Educational Psychology

LOL Lab
Language Of Learning Lab

Est. 2019

Meaning Between the Lines: Using Multidimensional Linguistics and Machine-Learning to Predict Reading Comprehension

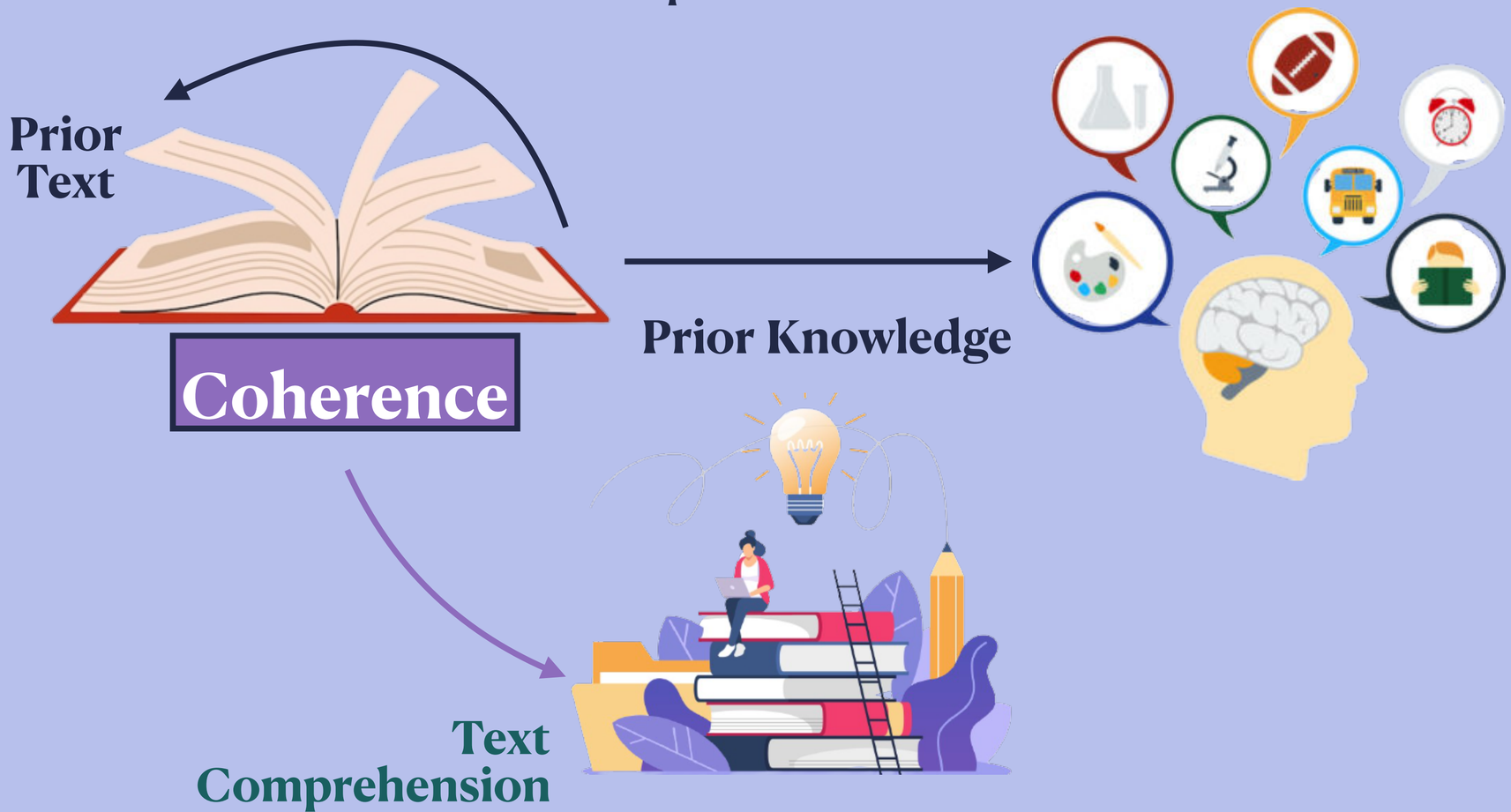
Lauren E. Flynn

SCiP
Society for Computation in Psychology

November 2022

successful text comprehension

when a reader has constructed a *coherent* & meaningful mental representation of a text



