How well do Student Nurses Write Case Studies?
A Cohesion-Centered Textual Complexity Analysis

EC-TEL 2017, Tallinn, Estonia

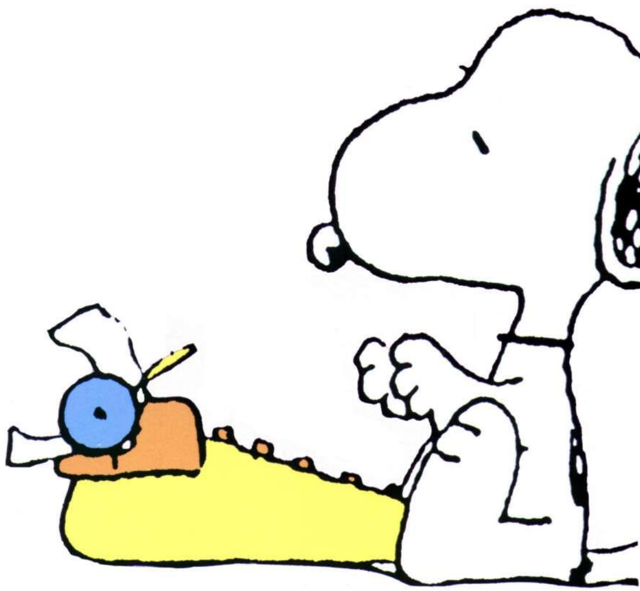
Mihai Dascalu[1,2,3], Philippe Dessus[3], Laurent Thuez[4] & Stefan Trausan-Matu[1,2]

[1] University Politehnica of Bucharest, Romania

[2] Academy of Romanian Scientists, Romania

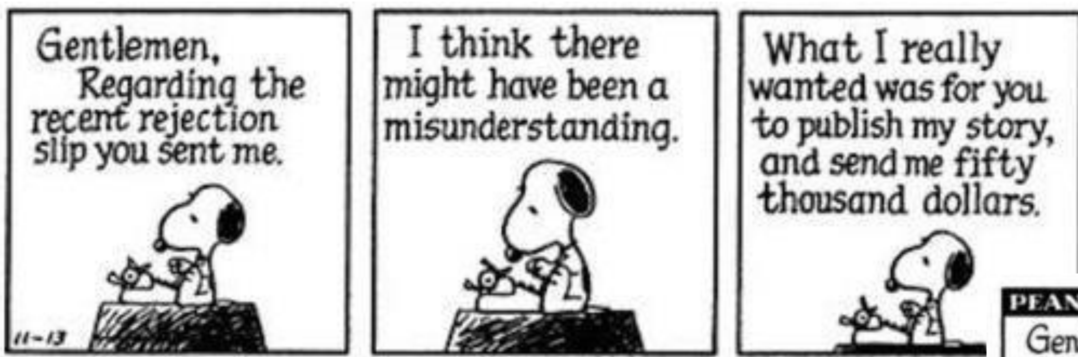
[3] Laboratoire des Sciences de l'Éducation, France

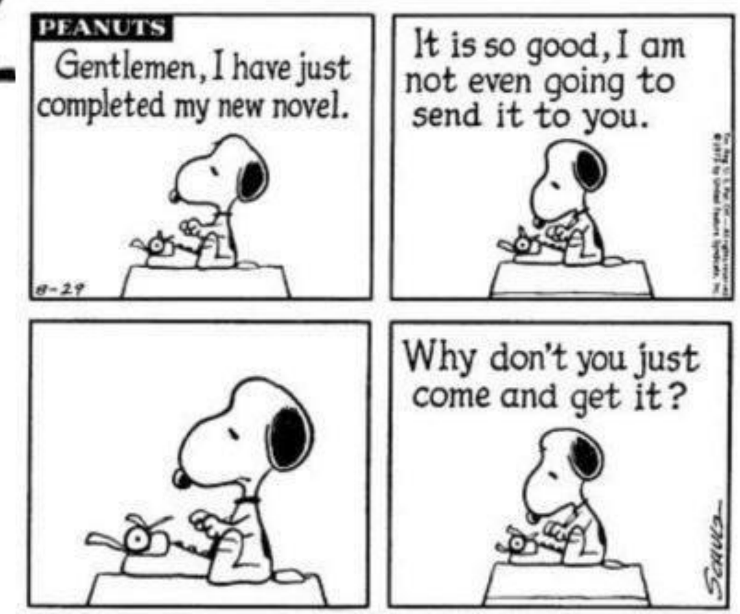
[4] IFSI, Centre Hospitalier Annecy-Genevois, France



Why study writing?

