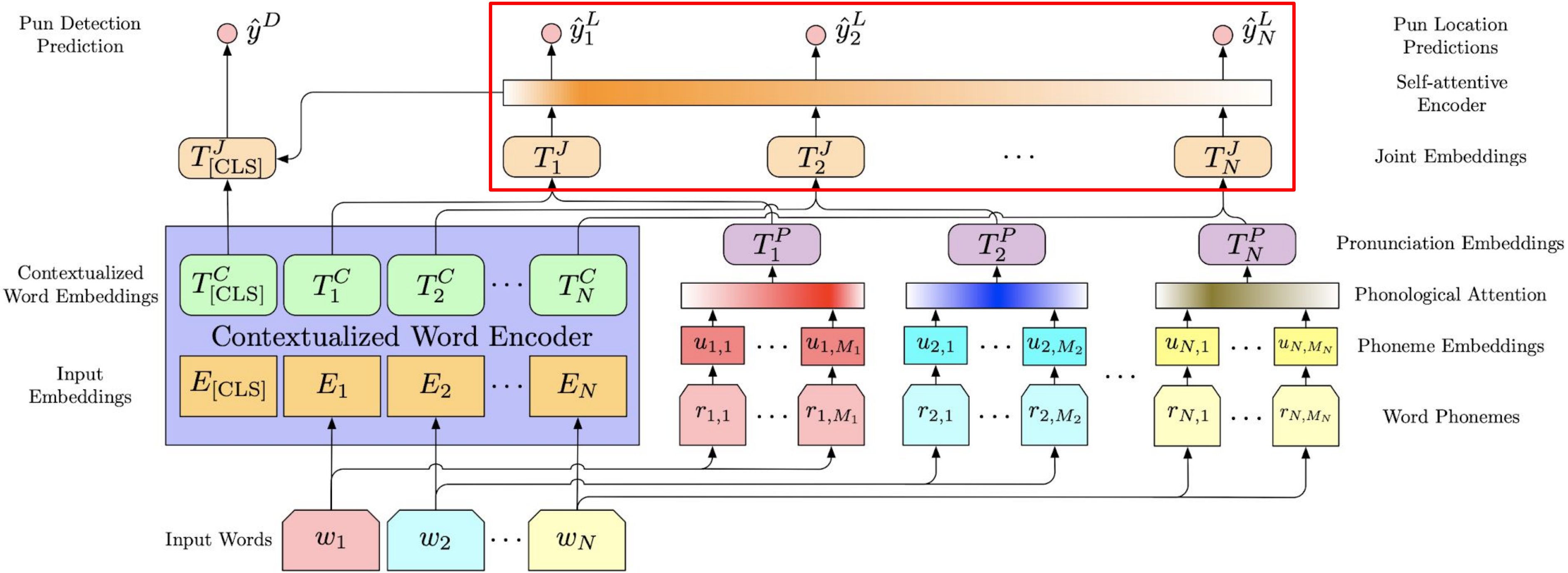
$$\alpha^P_{i,j} = \frac{v_{i,j}^\intercal \boxed{v_s}}{\sum_k v_{i,k}^\intercal v_s},$$

context vector

$$T^P_i = \sum_j \alpha_{i,j} u_{i,j}.$$

F_P (·) is a fully-connected layer and ***$u_{i,j}$*** represents the phoneme embeddings.