coherence vs. cohesion

coherence	cohesion
relational property between the reader and the text	property of text
how connected ideas are of readers' interpretations of text	how connected words and sentences are in discourse
necessary for comphension	can aid in comphension

think-aloud procedures

examines coherence-building
processes while reading

report thoughts

think-aloud cohesion

depends on reader successfully developing
coherent representations while reading

cohesion of think-alouds used to measure readers'
coherence of text

positively associated with:

reading skills

vocabulary knowledge

self-explanations

type of comprehension strategy used during think-alouds

explaining meaning of current text to one's self

promotes inference-generation, deep comprehension, & problem solving

movement towards leveraging NLP to assess constructed responses

current study

analyzes multidimensional linguistic features of self-explanations and machine-learning to predict readers' comprehension

lexical features + cohesive features + source-based features

using a diverse set of science texts

methods

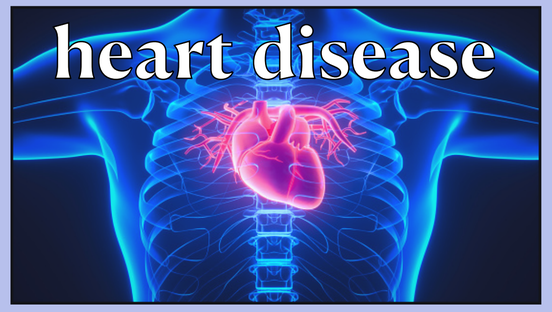
methods

N = 511

n = 118

n = 115

n = 278



shallow Q's
deep Q's

shallow Q's
deep Q's

shallow Q's
deep Q's

comprehension scoring

human raters scored the 8 open-ended comprehension questions for partial credit, ranging from 0-1 for both the textbase and bridging inferences scores

0.00	0.25	0.50	0.75	1.00

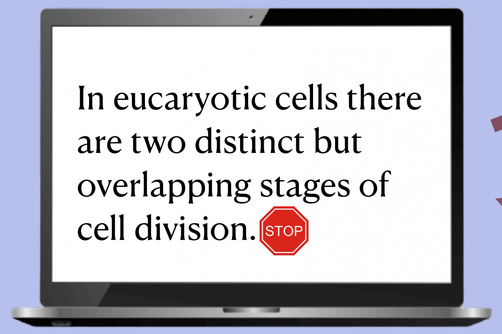
textbase (shallow)

bridging inferences (deep)

$$\frac{0.75 + 0.25}{2} = 0.50$$

(total comprehension score)

procedure

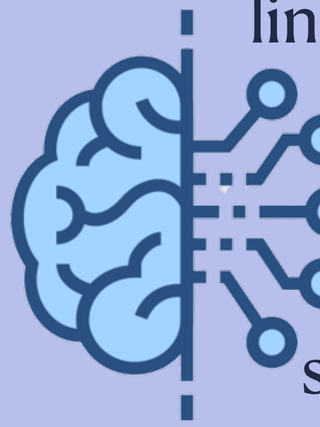


self-explain

lexical features
age of acquisition
academic content word frequency
concreteness of content words

cohesive features
noun type-token-ratio
Word2vec semantic similarity
sentence linking

source-based features
% noun type overlap
% of lemma overlap
total verb lemma overlap



linear regression
BRNN
random forest
svm linear
svm polynomial
svm radial

machine learning

Natural Language Processing (NLP)

lexical features

cohesive features

source-based features

Red blood cells are a necessity for the body. They bring carbon dioxide to the cells of the body. The body then turns the oxygen into carbon dioxide. The red blood cells take the carbon dioxide and have it removed

.....