Writing is important for success in school and in the workplace

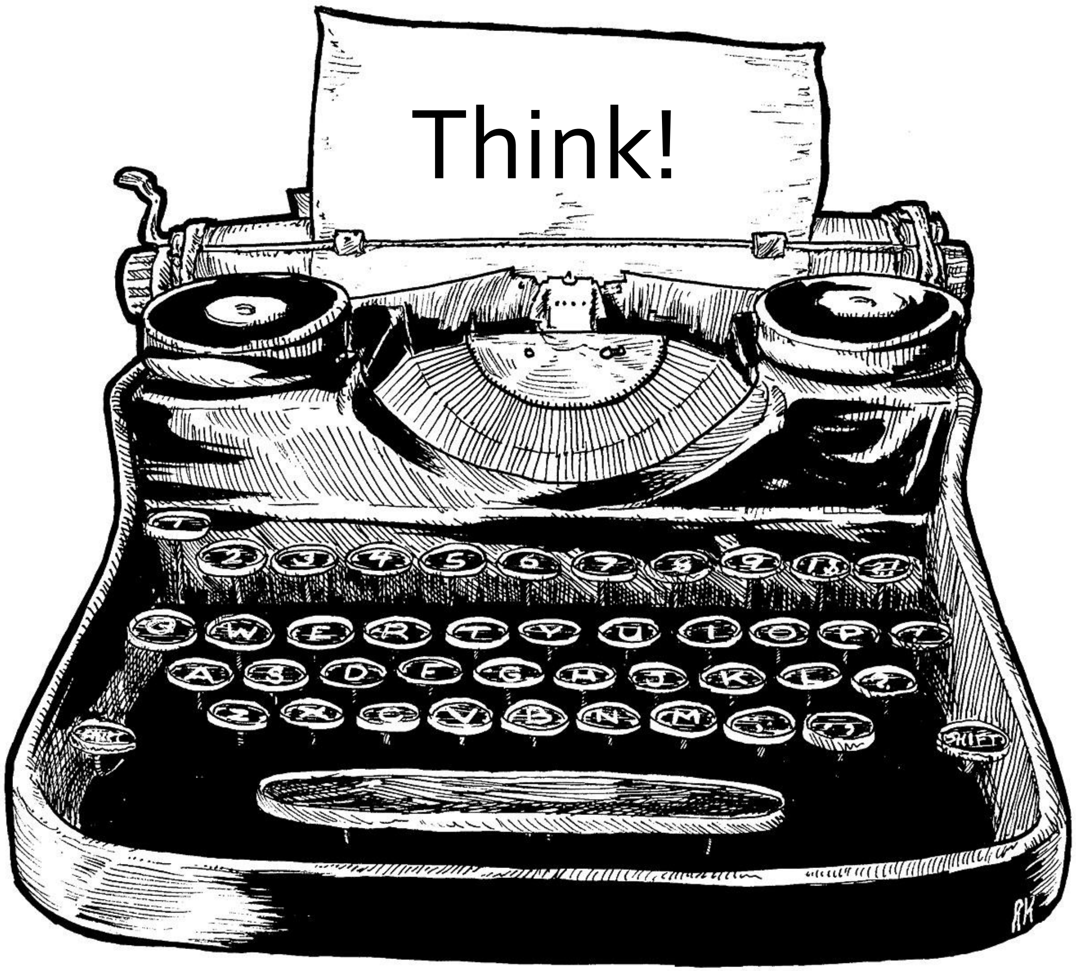




Over 90% of professionals cite writing as essential in the workplace

Light, 2001

Writing is a complex cognitive and social process



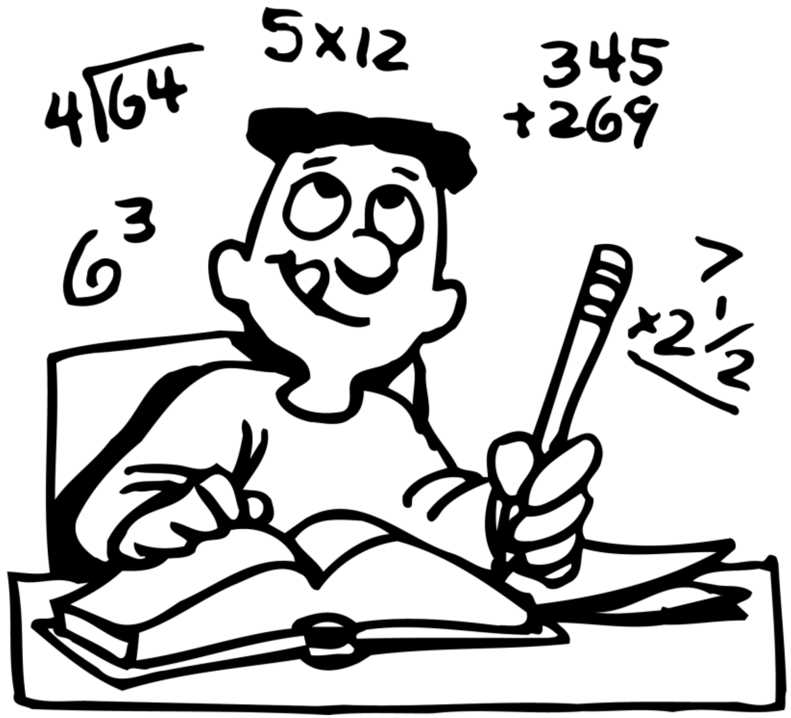
Flower & Hayes, 1980; Graham & Perin, 2007; Hayes, 1996; Torrance & Galbraith, 2006

Students underachieve on national tests of writing