Framework Architecture for Pun Location

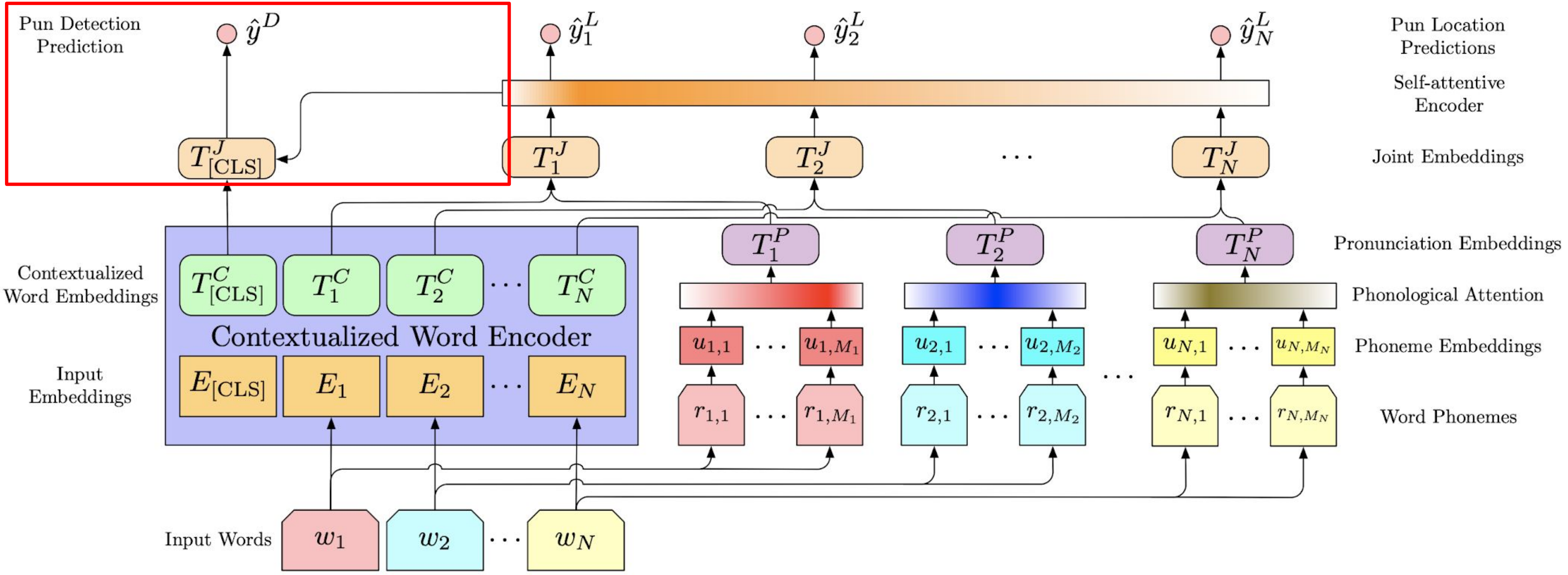


A self-attentive encoder blends contextualized word embeddings and pronunciation embeddings to capture the overall representation for each word.

SCALABLE ANALYTICS INSTITUTE

UCLA Samueli Computer Science

Framework Architecture for Pun Detection



The whole input embedding can be derived by concatenating the overall contextualized embedding and the self-attentive embedding.

Dataset and Evaluation

- The Experiments are conducted on two publicly available benchmark datasets **SemEval 2017 shared task 7** and **Pun of the Day (PTD)**.

Dataset	SemEval		PTD
	Homo	Hetero	
Examples w/ Puns	1,607	1,271	2,423
Examples w/o Puns	643	509	2,403
Total Examples	2,250	1,780	4,826

- We adopted **Precision, Recall and F1-score** to evaluate both pun detection and location task.

Main Experiment on SemEval-2017

SemEval task participants, extracting complicated linguistic features to train rule based and machine learning based classifiers.