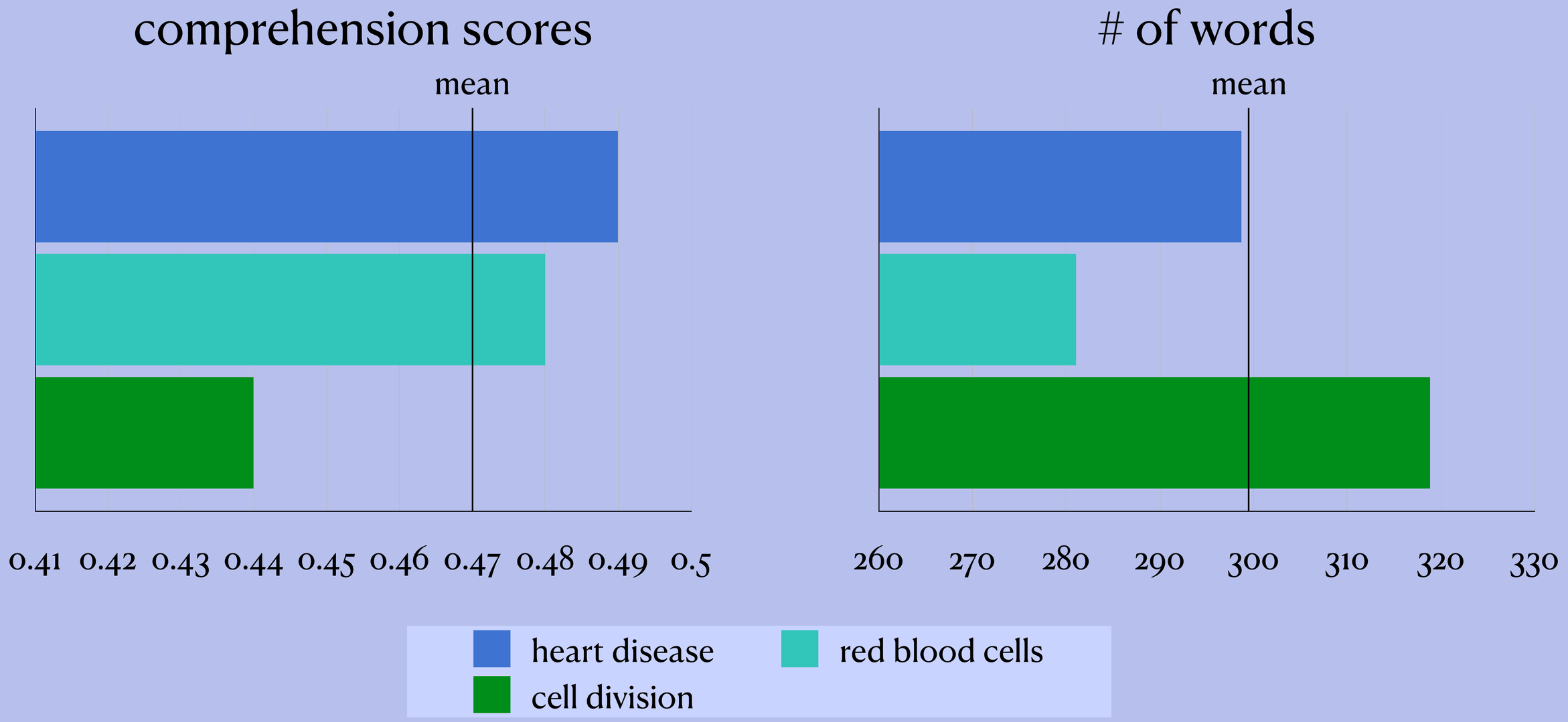
healthy red blood cells carry oxygen around your body. Sometimes not enough oxygen is transported. Not enough oxygen results in anemia

.....