Only 25% students in the U.S. exit high school proficient writers

National Assessment of Educational Progress, 2007; 2011

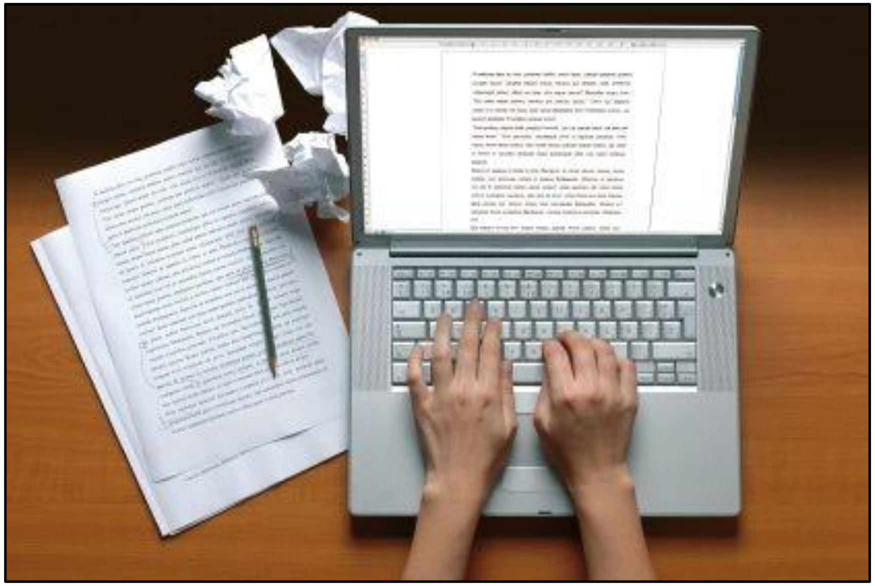
Deliberate Practice



Kellogg & Raulerson, 2007

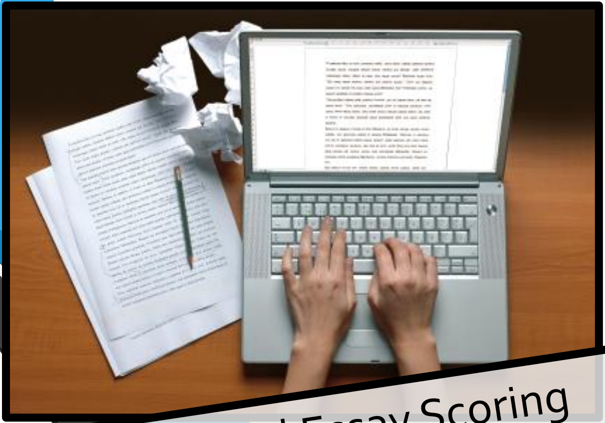
Computer-based Writing Instruction

- Provides students with *deliberate* practice on their writing
- Rely on automated essay scoring engines to assign:
  - Automated holistic scores
  - Relevant feedback to submitted essays