Model	Homographic Puns						Heterographic Puns					
	Pun Detection			Pun Location			Pun Detection			Pun Location		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Duluth	78.32	87.24	82.54	44.00	44.00	44.00	73.99	86.62	68.71	-	-	-
JU_CSE_NLP	72.51	90.79	68.84	33.48	33.48	33.48	73.67	94.02	71.74	37.92	37.92	37.92
PunFields	79.93	73.37	67.82	32.79	32.79	32.79	75.80	59.40	57.47	35.01	35.01	35.01
UWAV	68.38	47.23	46.71	34.10	34.10	34.10	65.23	41.78	42.53	42.80	42.80	42.80
Fermi	90.24	89.70	85.33	52.15	52.15	52.15	-	-	-	-	-	-
UWaterloo	-	-	-	65.26	65.21	65.23	-	-	-	79.73	79.54	79.64
Sense	-	-	-	81.50	74.70	78.00	-	-	-	-	-	-
CRF	87.21	64.09	73.89	86.31	55.32	67.43	89.56	70.94	79.17	88.46	62.76	73.42
Joint	91.25	93.28	92.19	83.55	77.10	80.19	86.67	93.08	89.76	81.41	77.50	79.40
CPR	91.42	94.21	92.79	88.80	85.65	87.20	93.35	95.04	94.19	92.31	88.24	90.23
PCPR	**94.18**	**95.70**	**94.94**	**90.43**	**87.50**	**88.94**	**94.84**	**95.59**	**95.22**	**94.23**	**90.41**	**92.28**

Main Experiment on SemEval-2017

Model	Homographic Puns						Heterographic Puns					
	Pun Detection			Pun Location			Pun Detection			Pun Location		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Duluth	78.32	87.24	82.54	44.00	44.00	44.00	73.99	86.62	68.71	-	-	-
JU_CSE_NLP	72.51	90.79	68.84	33.48	33.48	33.48	73.67	94.02	71.74	37.92	37.92	37.92
PunFields	79.93	73.37	67.82	32.79	32.79	32.79	75.80	59.40	57.47	35.01	35.01	35.01
UWAV	68.38	47.23	46.71	34.10	34.10	34.10	65.23	41.78	42.53	42.80	42.80	42.80
Fermi	90.24	89.70	85.33	52.15	52.15	52.15	-	-	-	-	-	-
UWaterloo	-	-	-	65.26	65.21	65.23	-	-	-	79.73	79.54	79.64
Sense	-	-	-	81.50	74.70	78.00	-	-	-	-	-	-
CRF	87.21	64.09	73.89	86.31	55.32	67.43	89.56	70.94	79.17	88.46	62.76	73.42
Joint	91.25	93.28	92.19	83.55	77.10	80.19	86.67	93.08	89.76	81.41	77.50	79.40
CPR	91.42	94.21	92.79	88.80	85.65	87.20	93.35	95.04	94.19	92.31	88.24	90.23
PCPR	**94.18**	**95.70**	**94.94**	**90.43**	**87.50**	**88.94**	**94.84**	**95.59**	**95.22**	**94.23**	**90.41**	**92.28**

Incorporates word sense emb into RNN

Main Experiment on SemEval-2017

Captures linguistic features such as POS tags, n-grams, and word suffix

Model	Homographic Puns						Heterographic Puns					
	Pun Detection			Pun Location			Pun Detection			Pun Location		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Duluth	78.32	87.24	82.54	44.00	44.00	44.00	73.99	86.62	68.71	-	-	-
JU_CSE_NLP	72.51	90.79	68.84	33.48	33.48	33.48	73.67	94.02	71.74	37.92	37.92	37.92
PunFields	79.93	73.37	67.82	32.79	32.79	32.79	75.80	59.40	57.47	35.01	35.01	35.01
UWAV	68.38	47.23	46.71	34.10	34.10	34.10	65.23	41.78	42.53	42.80	42.80	42.80
Fermi	90.24	89.70	85.33	52.15	52.15	52.15	-	-	-	-	-	-
UWaterloo	-	-	-	65.26	65.21	65.23	-	-	-	79.73	79.54	79.64
Sense	-	-	-	81.50	74.70	78.00	-	-	-	-	-	-
$\boxed{\text{CRF}}$	87.21	64.09	73.89	86.31	55.32	67.43	89.56	70.94	79.17	88.46	62.76	73.42
Joint	91.25	93.28	92.19	83.55	77.10	80.19	86.67	93.08	89.76	81.41	77.50	79.40
CPR	91.42	94.21	92.79	88.80	85.65	87.20	93.35	95.04	94.19	92.31	88.24	90.23
PCPR	**94.18**	**95.70**	**94.94**	**90.43**	**87.50**	**88.94**	**94.84**	**95.59**	**95.22**	**94.23**	**90.41**	**92.28**

