Methods for Identifying Relations and Associations in Text

Stuart G. Towns and Richard Watson Todd

King Mongkut's University of Technology Thonburi

DRAL 2017

Previous Research into Writing Quality

What does English Proficiency mean?

What linguistic features correlate with human judgements of writing quality?
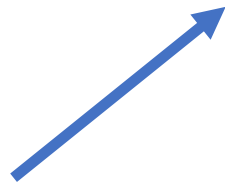
What is Basic English?

What makes good writing good?

What are the linguistic features found in proficient writing?
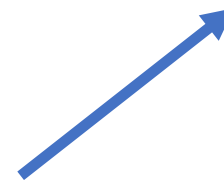
How do we teach students to be good writers?

How are linguistic features expressed in different levels of proficiency?

Linguistic Features for Writing Quality

Syntactic Complexity
- Clause Length
- # of words before the main verb

Lexical Diversity & Sophistication
- Type-Token Ratios
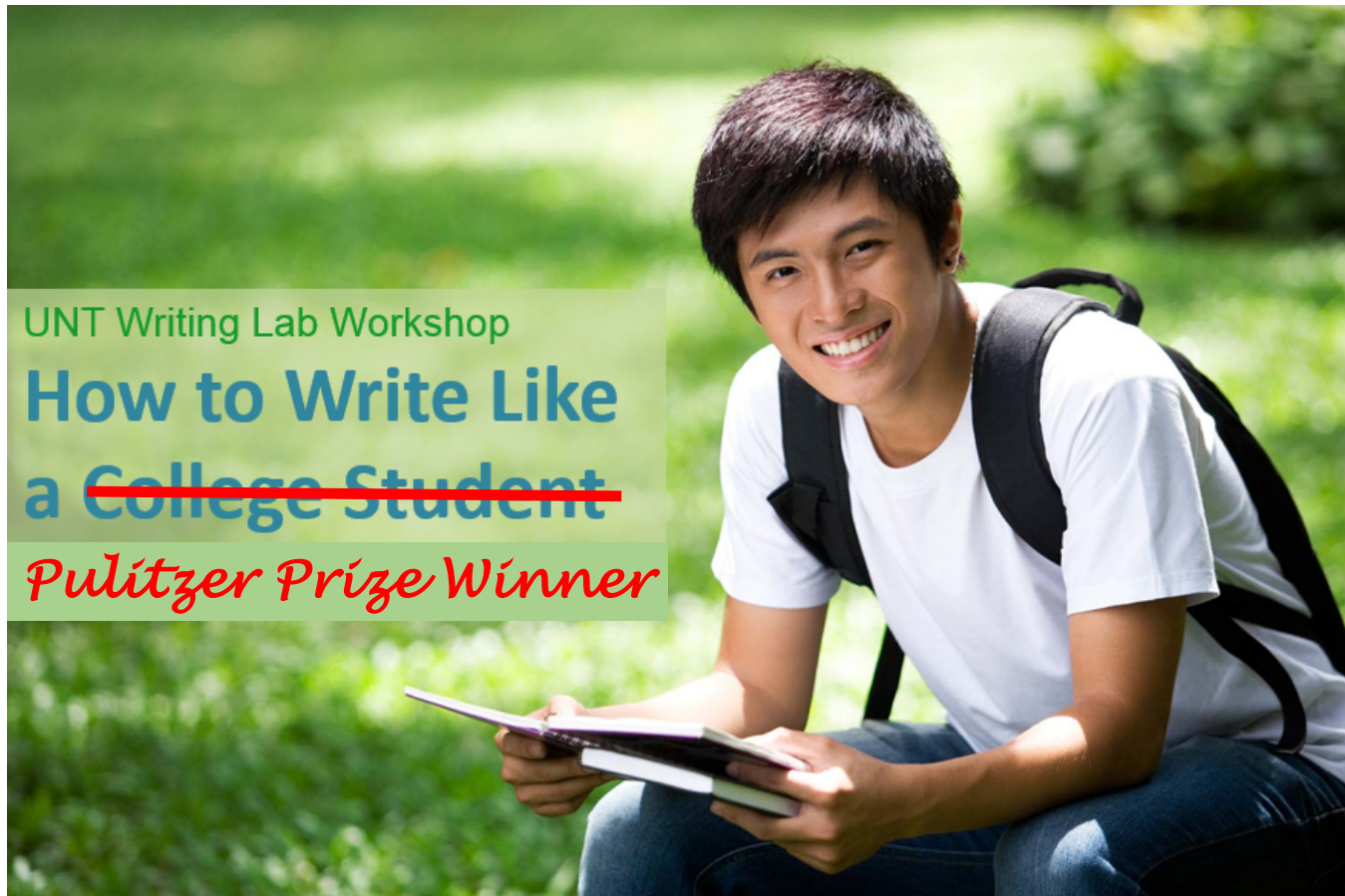- Word Rarity

