Lexical concreteness in narrative

Michael Flor
Educational Testing Service
Princeton
NJ 08541, USA
mflor@ets.org

Swapna Somasundaran
Educational Testing Service
Princeton
NJ 08541, USA
ssomasundaran@ets.org

Abstract

This study explores the relation between lexical concreteness and narrative text quality. We present a methodology to quantitatively measure lexical concreteness of a text. We apply it to a corpus of student stories, scored according to writing evaluation rubrics. Lexical concreteness is weakly-to-moderately related to story quality, depending on story-type. The relation is mostly borne by adjectives and nouns, but also found for adverbs and verbs.

1 Introduction

The influential writing-style guide, *The Elements of Style* (1999), (a.k.a. *Strunk and White*), recommends writers to '*prefer the specific to the general, the definite to the vague, the concrete to the abstract.*' This involves two related but distinct notions, two different senses of the word 'concrete' - *tangible* and *specific*. Tangibility, or the *concreteness/abstractness continuum* relates to objects and properties that afford sensorial perception - tangible things that can be seen, heard, smelled and touched. The *specificity* notion relates to the amount and level of detail that is conveyed in a story, to what extent things are presented in specific rather than general terms. The two notions go hand in hand, since to provide specific details the writer often has to mention more concrete objects and attributes and use less abstract terms. There are exceptions. Emotions and states of mind are usually not concrete (i.e. tangible) entities, though they are often specific. Numerical quantities (e.g. *6 million dollars, 30% of the population*) are quite specific but not quite sensorially concrete. Still, the importance of both concreteness and specificity for good writing is frequently mentioned in writer guides (Hacker and Sommers, 2014), in advice to college students (Maguire, 2012) and in recommendations for business writers (Matson, 2017).

College writing labs often suggest that students can improve their writing by including more concrete details in their essays.[1] Concreteness is also noted as an important aspect of writing literacy for K-12 education. The Common Core State Standards[2] (a federally recommended standard in the USA) specifies the following capability for students in Grade 6: "*Develop the topic with relevant facts, definitions, **concrete** details, quotations, or other information and examples.*" Despite its purported importance, few studies have measured lexical concreteness in stories, and no studies explored a quantitative relation between concreteness and story quality.

This work explores lexical concreteness in narrative essays. We use a quantitative measure, utilizing per-word concreteness ratings. We investigate whether better stories are more concrete and whether the story type (e.g. hypothetical situation versus personal narratives) influences the concreteness trends. We also perform a fine-grained analysis by parts-of-speech (nouns, verbs, adjectives and adverbs) to explore how their concreteness varies with story quality.

2 Related Work

The literature on using lexical concreteness for analysis of writing is rather limited.[3] Louis and Nenkova (2013) used imageability of words as a feature to model quality of science-journalism writing. For reading, concrete language was found to be more comprehensible and memorable than abstract language (Sadoski et al., 2000, 1993). Concreteness has also been related to reader engagement, promoting interest for expository materials (Sadoski, 2001).

---

[1] For example, see Purdue University and Roane State
[2] corestandards.org
[3] For recent research on specificity, see (Lugini and Litman, 2017; Li and Nenkova, 2015; Louis and Nenkova, 2011)

Researchers have also looked at developmental aspects of mastery in producing expository and narrative texts. Proportion of abstract nouns in language production increases with age and schooling, although it is more pronounced in expository than in narrative writing (Ravid, 2005). Berman and Nir-Sagiv (2007) have found that the proportion of very concrete nouns tends to decrease from childhood to adulthood, whereas the proportion of abstract nouns tends to increase, in both expository and narrative texts. Sun and Nippold (2012) conducted a study in which students ages 11-17 were asked to write a personal story. The essays were examined for the use of abstract nouns (e.g., *accomplishment, loneliness*) and metacognitive verbs (e.g., *assume, discover*). The use of both types of words significantly increases with age. Goth et al. (2010) analyzed fables created by sixth graders (age 12) and found that boys use more concrete terms than girls.