Main Experiment on SemEval-2017

Model	Homographic Puns						Heterographic Puns					
	Pun Detection			Pun Location			Pun Detection			Pun Location		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Duluth	78.32	87.24	82.54	44.00	44.00	44.00	73.99	86.62	68.71	-	-	-
JU_CSE_NLP	72.51	90.79	68.84	33.48	33.48	33.48	73.67	94.02	71.74	37.92	37.92	37.92
PunFields	79.93	73.37	67.82	32.79	32.79	32.79	75.80	59.40	57.47	35.01	35.01	35.01
UWAV	68.38	47.23	46.71	34.10	34.10	34.10	65.23	41.78	42.53	42.80	42.80	42.80
Fermi	90.24	89.70	85.33	52.15	52.15	52.15	-	-	-	-	-	-
UWaterloo	-	-	-	65.26	65.21	65.23	-	-	-	79.73	79.54	79.64
Sense	-	-	-	81.50	74.70	78.00	-	-	-	-	-	-
CRF	87.21	64.09	73.89	86.31	55.32	67.43	89.56	70.94	79.17	88.46	62.76	73.42
Joint	91.25	93.28	92.19	83.55	77.10	80.19	86.67	93.08	89.76	81.41	77.50	79.40
CPR	91.42	94.21	92.79	88.80	85.65	87.20	93.35	95.04	94.19	92.31	88.24	90.23
PCPR	**94.18**	**95.70**	**94.94**	**90.43**	**87.50**	**88.94**	**94.84**	**95.59**	**95.22**	**94.23**	**90.41**	**92.28**

Jointly models two tasks with RNNs and a CRF tagger

Main Experiment on SemEval-2017

Model	Homographic Puns						Heterographic Puns					
	Pun Detection			Pun Location			Pun Detection			Pun Location		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Duluth	78.32	87.24	82.54	44.00	44.00	44.00	73.99	86.62	68.71	-	-	-
JU_CSE_NLP	72.51	90.79	68.84	33.48	33.48	33.48	73.67	94.02	71.74	37.92	37.92	37.92
PunFields	79.93	73.37	67.82	32.79	32.79	32.79	75.80	59.40	57.47	35.01	35.01	35.01
UWAV	68.38	47.23	46.71	34.10	34.10	34.10	65.23	41.78	42.53	42.80	42.80	42.80
Fermi	90.24	89.70	85.33	52.15	52.15	52.15	-	-	-	-	-	-
UWaterloo	-	-	-	65.26	65.21	65.23	-	-	-	79.73	79.54	79.64
Sense	-	-	-	81.50	74.70	78.00	-	-	-	-	-	-
CRF	87.21	64.09	73.89	86.31	55.32	67.43	89.56	70.94	79.17	88.46	62.76	73.42
Joint	91.25	93.28	92.19	83.55	77.10	80.19	86.67	93.08	89.76	81.41	77.50	79.40
CPR	91.42	94.21	92.79	88.80	85.65	87.20	93.35	95.04	94.19	92.31	88.24	90.23
PCPR	**94.18**	**95.70**	**94.94**	**90.43**	**87.50**	**88.94**	**94.84**	**95.59**	**95.22**	**94.23**	**90.41**	**92.28**

Exploits only the contextualized word encoder without considering phonemes.

