“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.
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.

- 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.
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.
Heterographic Puns

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

- 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

DON’T GO BACON MY HEART

WOO HO! IT’S FRY DAY!

I’M FEELING POSITIVE TODAY
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

Pun Detection Prediction

Contextualized Word Embeddings

Input Embeddings

Input Words

Contextualized Word Encoder

Pun Location Predictions

Self-attentive Encoder

Joint Embeddings

Pronunciation Embeddings

Phonological Attention

Phoneme Embeddings

Word Phonemes
Here, we choose BERT to derive contextualized word embeddings without loss of generality.
We apply the attention mechanism to simultaneously identify important phonemes and derive the pronunciation embedding for each word.

\[ v_{i,j} = \tanh(F_P(u_{i,j})) \],

\[ \alpha_{i,j} = \frac{v_{i,j}^T v_s}{\sum_k v_{i,k}^T v_s} \],

\[ T_i^P = \sum_j \alpha_{i,j} u_{i,j} \].

\( F_P (\cdot) \) is a fully-connected layer and \( u_{i,j} \) represents the phoneme embeddings.
A self-attentive encoder blends contextualized word embeddings and pronunciation embeddings to capture the overall representation for each word.
The whole input embedding can be derived by concatenating the overall contextualized embedding and the self-attentive embedding.
The Experiments are conducted on two publicly available benchmark datasets SemEval 2017 shared task 7 and Pun of the Day (PTD).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>SemEval Homo</th>
<th>SemEval Hetero</th>
<th>PTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examples w/ Puns</td>
<td>1,607</td>
<td>1,271</td>
<td>2,423</td>
</tr>
<tr>
<td>Examples w/o Puns</td>
<td>643</td>
<td>509</td>
<td>2,403</td>
</tr>
<tr>
<td>Total Examples</td>
<td>2,250</td>
<td>1,780</td>
<td>4,826</td>
</tr>
</tbody>
</table>

We adopted Precision, Recall and F1-score to evaluate both pun detection and location task.
## Main Experiment on SemEval-2017

<table>
<thead>
<tr>
<th>Model</th>
<th>Homographic Puns</th>
<th>Heterographic Puns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pun Detection</td>
<td>Pun Location</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Duluth</td>
<td>78.32</td>
<td>87.24</td>
</tr>
<tr>
<td>JU_CSE_NLP</td>
<td>72.51</td>
<td>90.79</td>
</tr>
<tr>
<td>PunFields</td>
<td>79.93</td>
<td>73.37</td>
</tr>
<tr>
<td>UWAV</td>
<td>68.38</td>
<td>47.23</td>
</tr>
<tr>
<td>Fermi</td>
<td>90.24</td>
<td>89.70</td>
</tr>
<tr>
<td>UWaterloo</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sense</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CRF</td>
<td>87.21</td>
<td>64.09</td>
</tr>
<tr>
<td>Joint</td>
<td>91.25</td>
<td>93.28</td>
</tr>
<tr>
<td>CPR</td>
<td>91.42</td>
<td>94.21</td>
</tr>
<tr>
<td>PCPR</td>
<td><strong>94.18</strong></td>
<td><strong>95.70</strong></td>
</tr>
</tbody>
</table>

SemEval task participants, extracting complicated linguistic features to train rule based and machine learning based classifiers.
## Main Experiment on SemEval-2017