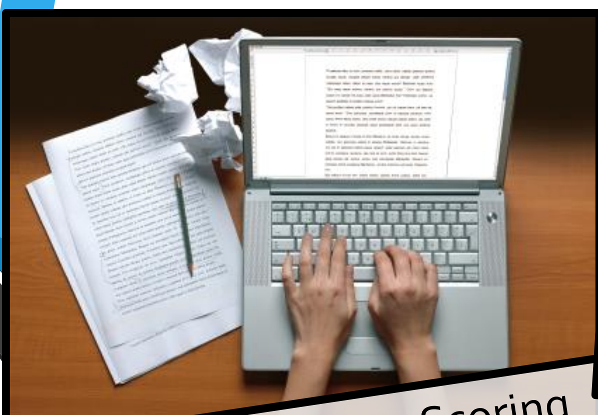
Allen, Jacovina, & McNamara, 2016

Computer-based Writing Instruction

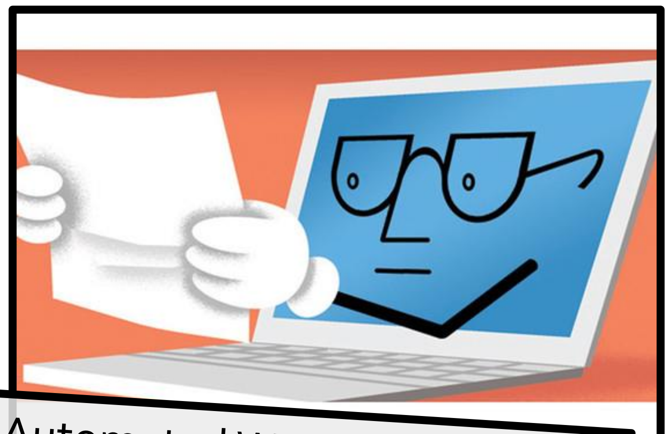


Automated Essay Scoring (AES)

Computer-based Writing Instruction

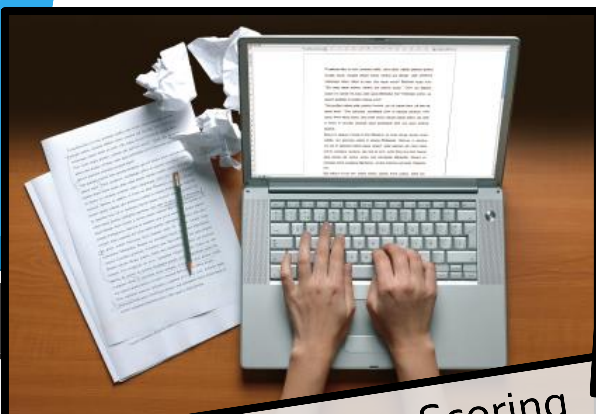


Automated Essay Scoring (AES)

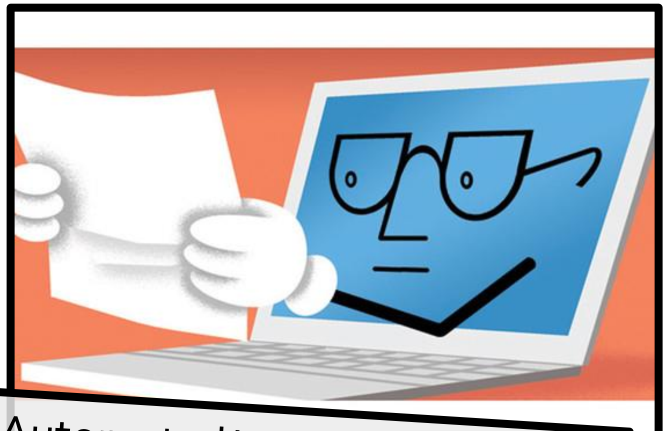


Automated Writing Evaluation (AWE)

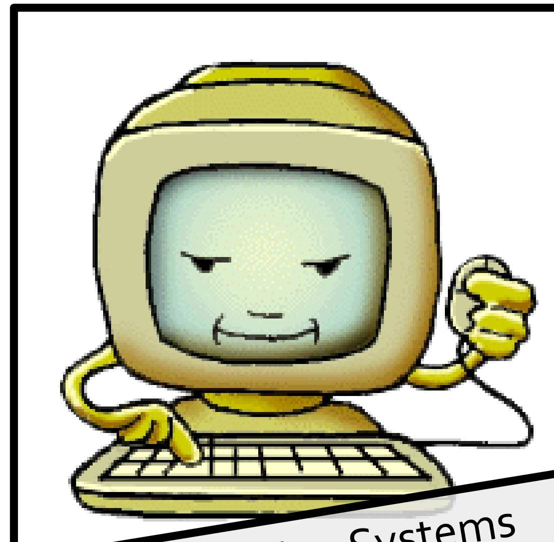
Computer-based Writing Instruction



Automated Essay Scoring
(AES)



Automated Writing Evaluation
(AWE)



Intelligent Tutoring Systems
(ITS)