Cohesion
- Repetition & Reference
- Paraphrases

Coherence

?

Research on Writing Quality

UNT Writing Lab Workshop

How to Write Like a ~~College Student~~

Pulitzer Prize Winner

Adolescent love among eccentrics

"Moonrise Kingdom" opens with no music — just the sound of raindrops falling on the roof of a preternaturally cozy house, which the camera gently leads the audience through as the family members inside go about their rainy-day business.

Bathed in apple reds, egg-yolk yellows and an air of studied eccentricity, the house is immediately recognizable as yet another habitat created by Wes Anderson, a film director whose obsession with material culture, nostalgia and nursery comforts borders on the fetishistic.

Of course, for viewers who happen to share Anderson's taste for boldly framed, bespoke productions — in which everything looks (and most probably is) lovingly handmade and artisanal, "Moonrise Kingdom" will simply offer yet another chance to live, at least for a little while, in the kind of universe only Anderson can cre-

stay out in the rain.

That opening-scene house has a name, by the way: Summer's End, which turns out to aptly capture a vaguely autumnal tale of young love that takes place in early September 1965 — a time of Ford Falcons and mothers who smoked.

The 12-year-old girl who lives in the house, Suzy Bishop (Kara Hayward), wears miniskirts with her scuffed saddle shoes, and daring blue eyeshadow that wouldn't be out of place on Carnaby Street. (Bearing a faint resemblance to Emma Watson, Hayward possesses a similarly compelling face for the screen, especially when she fixes the camera with a dark, adamantine stare.) In the past year, Suzy has struck up a pen-pal friendship with Sam Shakusky (Jared Gilman), a bespectacled, raccoon-tail-capped kid who as "Moonrise Kingdom" begins has just run away from Khaki Scout camp.

As explained by a nameless nar-
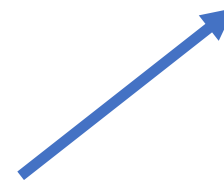

Kara Hayward and Jared Gilman star in Wes Anderson's "Moonrise Kingdom," which may become this summer's art house sleeper.

Linguistic Features for Writing Quality
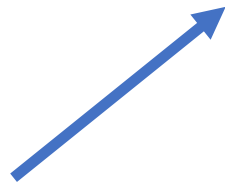
Syntactic Complexity
- Clause Length
- # of words before the main verb

Lexical Diversity & Sophistication
- Type-Token Ratios
- Word Rarity

Cohesion
- Repetition & Reference
- Paraphrases

Coherence

?

Connectedness of Concepts in a Text

Cohesion

explicit connected
concepts in the text

Coherence

implicit connections in
the mind of the reader

Reiterations

- Repetition
- Reference

Relations

- Synonyms
- Hyponyms
- Meronyms

Associations

- Any other
  connection

"Moonrise Kingdom" opens with no music — just the sound of raindrops falling on the roof of a preternaturally cozy house, which the camera gently leads the audience through as the family members inside go about their rainy day business.

Bathed in apple reds, egg yolk yellows and an air of studied eccentricity, the house is immediately recognizable as yet another habitat created by Wes Anderson, a film director whose obsession with material culture, nostalgia and nursery comforts borders on the fetishistic.

Of course, for viewers who happen to share Anderson's taste for boldly framed, bespoke productions — in which everything looks (and most probably is) lovingly handmade and artisanal, "Moonrise Kingdom" will simply offer yet another chance to live, at least for a little while, in the kind of universe only Anderson can create.

(You can almost smell the damp canvas and wood polish in that opening sequence.)

Those who long ago wrote off the writer-director as insufferably mannered and arcane — the usual term of art is "twee" — well, they're welcome to stay out in the rain.