Main Experiment on SemEval-2017

Model	Homographic Puns						Heterographic Puns					
	Pun Detection			Pun Location			Pun Detection			Pun Location		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Duluth	78.32	87.24	82.54	44.00	44.00	44.00	73.99	86.62	68.71	-	-	-
JU_CSE_NLP	72.51	90.79	68.84	33.48	33.48	33.48	73.67	94.02	71.74	37.92	37.92	37.92
PunFields	79.93	73.37	67.82	32.79	32.79	32.79	75.80	59.40	57.47	35.01	35.01	35.01
UWAV	68.38	47.23	46.71	34.10	34.10	34.10	65.23	41.78	42.53	42.80	42.80	42.80
Fermi	90.24	89.70	85.33	52.15	52.15	52.15	-	-	-	-	-	-
UWaterloo	-	-	-	65.26	65.21	65.23	-	-	-	79.73	79.54	79.64
Sense	-	-	-	81.50	74.70	78.00	-	-	-	-	-	-
CRF	87.21	64.09	73.89	86.31	55.32	67.43	89.56	70.94	79.17	88.46	62.76	73.42
Joint	91.25	93.28	92.19	83.55	77.10	80.19	86.67	93.08	89.76	81.41	77.50	79.40
CPR	91.42	94.21	92.79	88.80	85.65	87.20	93.35	95.04	94.19	92.31	88.24	90.23
PCPR	**94.18**	**95.70**	**94.94**	**90.43**	**87.50**	**88.94**	**94.84**	**95.59**	**95.22**	**94.23**	**90.41**	**92.28**

PCPR dramatically improves the pun location and detection performance, compared to the SOTA models, Joint and CPR.

Main Experiment on SemEval-2017

Model	Homographic Puns						Heterographic Puns					
	Pun Detection			Pun Location			Pun Detection			Pun Location		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Duluth	78.32	87.24	82.54	44.00	44.00	44.00	73.99	86.62	68.71	-	-	-
JU_CSE_NLP	72.51	90.79	68.84	33.48	33.48	33.48	73.67	94.02	71.74	37.92	37.92	37.92
PunFields	79.93	73.37	67.82	32.79	32.79	32.79	75.80	59.40	57.47	35.01	35.01	35.01
UWAV	68.38	47.23	46.71	34.10	34.10	34.10	65.23	41.78	42.53	42.80	42.80	42.80
Fermi	90.24	89.70	85.33	52.15	52.15	52.15	-	-	-	-	-	-
UWaterloo	-	-	-	65.26	65.21	65.23	-	-	-	79.73	79.54	79.64
Sense	-	-	-	81.50	74.70	78.00	-	-	-	-	-	-
CRF	87.21	64.09	73.89	86.31	55.32	67.43	89.56	70.94	79.17	88.46	62.76	73.42
Joint	91.25	93.28	92.19	83.55	77.10	80.19	86.67	93.08	89.76	81.41	77.50	79.40
CPR	91.42	94.21	92.79	88.80	85.65	87.20	93.35	95.04	94.19	92.31	88.24	90.23
PCPR	**94.18**	**95.70**	**94.94**	**90.43**	**87.50**	**88.94**	**94.84**	**95.59**	**95.22**	**94.23**	**90.41**	**92.28**

By applying the pronunciation-attentive representations, different words with similar pronunciations are linked, leading to a much better pinpoint of pun word for the heterographic dataset.

Main Experiment on SemEval-2017

Model	Homographic Puns						Heterographic Puns					
	Pun Detection			Pun Location			Pun Detection			Pun Location		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Duluth	78.32	87.24	82.54	44.00	44.00	44.00	73.99	86.62	68.71	-	-	-
JU_CSE_NLP	72.51	90.79	68.84	33.48	33.48	33.48	73.67	94.02	71.74	37.92	37.92	37.92
PunFields	79.93	73.37	67.82	32.79	32.79	32.79	75.80	59.40	57.47	35.01	35.01	35.01
UWAV	68.38	47.23	46.71	34.10	34.10	34.10	65.23	41.78	42.53	42.80	42.80	42.80
Fermi	90.24	89.70	85.33	52.15	52.15	52.15	-	-	-	-	-	-
UWaterloo	-	-	-	65.26	65.21	65.23	-	-	-	79.73	79.54	79.64
Sense	-	-	-	81.50	74.70	78.00	-	-	-	-	-	-
CRF	87.21	64.09	73.89	86.31	55.32	67.43	89.56	70.94	79.17	88.46	62.76	73.42
Joint	91.25	93.28	92.19	83.55	77.10	80.19	86.67	93.08	89.76	81.41	77.50	79.40
CPR	91.42	94.21	92.79	88.80	85.65	87.20	93.35	95.04	94.19	92.31	88.24	90.23
PCPR	**94.18**	**95.70**	**94.94**	**90.43**	**87.50**	**88.94**	**94.84**	**95.59**	**95.22**	**94.23**	**90.41**	**92.28**