<table>
<thead>
<tr>
<th>Model</th>
<th>Homographic Puns</th>
<th></th>
<th>Heterographic Puns</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pun Detection</td>
<td>Pun Location</td>
<td>F₁</td>
<td>Pun Detection</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F₁</td>
<td>P</td>
</tr>
<tr>
<td>Duluth</td>
<td>78.32</td>
<td>87.24</td>
<td>82.54</td>
<td>44.00</td>
</tr>
<tr>
<td>JU_CSE_NLP</td>
<td>72.51</td>
<td>90.79</td>
<td>68.84</td>
<td>33.48</td>
</tr>
<tr>
<td>PunFields</td>
<td>79.93</td>
<td>73.37</td>
<td>67.82</td>
<td>32.79</td>
</tr>
<tr>
<td>UWAV</td>
<td>68.38</td>
<td>47.23</td>
<td>46.71</td>
<td>34.10</td>
</tr>
<tr>
<td>Fermi</td>
<td>90.24</td>
<td>89.70</td>
<td>85.33</td>
<td>52.15</td>
</tr>
<tr>
<td>UWaterloo</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>81.50</td>
</tr>
<tr>
<td><strong>Sense</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CRF</td>
<td>87.21</td>
<td>64.09</td>
<td>73.89</td>
<td>86.31</td>
</tr>
<tr>
<td>Joint</td>
<td>91.25</td>
<td>93.28</td>
<td>92.19</td>
<td>83.55</td>
</tr>
<tr>
<td>CPR</td>
<td>91.42</td>
<td>94.21</td>
<td>92.79</td>
<td>88.80</td>
</tr>
<tr>
<td>PCPR</td>
<td><strong>94.18</strong></td>
<td><strong>95.70</strong></td>
<td><strong>94.94</strong></td>
<td><strong>90.43</strong></td>
</tr>
</tbody>
</table>

Incorporates word sense emb into RNN
## Main Experiment on SemEval-2017

<table>
<thead>
<tr>
<th>Model</th>
<th>Homographic Puns</th>
<th>Heterographic Puns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pun Detection</td>
<td>Pun Location</td>
</tr>
<tr>
<td></td>
<td>P    R    F₁</td>
<td>P    R    F₁</td>
</tr>
<tr>
<td>Duluth</td>
<td>78.32 87.24 82.54</td>
<td>44.00 44.00 44.00</td>
</tr>
<tr>
<td>JU_CSE_NLP</td>
<td>72.51 90.79 68.84</td>
<td>33.48 33.48 33.48</td>
</tr>
<tr>
<td>PunFields</td>
<td>79.93 73.37 67.82</td>
<td>32.79 32.79 32.79</td>
</tr>
<tr>
<td>UWAV</td>
<td>68.38 47.23 46.71</td>
<td>34.10 34.10 34.10</td>
</tr>
<tr>
<td>Fermi</td>
<td>90.24 89.70 85.33</td>
<td>52.15 52.15 52.15</td>
</tr>
<tr>
<td>UWWaterloo</td>
<td>-    -    -</td>
<td>65.26 65.21 65.23</td>
</tr>
<tr>
<td>Sense CRF</td>
<td>-    -    -</td>
<td>81.50 74.70 78.00</td>
</tr>
<tr>
<td>Joint</td>
<td>91.25 93.28 92.19</td>
<td>83.55 77.10 80.19</td>
</tr>
<tr>
<td>CPR</td>
<td>91.42 94.21 92.79</td>
<td>88.80 85.65 87.20</td>
</tr>
<tr>
<td>PCPR</td>
<td><strong>94.18</strong> 95.70 94.94</td>
<td><strong>90.43</strong> 87.50 88.94</td>
</tr>
</tbody>
</table>

**Captures linguistic features such as POS tags, n-grams, and word suffix**
<table>
<thead>
<tr>
<th>Model</th>
<th>Homographic Puns</th>
<th></th>
<th>Heterographic Puns</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pun Detection</td>
<td>Pun Location</td>
<td>Pun Detection</td>
<td>Pun Location</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F₁</td>
<td>P</td>
</tr>
<tr>
<td>Duluth</td>
<td>78.32</td>
<td>87.24</td>
<td>82.54</td>
<td>44.00</td>
</tr>
<tr>
<td>JU_CSE_NLP</td>
<td>72.51</td>
<td>90.79</td>
<td>68.84</td>
<td>33.48</td>
</tr>
<tr>
<td>PunFields</td>
<td>79.93</td>
<td>73.37</td>
<td>67.82</td>
<td>32.79</td>
</tr>
<tr>
<td>UWAV</td>
<td>68.38</td>
<td>47.23</td>
<td>46.71</td>
<td>34.10</td>
</tr>
<tr>
<td>Fermi</td>
<td>90.24</td>
<td>89.70</td>
<td>85.33</td>
<td>52.15</td>
</tr>
<tr>
<td>UWWaterloo</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>65.26</td>
</tr>
<tr>
<td>Sense</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>81.50</td>
</tr>
<tr>
<td>CRF</td>
<td>87.21</td>
<td>64.09</td>
<td>73.89</td>
<td>86.31</td>
</tr>
<tr>
<td><strong>Joint</strong></td>
<td>91.25</td>
<td>93.28</td>
<td>92.19</td>
<td>83.55</td>
</tr>
<tr>
<td>CPR</td>
<td>91.42</td>
<td>94.21</td>
<td>92.79</td>
<td>88.80</td>
</tr>
<tr>
<td>PCPR</td>
<td><strong>94.18</strong></td>
<td><strong>95.70</strong></td>
<td><strong>94.94</strong></td>
<td><strong>90.43</strong></td>
</tr>
</tbody>
</table>