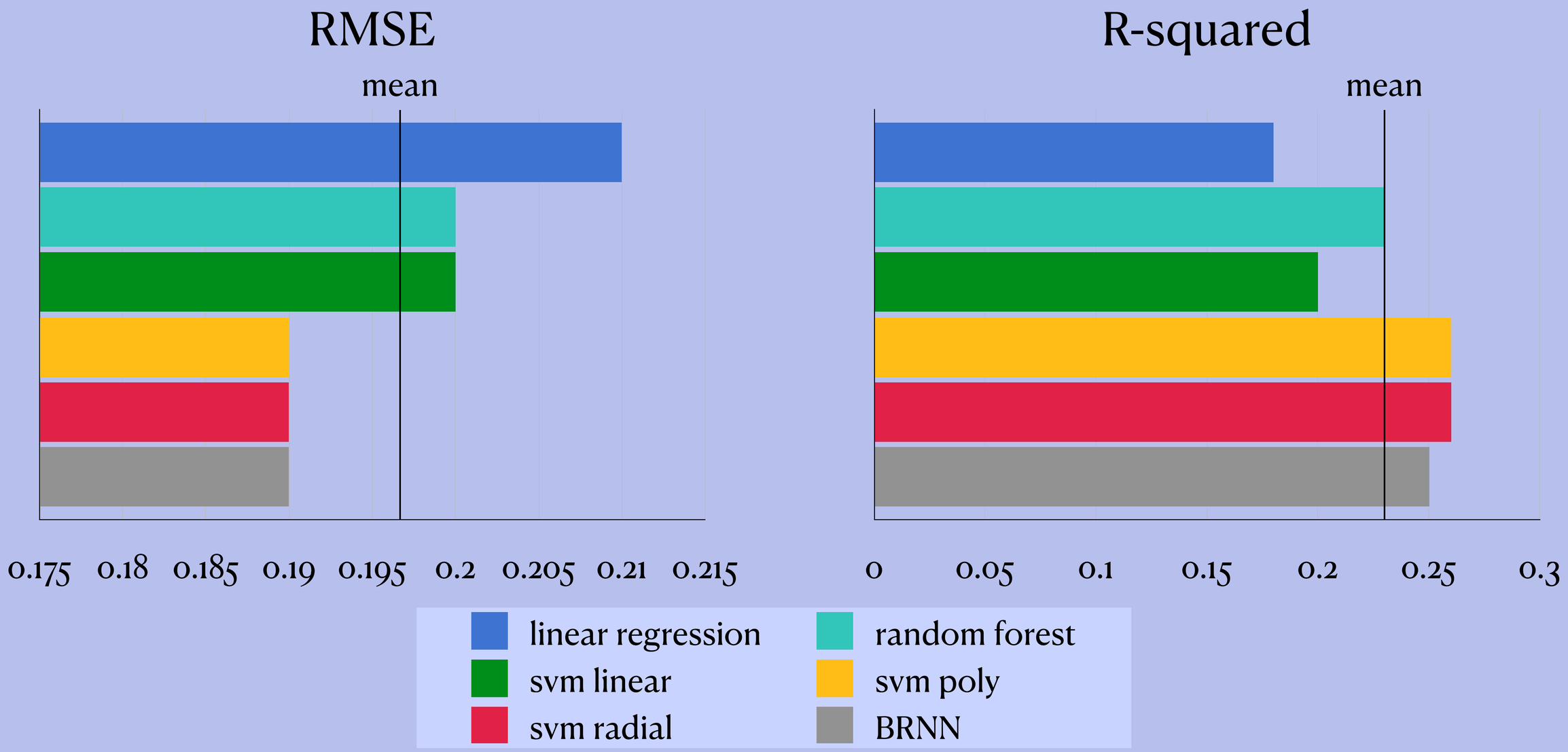
Anemia is a condition where not enough oxygen gets into the body. Anemia can make a person feel tired and weak. One time, my doctor told me I was anemic and it made me feel really tired. I guess, anemia must have something to do with problems with your red blood cells. Is blood a cell?