That opening scene house has a name, by the way: Summer's End, which turns out to aptly capture a vaguely autumnal tale of young love that takes place in early September 1965 — a time of Ford Falcons and mothers who smoked.

"Moonrise Kingdom" opens with no music — just the sound of raindrops falling on the roof of a preternaturally cozy house, which the camera gently leads the audience through as the family members inside go about their rainy day business.

Bathed in apple reds, egg yolk yellows and an air of studied eccentricity, the house is immediately recognizable as yet another habitat created by Wes Anderson, a film director whose obsession with material culture, nostalgia and nursery comforts borders on the fetishistic.

Of course, for viewers who happen to share Anderson's taste for boldly framed, bespoke productions — in which everything looks (and most probably is) lovingly handmade and artisanal, "Moonrise Kingdom" will simply offer yet another chance to live, at least for a little while, in the kind of universe only Anderson can create.

(You can almost smell the damp canvas and wood polish in that opening sequence.)

Those who long ago wrote off the writer-director as insufferably mannered and arcane — the usual term of art is "twee" — well, they're welcome to stay out in the rain.

That opening scene house has a name, by the way: Summer's End, which turns out to aptly capture a vaguely autumnal tale of young love that takes place in early September 1965 — a time of Ford Falcons and mothers who smoked.

"Moonrise Kingdom" opens with no music — just the sound of raindrops falling on the roof of a preternaturally cozy house, which the camera gently leads the audience through as the family members inside go about their rainy day business.

Bathed in apple reds, egg yolk yellows and an air of studied eccentricity, the house is immediately recognizable as yet another habitat created by Wes Anderson, a film director whose obsession with material culture, nostalgia and nursery comforts borders on the fetishistic.

Of course, for viewers who happen to share Anderson's taste for boldly framed, bespoke productions — in which everything looks (and most probably is) lovingly handmade and artisanal, "Moonrise Kingdom" will simply offer yet another chance to live, at least for a little while, in the kind of universe only Anderson can create.

(You can almost smell the damp canvas and wood polish in that opening sequence.)

Those who long ago wrote off the writer-director as insufferably mannered and arcane — the usual term of art is "twee" — well, they're welcome to stay out in the rain.

That opening scene house has a name, by the way: Summer's End, which turns out to aptly capture a vaguely autumnal tale of young love that takes place in early September 1965 — a time of Ford Falcons and mothers who smoked.

"Moonrise Kingdom" opens with no music — just the sound of raindrops falling on the roof of a preternaturally cozy house, which the camera gently leads the audience through as the family members inside go about their rainy day business.

Bathed in apple reds, egg yolk yellows and an air of studied eccentricity, the house is immediately recognizable as yet another habitat created by Wes Anderson, a film director whose obsession with material culture, nostalgia and nursery comforts borders on the fetishistic.

Of course, for viewers who happen to share Anderson's taste for boldly framed, bespoke productions — in which everything looks (and most probably is) lovingly handmade and artisanal, "Moonrise Kingdom" will simply offer yet another chance to live, at least for a little while, in the kind of universe only Anderson can create.

(You can almost smell the damp canvas and wood polish in that opening sequence.)

Those who long ago wrote off the writer-director as insufferably mannered and arcane — the usual term of art is "twee" — well, they're welcome to stay out in the rain.

That opening scene house has a name, by the way: Summer's End, which turns out to aptly capture a vaguely autumnal tale of young love that takes place in early September 1965 — a time of Ford Falcons and mothers who smoked.

"Moonrise Kingdom" opens with no music — just the sound of raindrops falling on the roof of a preternaturally cozy house, which the camera gently leads the audience through as the family members inside go about their rainy day business.

Bathed in apple reds, egg yolk yellows and an air of studied eccentricity, the house is immediately recognizable as yet another habitat created by Wes Anderson, a film director whose obsession with material culture, nostalgia and nursery comforts borders on the fetishistic.

Of course, for viewers who happen to share Anderson's taste for boldly framed, bespoke productions — in which everything looks (and most probably is) lovingly handmade and artisanal, "Moonrise Kingdom" will simply offer yet another chance to live, at least for a little while, in the kind of universe only Anderson can create.

(You can almost smell the damp canvas and wood polish in that opening sequence.)

