How do you pronounce your name?
Improving G2P with transliterations

Aditya Bhargava and Grzegorz Kondrak
University of Alberta
ACL-HLT 2011
Introduction

- Name pronunciations can be fickle
  - Speech synthesis systems must handle them
  - Best G2P system can't account for how I decide my name is pronounced
- Existing transliterations encode this info
  - Ample data that can be easily mined from the Web
Objective: apply transliterations

Gershwin

Гершвин

ガーシュウィン

/ˈdʒɜːʃwɪn/?

/gɜːʃwɪn/?

...?
Applying transliterations

- Assume existing G2P base systems
  - Produce n-best output lists
- Assume available transliteration
- Pick candidate output that is “most similar” to transliteration
Data

- G2P: Combilex
  - Provides “name” annotations
- Transliterations: NEWS Shared Task 2010 English-to-Hindi data
- Intersect data
Base systems

- **Festival** (Black et al., 1998)
  - CARTs
  - Popular end-to-end speech synthesis
- **Sequitur** (Bisani and Ney, 2008)
  - Generative joint n-grams
  - G2P only
- **DirecTL+** (Jiampojamarn et al., 2008)
  - Discriminative phrasal decoding
  - G2P only
Similarity

- Similarity measures:
  - ALINE phoneme-to-phoneme aligner score
    - Rule-based G2P converter for Hindi
  - M2M-Aligner alignment system score
    - Extension of learned edit distance algorithm
- Two overall approaches:
  - Use highest similarity score
  - Combine similarity score with system score
Similarity: results

The chart compares the word accuracy of different methods across three datasets: Festival, Sequitur, and DirecTL+. The methods compared are Base, ALINE, M2M, ALINE+Base, and M2M+Base.
Similarity: results
Similarity: results

- **Festival**
- **Sequitur**
- **DirecTL+**

Word accuracy

- Base
- ALINE
- M2M
- ALINE+Base
- M2M+Base
Similarity: results

The diagram shows the word accuracy results for different models across three datasets: Festival, Sequitur, and DirecTL+. The models compared are:

- Base
- ALINE
- M2M
- ALINE+Base
- M2M+Base

The y-axis represents the word accuracy percentage, and the x-axis represents the datasets.
Similarity: post mortem

- Difficult to do!
- Can't follow transliterations exactly
  - Differences in scripts
  - Differences in languages (phonologies)
  - Noisy data
- Need to smooth out this volatility
- Limited to one language
SVM re-ranking

- Many features
  - Similarity scores (M2M-Aligner)
  - Score differences
  - N-grams based on alignments between transcriptions and transliterations
    - Similar to features used in DirecTL+
SVM re-ranking

- Many features
  - Similarity scores (M2M-Aligner)
  - Score differences
  - N-grams based on alignments between transcriptions and transliterations
    - Similar to features used in DirecTL+
SVM re-ranking

- Allows many languages
  - English-to-{Bengali, Chinese, Hindi, Thai, Japanese, Kannada, Korean, Russian, Tamil}
  - Features repeated for each transliteration
SVM re-ranking

Input

n-best outputs

transliterations

Gershwin

/ʤɛrʃwɪn/

/ɡəʃwɪn/

/ɡɜːʃwɪn/

Gershwin

गर्शविन

गार्श्विन

Гершвин

/garʃwɪn/

/gaːʃwɪn/

/ɡɛrʃwɪn/
SVM re-ranking

- Festival
- Sequitur
- DirecTL+

Word accuracy
- Base
- SVM-score
- SVM-ngram
- SVM-all
SVM re-ranking

Word accuracy

- Festival
- Sequitur
- DirecTL+

Base
SVM-score
SVM-ngram
SVM-all
SVM re-ranking

- Festival
- Sequitur
- DirecTL+

Word accuracy

- Base
- SVM-score
- SVM-ngram
- SVM-all
SVM re-ranking

Word accuracy

Festival
Sequitur
DirecTL+

Base
SVM-score
SVM-ngram
SVM-all
Analysis

• SVM re-ranking gives significant improvements
• Festival and Sequitur get higher improvement
  – The better the base system, the harder it is to re-rank
  – $n$-gram features styled after DirecTL+
    • This benefits Festival and Sequitur
• **Similar features in a novel direction** can lead to improved performance
Analysis

- N-gram features most useful
  - Granular features
  - Includes unable-to-align feature
Multiple languages

![Graph showing the relationship between the number of available transliterations and absolute improvement in word accuracy.](image-url)
Future work

- Apply same re-ranking approach to different tasks (e.g. transliteration) and different data (e.g. transcriptions)
  - Very successful results so far
- Leverage noisy web transcriptions
- Incorporate supplemental information directly in system
Conclusion

- First use of transliterations for G2P
- Basic similarity-based methods don't work
- SVM re-ranking improves all tested base systems
- Multiple languages are vital
- Relevant scripts, etc. are online