Pronunciation embeddings also facilitate homographic pun detection, implying the potential of pronunciation for enhancing general language modeling. This is consistent with [1] that improves the quality of word embeddings by introducing pronunciation features.

[1] Wenhao Zhu et al. "Improve word embedding using both writing and pronunciation." PloS one, 2018.

Main Experiment on PTD

Exploits word representations with multiple stylistic features.

Applies a random forest model with Word2Vec and human-centric features.

Trains a CNN to learn essential feature automatically.

Improves the CNN by adjusting the filter size and adding a highway layer.

Model	P	R	F_1
MCL	83.80	65.50	73.50
HAE	83.40	88.80	85.90
PAL	86.40	85.40	85.70
HUR	86.60	94.00	90.10
CPR	98.12	**99.34**	98.73
PCPR	**98.44**	99.13	**98.79**

Main Experiment on PTD

- The contextualized word embeddings can implicitly reveal those contradictions of meanings and further improve pun modeling.

- Phonetical embeddings can be intuitively useful to recognize identically pronounced words for detecting heterographic puns.

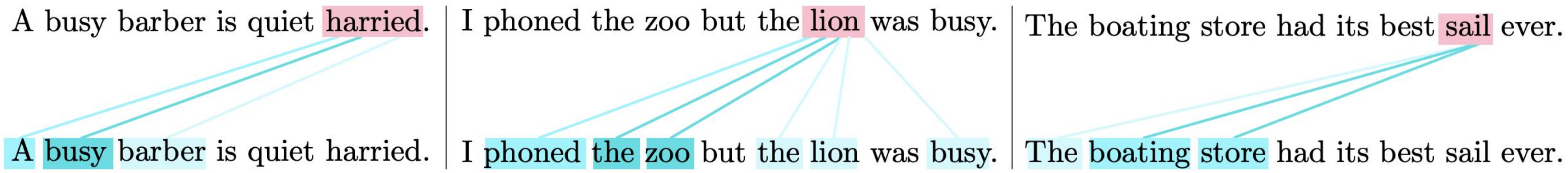
Model	P	R	F_1
MCL	83.80	65.50	73.50
HAE	83.40	88.80	85.90
PAL	86.40	85.40	85.70
HUR	86.60	94.00	90.10
CPR	98.12	**99.34**	98.73
PCPR	**98.44**	99.13	**98.79**

Ablation Study on SemEval-2017

Model	P	R	F_1
PCPR	**90.43**	**87.50**	**88.94**
w/o Pre-trained Phoneme Emb.	89.37	85.65	87.47
w/o Self-attention Encoder	89.17	86.42	87.70
w/o Phonological Attention	89.56	87.35	88.44

All these components are essential for PCPR to recognize puns.

Attention Visualization



Visualization of attention weights of each pun word (marked in pink) in the sentences. A deeper color indicates a higher attention weight.

Conclusion and Future Work

- In this paper, we propose a novel approach, PCPR, for pun recognition by leveraging a **contextualized word encoder** and modeling **phonemes as word pronunciations**.

- Extensive experiments prove the effectiveness of the attention mechanisms, contextualized embeddings and pronunciation embeddings.

- We release our implementations and pre-trained phoneme embeddings at https://github.com/joey1993/pun-recognition to facilitate future research.