Context

- Measure and quantify the complexity / difficulty of texts
- Present texts of steadily increasing complexity levels to learners
- Dimensions (Common Core State Standards):
  - Quantitative = measurable factors
  - Qualitative = meaning, structure, language conventionality & clarity
  - Reader/task orientation = prior knowledge, motivation & interests

ReaderBench

- Multi-lingual framework integrating advanced NLP techniques
- Provides a *unified vision* of predicting and assessing comprehension for supporting individual and collaborative learning
- Provide a scalable, easily extensible, multi-language platform:
    - English, French, Romanian, Dutch, Spanish, Italian*, Latin*
- Available online: readerbench.com

Multi-dimensional Analysis

- Surface metrics
- Syntax
- Cohesion
- Semantics
- Discourse structure

Surface Analysis

- Readability formulas (*Flesch Reading Ease, Gunning's Fog Index, Flesch Grade Level, Dale-Chall*)
- Entropy (character, word)
- Fluency (e.g., no. commas, no. words)
- Diction (e.g., word length, *no. syllables/word*, % hard words)
- Structure (e.g., no. paragraphs, size in characters/words/content words, no. sentences, length)

* EN only

Word Complexity

- *Mean syllable count per word*
- Mean polysemy count per word
- Mean distance to the hypernym tree root per word (WordNet)
- Mean difference between the inflected form, the lemma and the corresponding stem, per word
- *Age of Exposure*

* EN only

Syntax & Morphology

- POS occurences:
  - Nouns*
  - Prepositions
  - Adjectives