results

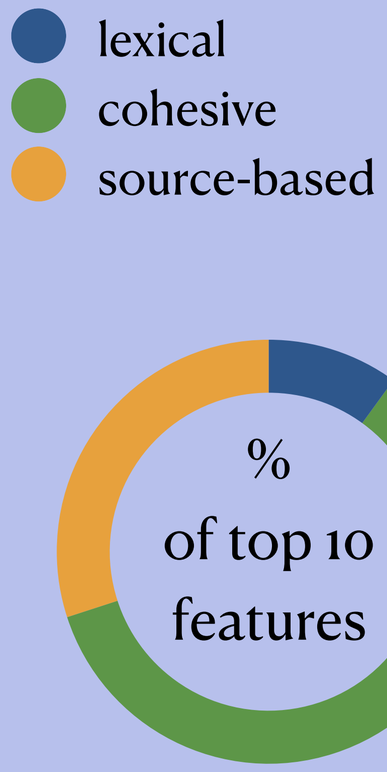
results: descriptives

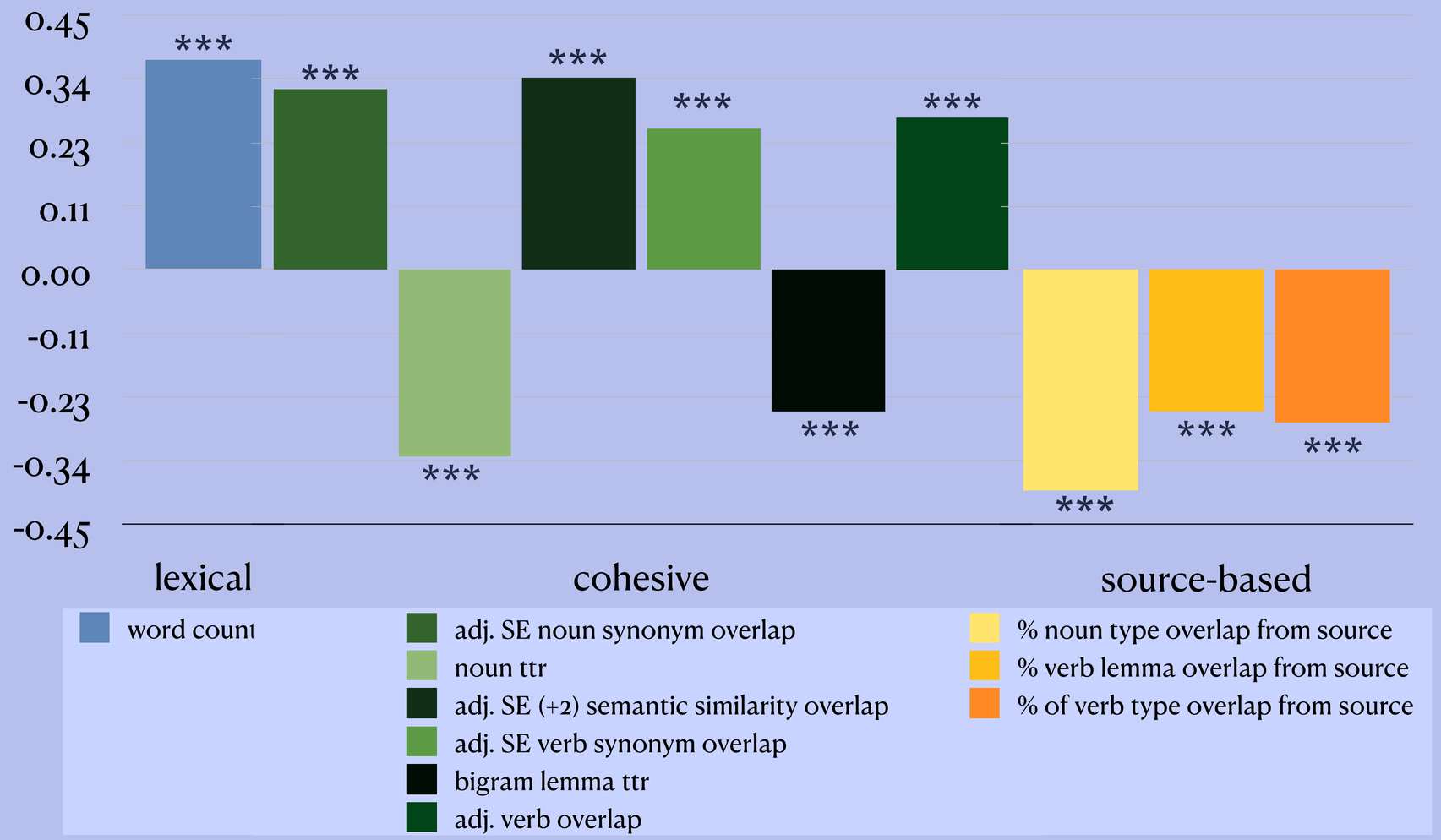


results: algorithm performance

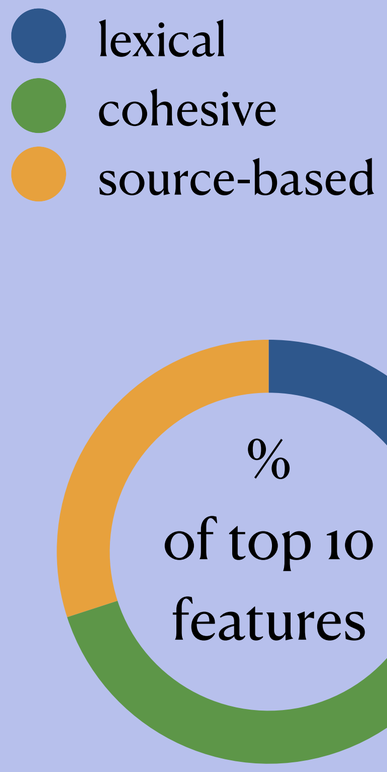


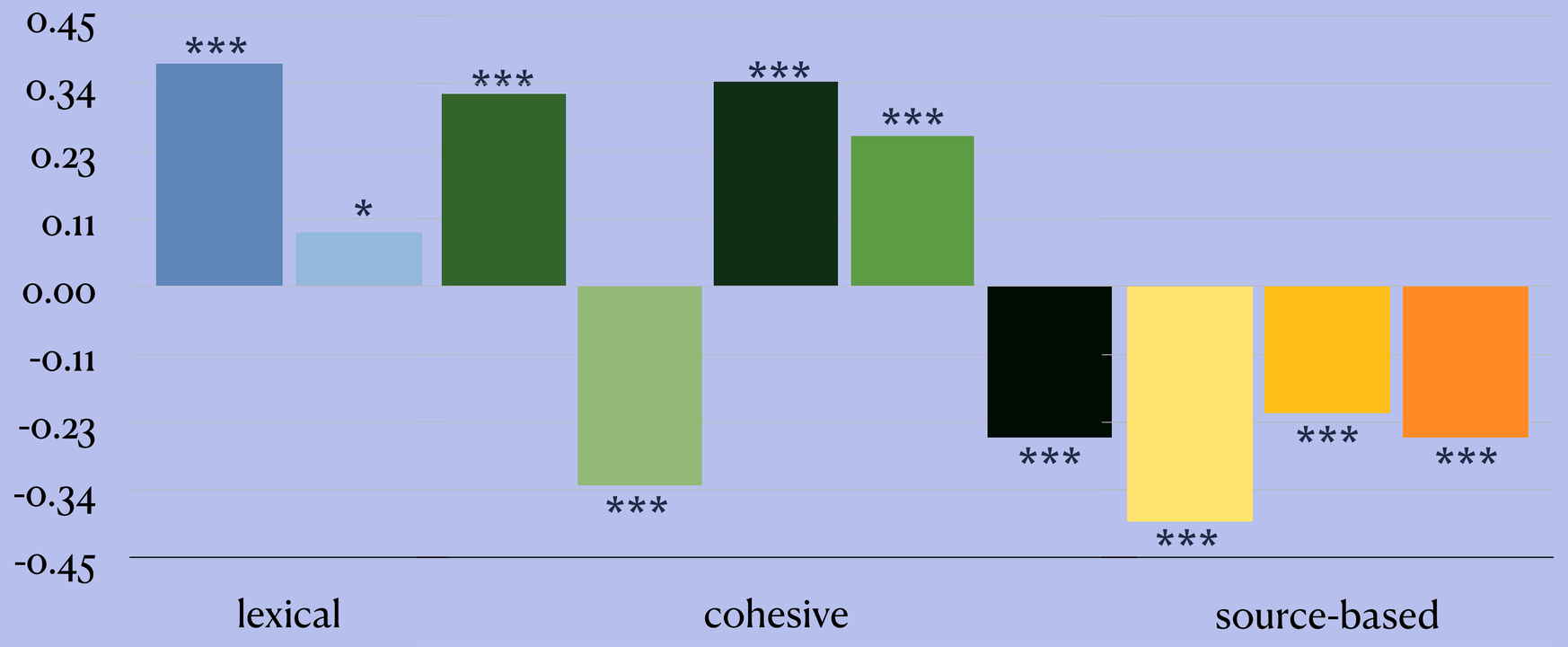
results: top 10 correlated features





results: model exploratory feature analysis





discussion

summary

leveraged multidimensional linguistic features

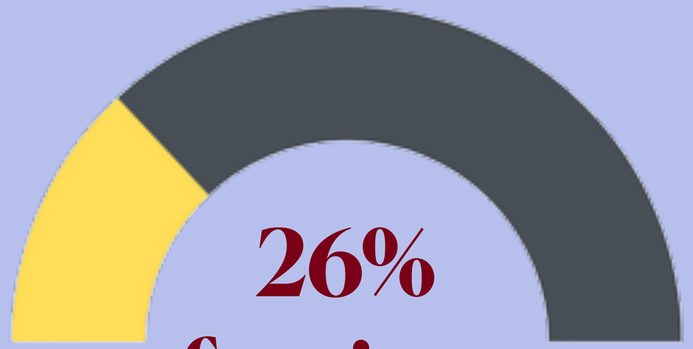
lexical features + cohesive features + source-based features