Jointly models two tasks with RNNs and a CRF tagger.
## Main Experiment on SemEval-2017

<table>
<thead>
<tr>
<th>Model</th>
<th>Homographic Puns</th>
<th></th>
<th></th>
<th></th>
<th>Heterographic Puns</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pun Detection</td>
<td>Pun Location</td>
<td></td>
<td>Pun Detection</td>
<td>Pun Location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F&lt;sub&gt;1&lt;/sub&gt;</td>
<td>P</td>
<td>R</td>
<td>F&lt;sub&gt;1&lt;/sub&gt;</td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Duluth</td>
<td>78.32</td>
<td>87.24</td>
<td>82.54</td>
<td>44.00</td>
<td>44.00</td>
<td>44.00</td>
<td>73.99</td>
<td>86.62</td>
</tr>
<tr>
<td>JU_CSE_NLP</td>
<td>72.51</td>
<td>90.79</td>
<td>68.84</td>
<td>33.48</td>
<td>33.48</td>
<td>33.48</td>
<td>73.67</td>
<td>94.02</td>
</tr>
<tr>
<td>PunFields</td>
<td>79.93</td>
<td>73.37</td>
<td>67.82</td>
<td>32.79</td>
<td>32.79</td>
<td>32.79</td>
<td>75.80</td>
<td>59.40</td>
</tr>
<tr>
<td>UWAV</td>
<td>68.38</td>
<td>47.23</td>
<td>46.71</td>
<td>34.10</td>
<td>34.10</td>
<td>34.10</td>
<td>65.23</td>
<td>41.78</td>
</tr>
<tr>
<td>Fermi</td>
<td>90.24</td>
<td>89.70</td>
<td>85.33</td>
<td>52.15</td>
<td>52.15</td>
<td>52.15</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>UWaterloo</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>65.26</td>
<td>65.21</td>
<td>65.23</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sense</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>81.50</td>
<td>74.70</td>
<td>78.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CRF</td>
<td>87.21</td>
<td>64.09</td>
<td>73.89</td>
<td>86.31</td>
<td>55.32</td>
<td>67.43</td>
<td>89.56</td>
<td>70.94</td>
</tr>
<tr>
<td>Joint</td>
<td>91.25</td>
<td>93.28</td>
<td>92.19</td>
<td>83.55</td>
<td>77.10</td>
<td>80.19</td>
<td>86.67</td>
<td>93.08</td>
</tr>
<tr>
<td><strong>CPR</strong></td>
<td>91.42</td>
<td>94.21</td>
<td>92.79</td>
<td>88.80</td>
<td>85.65</td>
<td>87.20</td>
<td>93.35</td>
<td>95.04</td>
</tr>
<tr>
<td><strong>PCPR</strong></td>
<td><strong>94.18</strong></td>
<td><strong>95.70</strong></td>
<td><strong>94.94</strong></td>
<td><strong>90.43</strong></td>
<td><strong>87.50</strong></td>
<td><strong>88.94</strong></td>
<td><strong>94.84</strong></td>
<td><strong>95.59</strong></td>
</tr>
</tbody>
</table>