  - Verbs
  - Pronouns*
  - Adverbs
- Max depth
- Size of parsing tree
- Semantic dependencies

Cohesion

- An important element of writing quality is text cohesion:
  - Cohesive links or ties between sentences from the same text or fragment of text
- Cohesion is an important element of text processing:
  - Create a unified and connected text
- Consists of local and global cohesion:
  - Local → Features that link short text segments of text
  - Global → Features that link larger segments of texts

Cohesion versus Textual Complexity

- Mike likes pancakes. The sky is blue. Your favorite cup is on the table.
- Mike likes pancakes. He also likes cake. Cupcakes are John’s favorites.
- Mike likes probabilities. He also likes factoring quadratic expressions. Log-linear interpolations are Mike's favorites.

- Cohesion in itself is not enough to distinguish text difficulty
- Lack of cohesion may (artificially) increase textual complexity

Semantics & Discourse

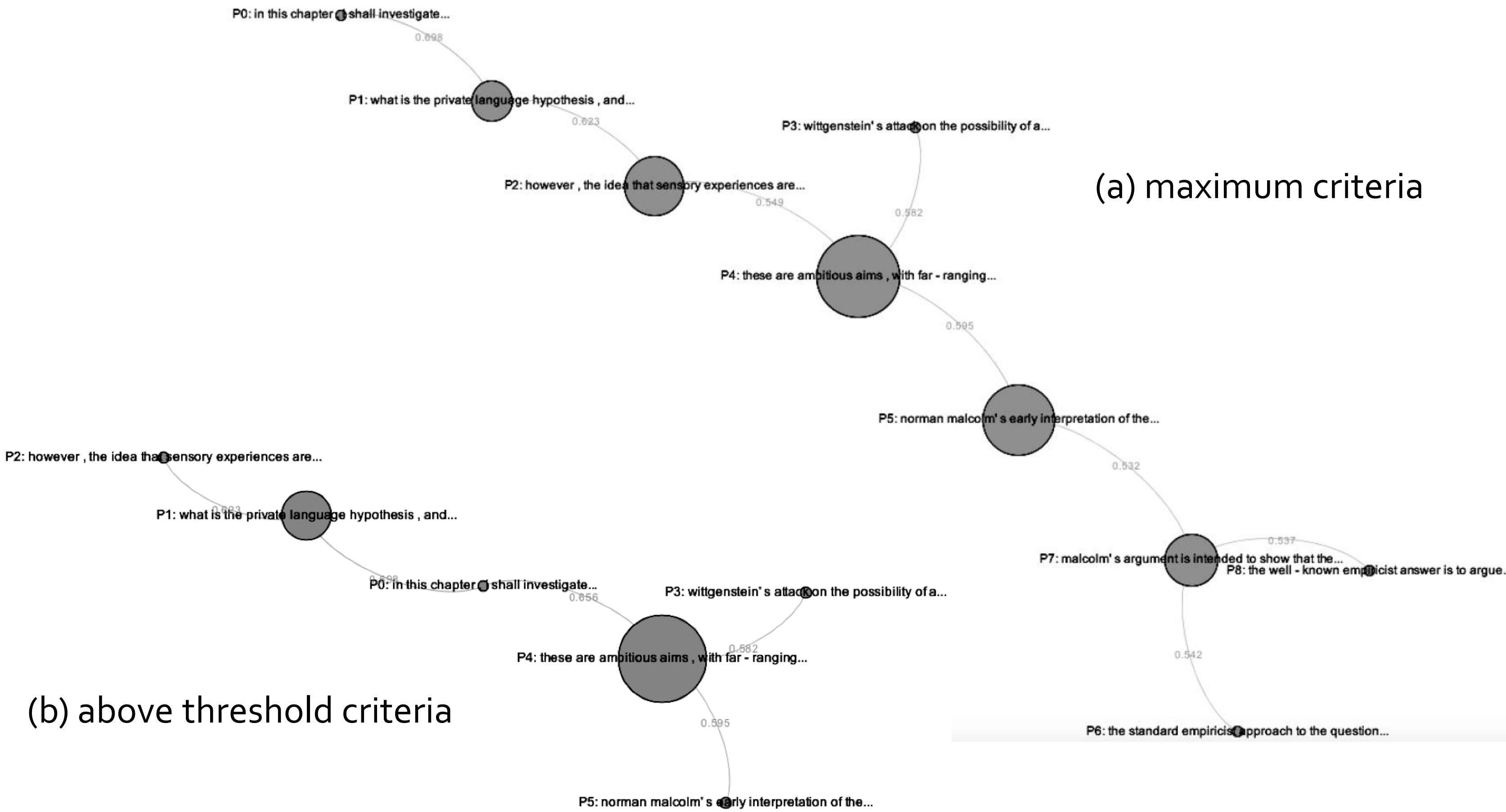
- Cohesion (Inner-block & Intra-block)
- Lexical chains (e.g., max span, avg. span, significance)
- *Co-references (e.g., no. chains, chain span)*
- *Entity-density features (e.g., overlapping nouns, no. entities / unique)*
- Discourse connectives
- Word features & vectors from linguistic resources (e.g., LIWC - Linguistic Inquiry and Word Count, ANEW - Affective Norms for English Words, *The General Inquirer: the Harvard IV-4*, *Lasswell*, *SenticNet*, *The Geneva Affect Label Coder*, *EmoLex*)
- Document cohesion flow

