

Linguistic Features in MELAB Writing Task Performances

Yeon Joo Jung, Scott A. Crossley & *Danielle S. McNamara

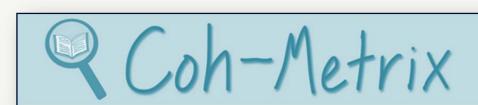
Georgia State University & *Arizona State University



INTRODUCTION

Human Judgments of L2 Writing Proficiency

- Significant role of linguistic features in human ratings of essay scores (e.g., Cummings et al., 2005, 2006; Engber, 1995; Grant & Ginther, 2000)
- Quantity of texts analyzed and range of linguistic features limited
- Still lacking a coherent understanding of linguistic features of L2 writing (Jarvis et al., 2003)
- Use of more advanced computational tools (e.g., Coh-Metrix) along with larger corpora needed to validate the role of linguistic features in L2 essay quality(Crossley & McNamara, 2010)



Coh-Metrix & Human Ratings of L2 Essay Quality

- Significant predictive power of some linguistic features in human judgments of L2 writing proficiency (Crossley & McNamara, 2012; Guo, Crossley & McNamara, 2013)
- Can help validate scoring rubrics by showing which linguistic features are more attended by human raters in assessing L2 essays (Guo, Crossley & McNamara, 2013)

Analysis of Performance Data from Large-Scale Language Tests

- Important source for characterizing L2 proficiency (Iwashita, Brown, T. McNamara & O'Hagan, 2008)
- Multi-dimensional analyses of performance data using multiple linguistic features needed to understand the development of L2 proficiency (Norris & Ortega, 1999)

RESEARCH OBJECTIVE

- To uncover the relationships between human judgments of L2 writing proficiency and language features that differ as a function of these judgments
- To provide strong empirical evidence in reference to the linguistic, rhetorical, and structural features of learners' performance on the writing task included in the Michigan English Language Assessment Battery (MELAB)
- To help validate the current MELAB composition rating scale

Research Question

What linguistic features, as measured by Coh-Metrix, distinguish MELAB test-taker writing performance as represented in a single holistic score given by expert raters on the basis of the ten-level MELAB composition rating scale?

ACKNOWLEDGEMENT

This presentation reports on research funded through CaMLA's Spaan Research Grant Program 2014

METHODS

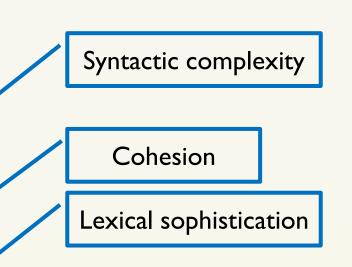
MELAB Writing Task

- Michigan English Language Assessment Battery
- Independent writing task
- 200-300 word-long essay about one of two topic choices
- 30-minute time limit

Human Ratings of MELAB Essays

- A locally developed 10-level holistic rating scale (i.e., 97, 93, 87,, 57, 53)
- 19 scores possible including midpoints between levels
- Rated independently by two trained raters
- Two scores averaged for the final score
- Composition descriptions for a successful essay

Topic is richly and fully developed. Flexible use of a wide range of syntactic (sentence-level) structures, accurate morphological (word forms) control. Organization is appropriate and effective, and there is excellent control of connection. There is a wide range of appropriately used vocabulary. Spelling and punctuation appear error free.



Corpus Collection

- 1,003 essays from the MELAB writing tests administered in 2013
- Stratified according to score level, gender & age
- Test-takers from 62 different L1 backgrounds
- 500 essays analyzed for this preliminary study

Variable Selection

Table ISummary of Pre-selected Coh-Metrix Indices for Regression Analysis

Category	Coh-Metrix measures	N of indices	
Basic text information	Text length	2	
Lexical sophistication	Word polysemy value	1	
	Word hypernymy value	1	
	Word frequency	3	
	Word information (familiarity, concreteness, imageability & meaningfulness)	4	
	Lexical diversity	1	
Syntactic complexity	N of words before the main verb	1	
	N of modifiers per noun phrase	1	
	Syntactic similarity	2	
	POS tags	6	
Cohesion	Lexical overlap	4	
	Causality	3	
	Connectives	2	
	Semantic similarity	3	

FINDINGS

Correlations between Score and Coh-Metrix Index: Training Set

Selected Coh-Metrix Indices for Regression Analysis: Training Set Category Coh-Metrix indices r value Number of words per text Basic text information .535 Lexical diversity for all words Lexical sophistication Lexical sophistication Word familiarity Word frequency (content words) Lexical sophistication Content word overlap -.317 Personal pronoun Syntactic complexity Number of modifiers per noun phrase Syntactic complexity .316 Word frequency (all words) Lexical sophistication Causal connectives Number of paragraphs per text Lexical sophistication Word meaningfulness Number of words before the main verb Syntactic complexity Hypernymy value Lexical sophistication Lexical sophistication Adverbs Semantic similarity (LSA paragraph to paragraph) .193 Semantic similarity (LSA sentence to sentence) -.176 Cohesion -.135 All connectives Cohesion -.133 Cohesion Noun overlap

Regression Analysis: Training Set

Table 2.

Entry	Coh-Metrix index added	r	r ²	В	В	S.E.
Entry I	Number of words per text	.535	.286	.054	.535	.00
Entry 2	Word frequency (content words)		.433	-22.592	389	2.440
Entry 3	Lexical diversity		.475	.098	.239	.019
Entry 4	Word meaningfulness		.498	090	158	.023
Entry 5	Semantic similarity (LSA paragraph to paragraph)		.515	6.974	.132	2.07
Entry 6	Number of modifiers per noun phrase		.527	7.072	.126	2.41
Entry 7	Content word overlap		.537	-17.807	121	6.76
Entry 8	Number of words before the main verb	.738	.545	.477	.096	.204
Entry 9	Causal connectives	.742	.551	049	084	.023

- $F(9, 325) = 44.268, p = .000, r = .742, r^2 = .551$
- 9 indices found to be significant predictors of the essay scores
- 55.1% of the variance in the scores explained by the reported model

Regression Analysis: Test Set

- The regression model extended to the test set (166 essays): r = .735, $r^2 = .540$
- 54% of the variance of human scores explained by the model
- Generalizability of the model

Preliminary Conclusions

- Provides evidence that linguistic features can predict human ratings of the essays in the MELAB writing task
- Contributes to the validation of raters' use of the MELAB composition rating scale by verifying which linguistic features are meaningfully related to the writing aspects specified in the scale
- Linguistic features associated with text length and lexical sophistication have greater predictive value
- Findings extendable to the MELAB test-taker population

CONTACT

Yeon Joo Jung, PhD student

Department of Applied Linguistics & ESL, Georgia State University yjung9@student.gsu.edu

