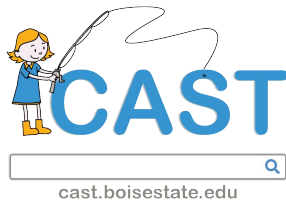


Incorporating Word-Level Phonemic Decoding into Readability Assessment

LREC-COLING 2024
Turin, Italy
May 20-25, 2024
Lingotto Conference Centre



Christine Pinney



Dr. Casey Kennington



Dr. Katherine Landau Wright



Dr. Maria Soledad Pera



Dr. Jerry Alan Fails

This research is part of CAST (Child Adaptive Search Tool), which has the goal to empower emergent searchers and meet the specific search needs of children. This research is supported by National Science Foundation Award #1763649

Agenda

- Motivation
- Related Work & Background
- Method & Data
- Experimental Procedure
- Results & Findings
- Conclusions & Implications

Motivation

- Current approaches in *automatic readability assessment* (ARA) of children's text lack *interpretability* and fail to reflect the *orthographic instruction carried out in classrooms*.
- Large language models and transformer architectures do not offer the interpretability required by the audience of readability tools: teachers and educators.
- Other ARA feature sets focus on syntactic and semantic features of text but do not address phonetic aspects relevant to readability.

Motivation

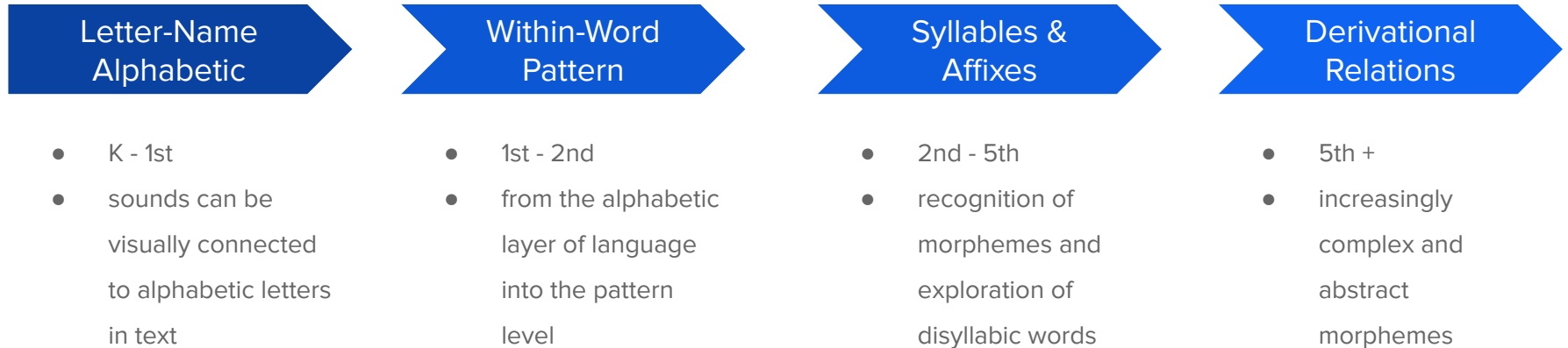
- To learn how to read, children must first learn to decode written letters and phonemes and their connection to speech sounds.
- Our **primary research question**: How do phonemic/phonetic features that characterize the decoding process contribute to current feature-based readability assessment methods, focusing on children who are in the *learning how to read* category of readers?

Related Work & Background

- Large language models and transformer architectures in conjunction with hand-crafted feature sets is a powerful methodological structure for readability assessment.
- For evaluation and comparison purposes, we focus on the approach carried out by Lee et al. (2021).
- Phonemic/phonetic analysis has also been shown to be suitable at predicting readability for lower grade levels (Reyes, 2019).

Related Work & Background

- Words Their Way (WTW), a method of orthographic instruction used in classrooms, provides stages of reading and spelling development (Bear et al., 2020).



Method & Data

- We created functions that map to educational spelling/reading development stages and associated linguistic patterns.
- Not all patterns are mapped, but there are features for each stage.

| Stage | Feature | Examples |
|------------------------|---|--|
| Derivational Relations | Greek Roots Latin Roots Advanced Suffixes Assimilated Prefixes | hydrate , epidermis circum ference, respiration aud ible , hesitancy , optician il logical, ag gregate, ob scure |
| Syllables & Affixes | VV Syllable Juncture VCCV Syllable Juncture VCCV Doublet Syllable Juncture VCCCV Syllable Juncture VVCV Syllable Juncture Compound Words Inflectional Endings for Adjectives Advanced Inflectional Endings | cre-ate , li-ar , pi-ano chap-ter , pub-lic , san-dal bliz-zard , pat-tern , mam-mal pump-kin , dol-phin , bot-tle sea-son , eas-y , float-ed bed room, head light, snow flake help less , bod ily , care ful hopp ed , hopp ing , hopp ing |
| Within Word Pattern | Basic Inflectional Endings Complex Consonants | plants, beach es , pick ed lunch , fudge , knock |
| Letter-Name Alphabetic | CVC Short Vowels | camp , test , stomp , shrunk |