How are concrete and abstract words identified and measured is an important methodological point. Goth et al. (2010) used the Coh-Metrix tool (Graesser et al., 2004), which measured individual word concreteness "*using the hypernym depth values retrieved from the WordNet lexical taxonomy, and averaged across noun and verb categories.*" Berman and Nir-Sagiv (2007) rated nouns manually, using a four-level ordinal ranking. The most concrete (level 1) included objects and specific people; level 2 - categorial nouns, roles and locations (*teacher, city, people*). Higher abstractions were: level 3 - rare nouns (e.g., *rival, cult*), and abstract but common terms such as *fight, war*; level 4: low frequency abstract nouns (e.g. *relationship, existence*). Sun and Nippold (2012) used a dichotomous distinction (abstract/non-abstract) while manually rating all nouns in their data set. Abstract nouns were defined as intangible entities, inner states and emotions.

In psycholinguistic research, the notion of word concreteness became prominent due to the dual-coding theory of word representation (Pavio, 2013, 1971). Experimental studies often utilize the MRC database (Coltheart, 1981), which provides lexical concreteness ratings norms for 4,292 words. Such ratings were obtained experimentally, averaging across ratings provided by multiple participants in rating studies. Recently, Brysbaert et al. (2013) provided concreteness norms for 40,000 English lemmas. This new

Prompt	Count essays	Text Type
A Fork in the Road	47	Fictional
At First Glance	69	Fictional
Finding Your Way Home	2	Fictional
Message in a Bottle	31	Fictional
Movie Sequel	12	Fictional
Pitch Session	6	Fictional
Special Object	37	Fictional
The Antique Trunk	8	Fictional
The Quest	6	Fictional
Different Country	47	Hypothetical
Electricity-Free	32	Hypothetical
Living Art	3	Hypothetical
Trading Places	22	Hypothetical
Weirdest Day Ever!	78	Hypothetical
You are the Teacher	121	Hypothetical
Travel	75	Personal
Memorable School Day	153	Personal
Proudest Moment	191	Personal
Totals	171	Fictional
	303	Hypothetical
	466	Personal

Table 1: Essay counts for 18 prompts and their text-type classifications.

database opens the possibility for wide-coverage automated analysis of texts for estimating concreteness/abstractness. We utilize this resource for analyzing stories produced by students, and investigate the relation between concreteness and quality of narrative.

3 Data

We used a corpus of narrative essays[4] provided by Somasundaran et al. (2018). The corpus consists of 940 narrative essays written by school students from grade levels 7-12. Each essay was written in response to one of 18 story-telling prompts. The total size of the corpus is 310K words, and average essay length is 330 words.

The writing prompts were classified according to the type of story they are calling for, using the three-fold schema from Longobardi et al. (2013) - Fictional, Hypothetical and Personal. Table 1 presents the prompt titles, story types and essay counts. Example prompts are shown in the appendix.

---

[4]i.e. stories, not expository or persuasive writing

3.1 Essay scores

All essays were manually scored by experienced research assistants (Somasundaran et al., 2018), using a rubric that was created by education experts and teachers, and presented by Smarter Balanced assessment consortium, an assessment aligned to U.S. State Standards for grades K-12 (Smarter Balanced, 2014b,a).

The essays were scored along three traits (dimensions): Organization, Development and Conventions. *Organization* is concerned with event coherence, whether the story has a coherent start and ending, and whether there is a plot to hold all the pieces of the story together. It is scored on a scale of 0-4 integer points. *Development* evaluates whether the story provides vivid descriptions, and whether there is character development. It is also scored on a scale of 0-4 integer points, with 4 being the highest score. The *Conventions* dimension evaluates language proficiency, and is concerned with aspects of grammar, mechanics, and punctuation. Scores are on a scale of 0-3 integer points (3 is the highest score).

Somasundaran et al. (2018) computed Narrative and Total composite scores for each essay. The Narrative score (range 0-8) is the sum of Organization and Development scores. Total score (range 0-11) is the sum of Organization, Development and Conventions. Not surprisingly, the Organization, Development, Narrative and Total scores are highly intercorrelated (0.88 and higher, see Table 3 in Somasundaran et al. (2018)). For the present study, we used the Narrative scores, focusing on essay narrative quality and de-emphasizing grammar and mechanics.