Exploits only the contextualized word encoder without considering phonemes.
## Main Experiment on SemEval-2017

<table>
<thead>
<tr>
<th>Model</th>
<th>Homographic Puns</th>
<th></th>
<th>Heterographic Puns</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pun Detection</td>
<td>Pun Location</td>
<td>Pun Detection</td>
<td>Pun Location</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F₁</td>
<td>P</td>
</tr>
<tr>
<td>Duluth</td>
<td>78.32</td>
<td>87.24</td>
<td>82.54</td>
<td>44.00</td>
</tr>
<tr>
<td>JU_CSE_NLP</td>
<td>72.51</td>
<td>90.79</td>
<td>68.84</td>
<td>33.48</td>
</tr>
<tr>
<td>PunFields</td>
<td>79.93</td>
<td>73.37</td>
<td>67.82</td>
<td>32.79</td>
</tr>
<tr>
<td>UWAV</td>
<td>68.38</td>
<td>47.23</td>
<td>46.71</td>
<td>34.10</td>
</tr>
<tr>
<td>Fermi</td>
<td>90.24</td>
<td>89.70</td>
<td>85.33</td>
<td>52.15</td>
</tr>
<tr>
<td>UWaterloo</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>65.26</td>
</tr>
<tr>
<td>Sense</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>81.50</td>
</tr>
<tr>
<td>CRF</td>
<td>87.21</td>
<td>64.09</td>
<td>73.89</td>
<td>86.31</td>
</tr>
<tr>
<td>Joint</td>
<td>91.25</td>
<td>93.28</td>
<td>92.19</td>
<td>83.55</td>
</tr>
<tr>
<td>CPR</td>
<td>91.42</td>
<td>94.21</td>
<td>92.79</td>
<td>88.80</td>
</tr>
<tr>
<td>PCPR</td>
<td>94.18</td>
<td>95.70</td>
<td>94.94</td>
<td>90.43</td>
</tr>
</tbody>
</table>

PCPR dramatically improves the pun location and detection performance, compared to the SOTA models, Joint and CPR.
<table>
<thead>
<tr>
<th>Model</th>
<th>Homographic Puns</th>
<th>Heterographic Puns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pun Detection</td>
<td>Pun Location</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Duluth</td>
<td>78.32</td>
<td>87.24</td>
</tr>
<tr>
<td>JU_CSE_NLP</td>
<td>72.51</td>
<td>90.79</td>
</tr>
<tr>
<td>PunFields</td>
<td>79.93</td>
<td>73.37</td>
</tr>
<tr>
<td>UWAV</td>
<td>68.38</td>
<td>47.23</td>
</tr>
<tr>
<td>Fermi</td>
<td>90.24</td>
<td>89.70</td>
</tr>
<tr>
<td>UWaterloo</td>
<td>79.73</td>
<td>79.54</td>
</tr>
<tr>
<td>Sense</td>
<td>87.21</td>
<td>64.09</td>
</tr>
<tr>
<td>CRF</td>
<td>91.25</td>
<td>93.28</td>
</tr>
<tr>
<td>Joint</td>
<td>91.42</td>
<td>94.21</td>
</tr>
<tr>
<td>CPR</td>
<td>94.18</td>
<td>95.70</td>
</tr>
<tr>
<td>PCPR</td>
<td>94.23</td>
<td>90.41</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Homographic Puns</th>
<th></th>
<th></th>
<th>Heterographic Puns</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pun Detection</td>
<td>Pun Location</td>
<td></td>
<td>Pun Detection</td>
<td>Pun Location</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F₁</td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Duluth</td>
<td>78.32</td>
<td>87.24</td>
<td>82.54</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>JU_CSE_NLP</td>
<td>72.51</td>
<td>90.79</td>
<td>68.84</td>
<td>33.48</td>
<td>33.48</td>
</tr>
<tr>
<td>PunFields</td>
<td>79.93</td>
<td>73.37</td>
<td>67.82</td>
<td>32.79</td>
<td>32.79</td>
</tr>
<tr>
<td>UWAV</td>
<td>68.38</td>
<td>47.23</td>
<td>46.71</td>
<td>34.10</td>
<td>34.10</td>
</tr>
<tr>
<td>Fermi</td>
<td>90.24</td>
<td>89.70</td>
<td>85.33</td>
<td>52.15</td>
<td>52.15</td>
</tr>
<tr>
<td>UWaterloo</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>65.26</td>
<td>65.21</td>
</tr>
<tr>
<td>Sense</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>81.50</td>
<td>74.70</td>
</tr>
<tr>
<td>CRF</td>
<td>87.21</td>
<td>64.09</td>
<td>73.89</td>
<td>86.31</td>
<td>55.32</td>
</tr>
<tr>
<td>Joint</td>
<td>91.25</td>
<td>93.28</td>
<td>92.19</td>
<td>83.55</td>
<td>77.10</td>
</tr>
<tr>
<td>CPR</td>
<td>91.42</td>
<td>94.21</td>
<td>92.79</td>
<td>88.80</td>
<td>85.65</td>
</tr>
<tr>
<td>PCPR</td>
<td><strong>94.18</strong></td>
<td><strong>95.70</strong></td>
<td><strong>94.94</strong></td>
<td><strong>90.43</strong></td>
<td><strong>87.50</strong></td>
</tr>
</tbody>
</table>