* EN only

Document Cohesion Flow

- Automated *cohesion* measures:
  - Lexical chains & semantic distances in ontologies – Wordnet
  - Semantic models: Latent Semantic Analysis [LSA], Latent Dirichlet Allocation [LDA] & word2vec
- Assess semantic global cohesion to capture text organization in terms of paragraph links
- Measure a document's structure derived from the order of paragraphs and of the manner in which they combine to hold the text together
- Develop a number of indices based on
  - *Building criteria*: Maximum value, above the threshold
  - *Ordering*: position and distance accuracy, adjacency accuracy, flow cohesion, correlations, max order sequence
- A text with strong cohesion flow should be more coherent

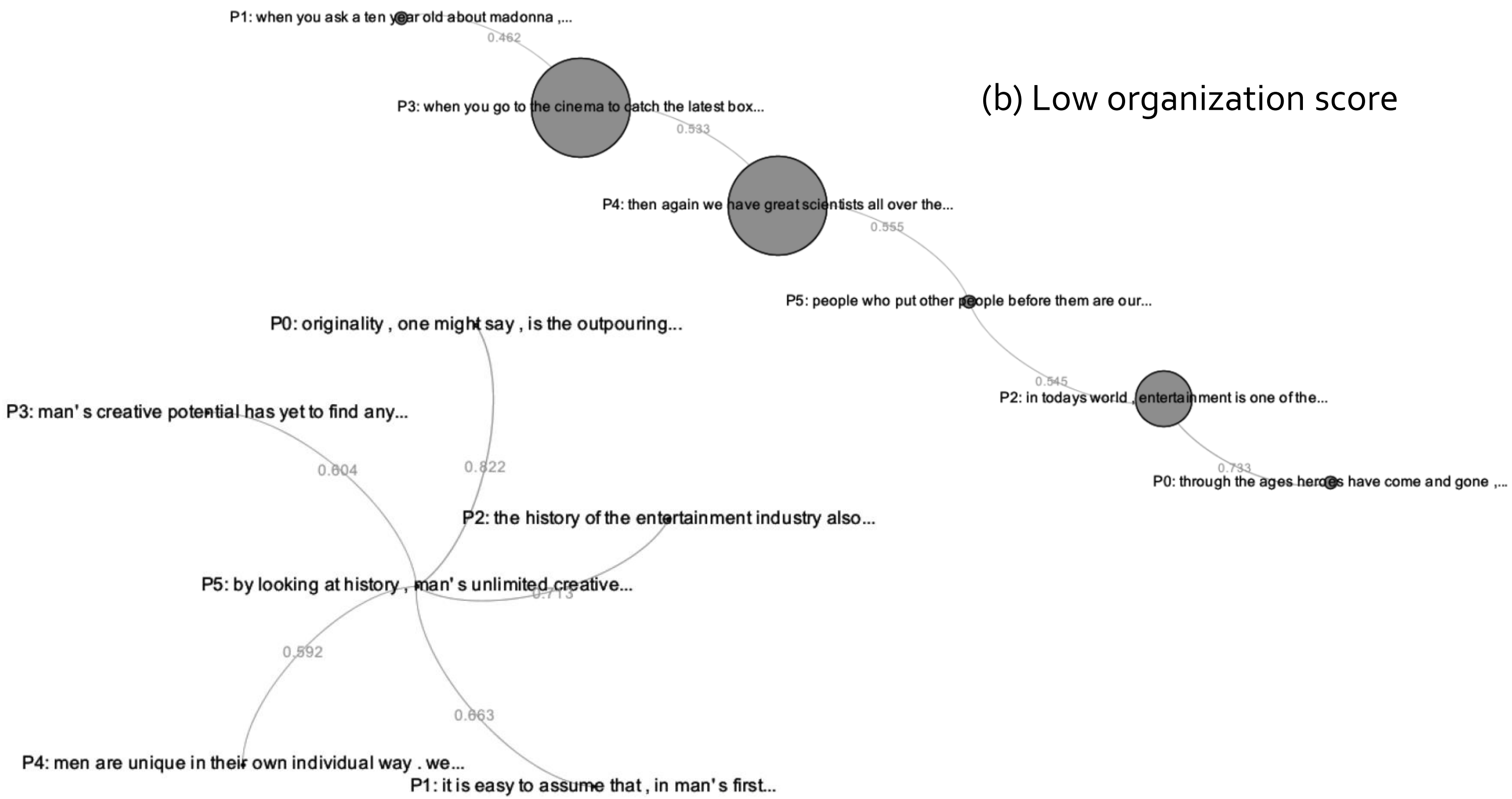
Cohesion Flow using LDA Semantic Similarity



(a) maximum criteria

(b) above threshold criteria

Differences between Essays Scored with High/Low Organization

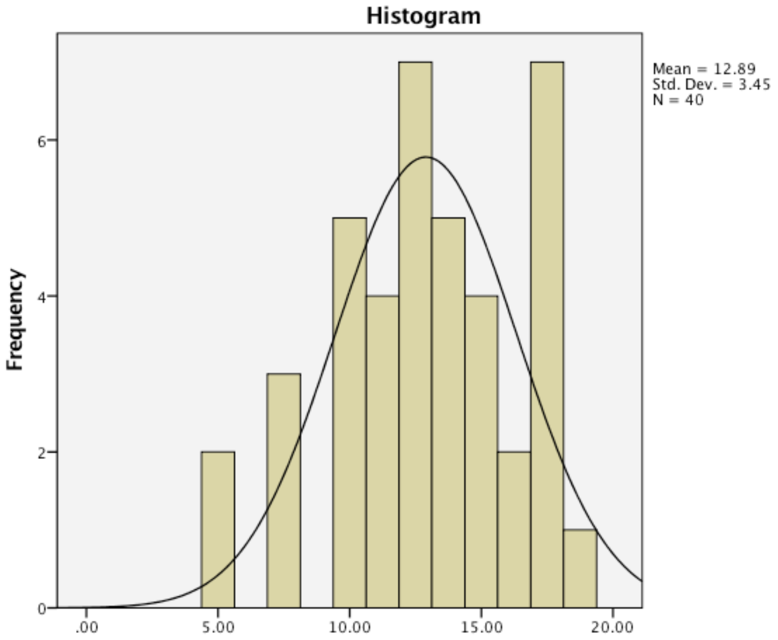


(b) Low organization score

(b) High organization score

French Experiment

- 40 essay written by 1st-year nurse students:
  - Case studies of infectious diseases and hygiene
  - Mean length: 1,342 words (SD = 293);
  - Min: 680 words; Max: 2,179 words
- Semantic models trained on *Le Monde* + 9 textbooks on infectious diseases and hygiene (273 pages ≅ 133,000 words)
- 2 categories: low / high performance students (<13 / ≥13)



