

G2PE: A SIMPLE PYTHON MODULE FOR ENGLISH GRAPHEME TO PHONEME CONVERSION

Kyubyong @ Kakao Brain



G2P AND TTS ...

-
-



G2P AND TTS ...

- g2p for TTS
-



G2P AND TTS ...

- g2p for TTS
- Do g2p modules really help for TTS?



ISSUES

-
-
-
-
-
-
-
-



ISSUES

- heteronyms such as lead, minute, refuse
-
-
-
-
-
-
-
-



ISSUES

- heteronyms such as lead, minute, refuse
- numbers
-
-
-
-
-
-
-



ISSUES

- heteronyms such as lead, minute, refuse
- numbers
- symbols
-
-
-
-
-



ISSUES

- heteronyms such as lead, minute, refuse
- numbers
- symbols
- foreign letters
-
-
-
-



ISSUES

- heteronyms such as lead, minute, refuse
- numbers
- symbols
- foreign letters
- phonetic vs. phonemic
-
-
-



ISSUES

- heteronyms such as lead, minute, refuse
- numbers
- symbols
- foreign letters
- phonetic vs. phonemic
- Often heteronyms can be disambiguated with their POS.
-
-



ISSUES

- heteronyms such as lead, minute, refuse
- numbers
- symbols
- foreign letters
- phonetic vs. phonemic
- Often heteronyms can be disambiguated with their POS.
- But, not always: bass (fish) vs. bass (instrument)
-



ISSUES

- heteronyms such as lead, minute, refuse
- numbers
- symbols
- foreign letters
- phonetic vs. phonemic
- Often heteronyms can be disambiguated with their POS.
- But, not always: bass (fish) vs. bass (instrument)
- Machine learning can help but be expensive.



UNIVERSAL ALPHABETS

-
-
-
-



UNIVERSAL ALPHABETS

- IPA: diacritics
-
-
-



UNIVERSAL ALPHABETS

- IPA: diacritics
- APA: ascii
-
-



UNIVERSAL ALPHABETS

- IPA: diacritics
- APA: ascii
- SAMPA
-



UNIVERSAL ALPHABETS

- IPA: diacritics
- APA: ascii
- SAMPA
- X-SAMPA



ENGLISH PRONUNCIATION DICTIONARIES

-

-

-

-



ENGLISH PRONUNCIATION DICTIONARIES

- CMU pronouncing dictionary: North American English
-
-
-



ENGLISH PRONUNCIATION DICTIONARIES

- CMU pronouncing dictionary: North American English
- wiktionary (cf.
[https://github.com/Kyubyong/pron_dictionaries])
-
-



ENGLISH PRONUNCIATION DICTIONARIES

- CMU pronouncing dictionary: North American English
- wiktionary (cf.
[https://github.com/Kyubyong/pron_dictionaries])
- GB or other regional English?
-



ENGLISH PRONUNCIATION DICTIONARIES

- CMU pronouncing dictionary: North American English
- wiktionary (cf. [\[https://github.com/Kyubyong/pron_dictionaries\]](https://github.com/Kyubyong/pron_dictionaries))
- GB or other regional English?
- Can you really predict how one reads?



ALGORITHM

-
-
-
-
-
-



ALGORITHM

- 0. Spell out Arabic numbers and symbols
-
-
-
-
-



ALGORITHM

- 0. Spell out Arabic numbers and symbols
- 1. Pos tagging with nltk
-
-
-
-



ALGORITHM

- 0. Spell out Arabic numbers and symbols
- 1. Pos tagging with nltk
- 2. Disambiguate heteronyms using 1.
-
-
-



ALGORITHM

- 0. Spell out Arabic numbers and symbols
- 1. Pos tagging with nltk
- 2. Disambiguate heteronyms using 1.
- 3. Look up the dictionary
-
-



ALGORITHM

- 0. Spell out Arabic numbers and symbols
- 1. Pos tagging with nltk
- 2. Disambiguate heteronyms using 1.
- 3. Look up the dictionary
- 4. For OOVs, predict with neural networks
-



ALGORITHM

- 0. Spell out Arabic numbers and symbols
- 1. Pos tagging with nltk
- 2. Disambiguate heteronyms using 1.
- 3. Look up the dictionary
- 4. For OOVs, predict with neural networks
- how important?



PREDICTION MODEL

-
-
-
-
-
-
-



PREDICTION MODEL

- Encode with a GRU layer
-
-
-
-
-
-



PREDICTION MODEL

- Encode with a GRU layer
- Decode with another GRU layer
-
-
-
-
-



PREDICTION MODEL

- Encode with a GRU layer
- Decode with another GRU layer
- Start decoding from the last hidden values of the encoder
-
-
-
-



PREDICTION MODEL

- Encode with a GRU layer
- Decode with another GRU layer
- Start decoding from the last hidden values of the encoder
- Model size: 3.2M
-
-
-



PREDICTION MODEL

- Encode with a GRU layer
- Decode with another GRU layer
- Start decoding from the last hidden values of the encoder
- Model size: 3.2M
- num. parameters: 834k
-
-



PREDICTION MODEL

- Encode with a GRU layer
- Decode with another GRU layer
- Start decoding from the last hidden values of the encoder
- Model size: 3.2M
- num. parameters: 834k
- Train in PyTorch
-



PREDICTION MODEL

- Encode with a GRU layer
- Decode with another GRU layer
- Start decoding from the last hidden values of the encoder
- Model size: 3.2M
- num. parameters: 834k
- Train in PyTorch
- Infer in NumPy to avoid the dependence on pytorch



TRIVIA

-

-

-

-

-

-

-



TRIVIA

- You need to extract correct last vectors of rnns (cf. [here](#))

-

-

-

-

-

-



TRIVIA

- You need to extract correct last vectors of rnns (cf. [here](#))
- Decode with another GRU layer
-
-
-
-
-



TRIVIA

- You need to extract correct last vectors of rnns (cf. [here](#))
- Decode with another GRU layer
- Start decoding from the last hidden values of the encoder
-
-
-
-



TRIVIA

- You need to extract correct last vectors of rnns (cf. [here](#))
- Decode with another GRU layer
- Start decoding from the last hidden values of the encoder
- Model size: 3.2M
-
-
-



TRIVIA

- You need to extract correct last vectors of rnns (cf. [here](#))
- Decode with another GRU layer
- Start decoding from the last hidden values of the encoder
- Model size: 3.2M
- num. parameters: 834k
-
-



TRIVIA

- You need to extract correct last vectors of rnns (cf. [here](#))
- Decode with another GRU layer
- Start decoding from the last hidden values of the encoder
- Model size: 3.2M
- num. parameters: 834k
- [Train in PyTorch](#)
-



TRIVIA

- You need to extract correct last vectors of rnns (cf. [here](#))
- Decode with another GRU layer
- Start decoding from the last hidden values of the encoder
- Model size: 3.2M
- num. parameters: 834k
- [Train in PyTorch](#)
- [Infer in NumPy](#) to avoid the dependence on