Method & Data

- IPA (International Phonetic Alphabet) translations allow the functions to capture features that manifest as multiple spellings mapped to single speech sounds.
 - The “shun” suffix has many spellings, including *-sion*, *-tion*, and *-cian*.
- Syllable counts
- Parts of speech

Method & Data

- Data:
 - **Reading AtoZ (RAZ):** texts curated and labeled by education professionals for children in kindergarten to 5th grade.
 - **Weebit:** texts gathered from online resources for children in 2nd to 10th grade.
 - **Science:** texts from scientific resources for children in kindergarten to 12th grade.
- We limit the grade range to kindergarten through 5th grade, as this best represents children within the *learning how to read* category of readers.

Experimental Procedure

- Normalized data and collected IPA translations, syllable counts, and parts of speech
- Applied functions to text to identify patterns characteristic of WTW development stages
 - Averaged counts of words containing features, represented as vector
- Classified readability of text by grade level
- Accuracy and F1 scores

Experimental Procedure

- Evaluated/compared with similar feature-based approaches, specifically the semantic and syntactic feature set presented by Lee et al. (2021)
- Multiclass classification with all grade levels and binary classification at grade boundaries
 - Independent feature sets and a combined feature set
 - Random Forest Classifier, Logistic Regression, Support Vector Machine, and Multilayer Perceptron

Results & Findings

- The Lee feature set contained 190 features, while the WTW feature set had only 15 features and still performed within approximately 10% of the Lee feature set in terms of accuracy and F1 scores.

| Dataset | Feature Set | Set Length | Accuracy | F1 Score |
|---------------|-------------|------------|----------|----------|
| RAZ | Lee | 190 | 81% | 73% |
| | Lee & WTW | 205 | 81% | 71% |
| | WTW | 15 | 71% | 65% |
| <u>WeeBit</u> | Lee | 190 | 61% | 60% |
| | Lee & WTW | 205 | 61% | 60% |
| | WTW | 15 | 57% | 57% |
| Science | Lee | 190 | 54% | 52% |
| | Lee & WTW | 205 | 50% | 42% |
| | WTW | 15 | 40% | 37% |

Results & Findings

- Combining the feature sets improved performance at most grade boundaries.
- The WTW feature set performed better than the Lee feature set at the 3rd to 4th and 4th to 5th grade boundaries.
- Feature importance analysis revealed importance of WTW feature set in combined set.

| Grade Boundary | Feature Set | Accuracy | F1 Score |
|-----------------------------------|-------------|----------|----------|
| Kindergarten 1 st | Lee | 98% | 98% |
| | Lee & WTW | 99% | 99% |
| | WTW | 92% | 92% |
| 1 st 2 nd | Lee | 87% | 82% |
| | Lee & WTW | 87% | 82% |
| | WTW | 81% | 72% |
| 2 nd 3 rd | Lee | 85% | 84% |
| | Lee & WTW | 86% | 85% |
| | WTW | 76% | 73% |
| 3 rd 4 th | Lee | 76% | 73% |
| | Lee & WTW | 78% | 76% |
| | WTW | 81% | 79% |
| 4 th 5 th | Lee | 71% | 72% |
| | Lee & WTW | 74% | 72% |
| | WTW | 74% | 72% |

Conclusions & Implications

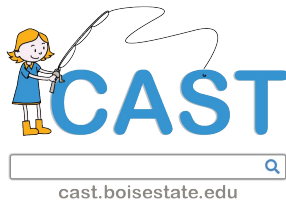
- Features designed to capture the phonetic patterns characteristic of educational development stages for reading and spelling provide valuable information in ARA not represented in syntactic and semantic feature sets.
- When working to find solutions for real-world tasks like ARA through the use of machine learning tools, we should consider the effect of chosen approaches and resulting output on the users with real-world investment in those tools.
 - Educators need transparency and interpretability.

Thanks for Listening!



Full Paper and References
available here.

LREC-COLING 2024



Christine Pinney



Dr. Casey Kennington



Dr. Katherine Landau Wright



Dr. Maria Soledad Pera



Dr. Jerry Alan Fails

This research is part of CAST (Child Adaptive Search Tool), which has the goal to empower emergent searchers and meet the specific search needs of children. This research is supported by National Science Foundation Award #1763649