and machine-learning algorithms

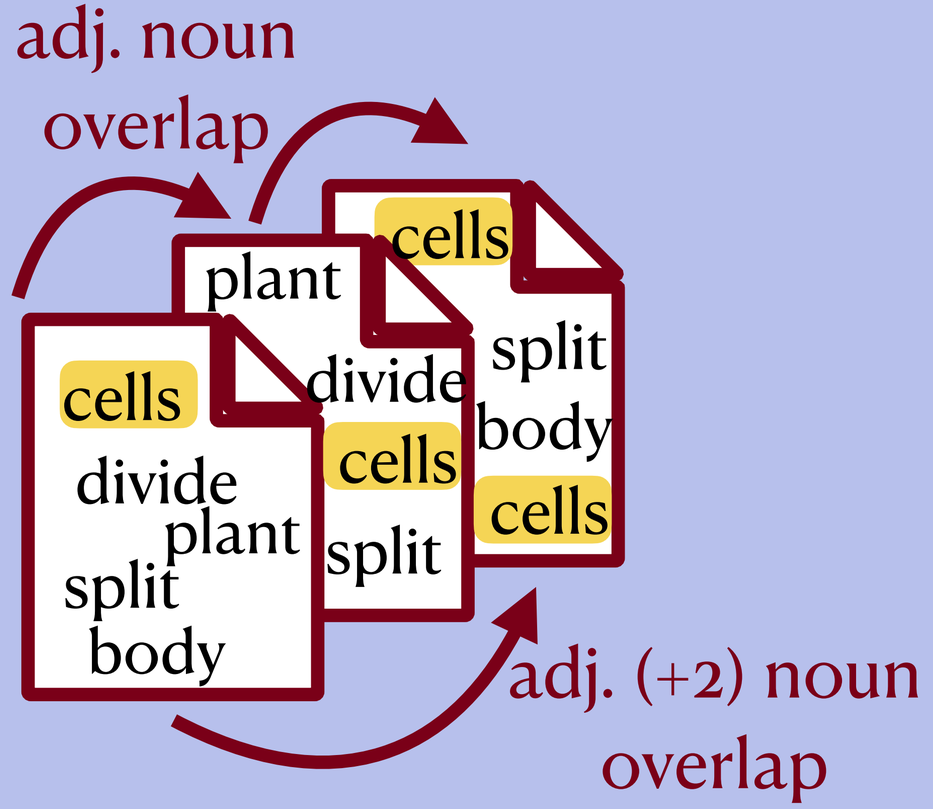
- linear regression
- svm linear
- svm radial
- random forest
- svm poly
- BRNN

to predict overall comprehension of science texts

study significance



svm radial algorithm was the best performing



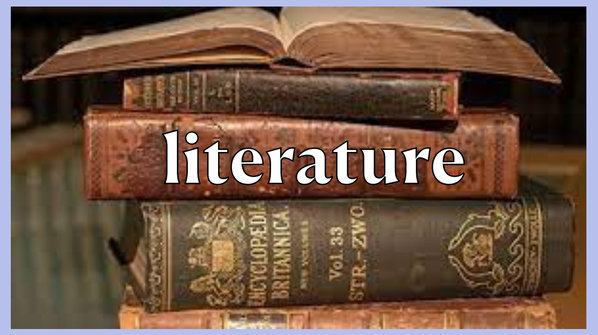
noun overlap (cohesion & source-based) were most important to model performance

moving forward

more fine-grained linguistic analyses to determine linguistic category most indicative of student comprehension

lexical features **vs.** cohesive features **vs.** source-based features

expand scope to other genres of text



moving forward

more fine-grained analyses to assess linguistic feature importance to indications of shallow or deep comprehension

shallow processing (textbase-level comprehension)

deep processing (situation-model comprehension)

e.g., bridging inferences

examine additional linguistic dimensions

sentiment features

+

syntactic features

questions?

email: flynn598@umn.edu