Those who long ago wrote off the writer-director as insufferably mannered and arcane — the usual term of art is "twee" — well, they're welcome to stay out in the rain.

That opening scene house has a name, by the way: Summer's End, which turns out to aptly capture a vaguely autumnal tale of young love that takes place in early September 1965 — a time of Ford Falcons and mothers who smoked.

Associations: PPWs vs Bloggers

Author	Blog or PPW?	% Assoc.
HD	PPW	43%
DA	PPW	41%
MO	PPW	41%
MG	PPW	32%
CT	Blog	30%
CN	Blog	30%
FI	Blog	24%
AF	Blog	22%

Moonrise Kingdom

Author	Blog or PPW?	% Assoc.
MO	PPW	56%
DA	PPW	40%
HD	PPW	38%
MG	PPW	31%
AF	Blog	26%
FB	Blog	23%
FI	Blog	16%
CN	Blog	15%

Ghost Protocol

Methods for Identifying
Relations and Associations in Text

Relations:
- Oxford American Writer's Thesaurus
- WordNet (Princeton)
- UCREL Semantic Analysis System

Associations:
- Word association database at Small World of Words
- Near Neighbors LSA tool hosted at the University of Colorado
- MI scores from the Corpus of Contemporary American English (COCA)

Creating a Word List

These criteria were followed to identify 73 words to be analyzed:

- Must be content words only, excluding adverbs, proper nouns, auxiliary verbs, phrasal verbs, or other idioms (e.g. *of course, by the way*)

- Must be separate entries in dictionaries and thesauruses (e.g. *rainy-day* was broken up into separate words but *egg-yolk* was not)

- Must not be one of the top 250 most common English words

Three Phases of this Study

- Phase 1: Identify all possible word pair connections per source, then determine which word pair connections are relevant to our text.

- Phase 2: Determine which of the six sources returned the most valid results.

- Phase 3: Sum the results from all six sources to see if the word pairs matched the researcher's intuition of connectedness between concepts in the text.

Phase 1

Example: Finding relevant connections

audience – house

In the Thesaurus: The house is the audience at a theatre

In our text: The audience is watching the movie that shows a house

Phase 1

Identify all possible word pair connections per source, then determine which word pair connections are relevant to our text.

Source	Identified	Relevant	%Relevant
Thesaurus	21	6	29%
WordNet	32	12	38%
Semantic Tagging	21	10	48%
Small World	45	23	51%
LSA	34	19	56%
MI Scores	28	15	54%

(Note: 73 words -> 2,628 possible combinations)

Phase 2

Determine which of the six sources returned the most valid results.

- Assumption: An exceptionally well-written text should have high-connectedness and should be well-organized. I.e., the connected pairs should be closer together than a random order.

- Point Bi-Serial Correlation between distance between the words in a word pair and whether or not it was a match.

Phase 2

Point Biserial Correlation between distances and relevance

Source	PBS Correlation
Thesaurus	-0.17**
WordNet	-0.04*
Semantic Tagging	-0.10**
Small World of Words	-0.24**
Latent Semantic Analysis	-0.02
COCA MI Scores	-0.12**

* $p < .05$ ** $p < .00001$

Phase 3

Sum the results from all six sources to see if the word pairs matched the researcher's intuition of connectedness between concepts.

Found by...	# Word Pairs	% Relevant
6 sources	0	n/a
5 Sources	0	n/a
4 Sources	3	100%
3 Sources	8	88%
2 Sources	27	52%
1 Source	91	29%
0 Sources	2,499	0%

audience – viewers
reds – yellows
damp - rain

Phase 1 Finding Summary

Phase 1: The association database (Small World) and corpus-based sources (LSA and MI) returned better results than the other sources, as it identified:

- More word pairs

- Higher percentage of relevant word pairs

- Unique reiterations, relations, and associations

Phase 2 Finding Summary

- The Small World was the most valid source (with respect to distance between words), with WordNet and LSA being the least valid.

  However, WordNet and LSA identified the most number of movie terms.

- Limitation: This study only considered word pairs, not word chains or word networks.

Phase 3 Finding Summary

The more often a word pair was found, the more likely it was a relevant word pair for our text.

These methods reduced the number of potential word pairs from 2,628 to a more manageable 129.

These six methods therefore show promise for future full automation of the identification of relations and associations in a text.

Thank you!