Variable Selection

- Removed non-normal indices
- Removed variables that demonstrated multicollinearity ($r > .90$)
- Levene's test of equality of error variances not significant ($p > .05$)
- 3 indices chosen (15:1 ratio)
- Regression analysis:
  - One index accounted for 25% of variance in students' essay scores $F(1, 38) = 12.367$, $p < .001$, $r = .496$, $R^2 = .246$: *Document cohesion flow adjacent accuracy using Wu-Palmer distance and maximum criterion*

Correlations between *ReaderBench* Textual Complexity Indices and Essay Scores

Indices	*r*	*p*
Document cohesion flow adjacent accuracy using Wu-Palmer distance and maximum criterion	.496	.001
Document cohesion flow adjacent accuracy using path distance and above avg + stdev criterion	.451	.004
Content words (i.e., nouns, verbs, adjective and adverbs that are not considered stop-words by providing contextual information)	.448	.004
Average start-middle cohesion using path distance	-.446	.004
Average paragraph-document cohesion using path distance	-.436	.005
Average 'voice' paragraph entropy	.431	.005
Average paragraph-document cohesion using Wu-Palmer distance	-.405	.010

Tests of between-subjects Effects for Significantly Different Indices

Dependent Variable	*Mean (SD) low*	*Mean (SD) high*	*F*	*Sig.*	*Partial η^2*
Document cohesion flow adjacent accuracy using Wu-Palmer distance and maximum criterion	0.98 (0.61)	1.74 (0.54)	17.33	<.001	0.313
Document cohesion flow adjacent accuracy using path distance and above avg + stdev criterion	1.07 (0.63)	2.07 (0.85)	18.07	<.001	0.322
Content words	472.79 (122.10)	655.24 (139.67)	19.16	<.001	0.335
Average start-middle cohesion using path distance	0.48 (0.07)	0.41 (0.07)	9.03	.005	0.192
Average paragraph-document cohesion using path distance	0.76 (0.02)	0.74 (0.02)	6.27	.017	0.142
Average 'voice' paragraph entropy	1.14 (0.11)	1.26 (0.10)	15.15	<.001	0.285
Average paragraph-document cohesion using Wu-Palmer distance	0.863 (0.015)	0.855 (0.010)	4.18	.048	0.099

Essay Features

- Essays with higher scores tend to:
  - Be longer & contain more content words
  - Introduce more varied concepts, additional ideas (more ‘voices’)
    => decrease in global cohesion in terms of paragraph-document cohesion, start-middle cohesion
  - Higher entropy = presence of additional semantic chains
  - Have a better organization in terms of paragraph structure & a more suitable cohesion flow among adjacent paragraphs
    => a more coherent discourse.

Discriminant Function Analysis Results

- Retained two variables as significant predictors:
  - Content words
  - Document cohesion flow adjacent accuracy using path distance and above avg + stdev criterion)
- Accuracy: 82.5%, $\chi^2(df = 2, n = 40) = 19.015, p < .001$
- LOOCV Accuracy: 77.5%
- Cohen's Kappa of .652 – substantial agreement

Discussion

- *ReaderBench* Features:
  - Multi-dimensional analysis & Holistic view of AES
  - Intense use of NLP techniques
  - Integrated an extensible model by combining evaluation factors
  - Multi-language support
- Human categorization of professional case studies can be partly predicted by considering document flow features
- Limitations & extensions:
  - Low number of essays
  - Binary essay categorization can be refined with more classes
  - Integrate word specificity measures

The Team behind *ReaderBench*

Mihai Dascălu[1,2,3], Ștefan Trăușan-Matu[1], Philippe Dessus[2], Danielle McNamara[3]

Scott Crossley[4], Nicolae Nistor[5], Laura K. Allen[3], Daniela Gifu[6], Maryse Bianco[2], Traian Rebedea[2] , Elvira Popescu[7]

Gabriel Gutu[1], Stefan Ruseti[1], Ionuț Cristian Paraschiv[1], Dorinela Sirbu[1], Marilena Panaite[1], Irina Toma[1]

[1] University Politehnica of Bucharest, Romania

[2] University Grenoble Alpes, France

[3] Institute for the Science of Teaching & Learning, USA

[4] Georgia State University, USA

[5] Ludwig Maximilian University, Munich, Germany

[6] University Alexandru Ioan Cuza, Iasi, Romania

[7] University of Craiova, Craiova, Romania

Thank you!

This research was partially supported by the FP7 2008-212578 LTfLL project, by the 644187 EC H2020 *Realising an Applied Gaming Eco-system* (RAGE) project, as well as by University Politehnica of Bucharest through the "Excellence Research Grants" Programs UPB–GEX 12/26.09.2016.