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.

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.

<table>
<thead>
<tr>
<th>Model</th>
<th>P</th>
<th>R</th>
<th>$F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCL</td>
<td>83.80</td>
<td>65.50</td>
<td>73.50</td>
</tr>
<tr>
<td>HAE</td>
<td>83.40</td>
<td>88.80</td>
<td>85.90</td>
</tr>
<tr>
<td>PAL</td>
<td>86.40</td>
<td>85.40</td>
<td>85.70</td>
</tr>
<tr>
<td>HUR</td>
<td>86.60</td>
<td>94.00</td>
<td>90.10</td>
</tr>
<tr>
<td>CPR</td>
<td>98.12</td>
<td>99.34</td>
<td>98.73</td>
</tr>
<tr>
<td>PCPR</td>
<td><strong>98.44</strong></td>
<td>99.13</td>
<td><strong>98.79</strong></td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Model</th>
<th>P</th>
<th>R</th>
<th>$F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCL</td>
<td>83.80</td>
<td>65.50</td>
<td>73.50</td>
</tr>
<tr>
<td>HAE</td>
<td>83.40</td>
<td>88.80</td>
<td>85.90</td>
</tr>
<tr>
<td>PAL</td>
<td>86.40</td>
<td>85.40</td>
<td>85.70</td>
</tr>
<tr>
<td>HUR</td>
<td>86.60</td>
<td>94.00</td>
<td>90.10</td>
</tr>
<tr>
<td>CPR</td>
<td>98.12</td>
<td><strong>99.34</strong></td>
<td>98.73</td>
</tr>
<tr>
<td>PCPR</td>
<td><strong>98.44</strong></td>
<td>99.13</td>
<td><strong>98.79</strong></td>
</tr>
</tbody>
</table>
Ablation Study on SemEval-2017

<table>
<thead>
<tr>
<th>Model</th>
<th>P</th>
<th>R</th>
<th>F₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCPR</td>
<td>90.43</td>
<td>87.50</td>
<td>88.94</td>
</tr>
<tr>
<td>w/o Pre-trained Phoneme Emb.</td>
<td>89.37</td>
<td>85.65</td>
<td>87.47</td>
</tr>
<tr>
<td>w/o Self-attention Encoder</td>
<td>89.17</td>
<td>86.42</td>
<td>87.70</td>
</tr>
<tr>
<td>w/o Phonological Attention</td>
<td>89.56</td>
<td>87.35</td>
<td>88.44</td>
</tr>
</tbody>
</table>

All these components are essential for PCPR to recognize puns.
A busy barber is quiet *harried*. I phoned the zoo but the *lion* was busy. The boating store had its best *sail* ever.

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](https://github.com/joey1993/pun-recognition) to facilitate future research.