POS	Count	Missing values
nouns	64,374	2,113 (3.3%)
verbs	66,718	753 (1.1%)
adjectives	19,090	658 (3.45%)
adverbs	19,399	212 (1.1%)
all content words	169,581	3,736 (2.2%)

Table 2: Content word counts by part-of-speech, with counts and proportion of tokens that did not have concreteness scores, for 940 essays.

3.2 Concreteness scores

We focus on concreteness of only the content words in the essays and ignore all function words. Each essay in the data set was POS-tagged with the Apache OpenNLP [5] tagger, and further analysis filtered in only nouns, verbs, adjective and adverbs. Those content words were checked against the database of concreteness scores (Brysbaert et al., 2013). The database provides real-valued ratings in the 1-5 range, from very abstract (score 1) to very concrete (score 5.0). For words that were not matched in the database, we checked if the lemma or an inflectional variant of the word was present in the database (using an in-house morphological toolkit). The database does not include names, but the essays often include names of persons and places. For our scoring, any names (identified by POS-tags *NNP* or *NNPS*), that were not found in the database, were assigned a uniform concreteness score of 4.0.

Concreteness scores were accumulated for each essay as described above. Average and median concreteness score was computed for each essay, separately for each of the categories (nouns, verbs, adjective and adverbs), and also jointly for all content-words. The total numbers of content words are given in Table 2. The concreteness-ratings coverage for our data is 97.8%.

4 Results

Pearson correlations of essay scores with per-essay levels of concreteness are presented in Table 3. Overall, the correlation of average-concreteness with essay score is r=0.222, which is considered a weak correlation (Evans, 1996). Breakdown by parts of speech shows that adjectives have the highest correlation of concreteness with score (0.297), followed by that for nouns (0.251), and adverbs (0.231). The correlation is weakest for verbs, only 0.122. Results for median-concreteness per essay show a similar pattern, though nouns now overtake adjectives.

	Average C.	Median C.
nouns	0.251	0.284
verbs	0.122	0.113
adjectives	0.297	0.242
adverbs	0.231	0.132
all content words	0.222	0.188

Table 3: Pearson correlations of essay narrative scores with per-essay levels of concreteness, for 940 essays. All correlations are significant, $p < .001$. C.=concreteness score

---

[5] http://opennlp.apache.org

(A) Prompt	N	Nouns	Verbs	Adjectives	Adverbs	All CW
Travel	75	0.400**	-0.017	0.401**	0.268*	0.371**
At First Glance	69	0.404**	0.006	0.326**	0.286*	0.240†
Memorable School Day	153	0.080	0.040	0.212**	0.239**	0.089
Proudest Moment	191	0.207**	0.072	0.118	0.060	0.137
Weirdest Day Ever	78	0.125	0.326**	0.355**	0.330**	0.322**
You are the Teacher	121	0.218*	0.102	0.298**	0.131	0.071
(B) Story type						
Fictional	171	0.465**	0.164†	0.417**	0.384**	0.413**
Hypothetical	303	0.263**	0.222**	0.287**	0.143*	0.217**
Personal	466	0.199**	0.045	0.237**	0.209**	0.138**

Table 4: Pearson correlations of essay narrative scores with per-essay average levels of concreteness; (A) for prompts that have above 60 essays, (B) for all essays, grouped by story-type. Significance of correlation **: $p < 0.01$, *: $p < .03$, $\dagger$: $p < .05$. CW=content words.

Next, we present the correlations of concreteness levels with essay scores for each of the six prompts that have more than 50 essays (Table 4A). For two of the prompts, *Travel* and *At First Glance*, average concreteness of nouns is moderately correlated with essay narrative score ($r = 0.4$). For four prompts, adjectives show weak correlation with essay scores (from 0.21 to 0.35), while for the *Travel* prompt, average concreteness of adjectives is moderately correlated with essay narrative score ($r=0.4$). For four prompts, the average concreteness of adverbs is weakly correlated with essay score (0.24 to 0.33). For verbs, only one prompt, *Weirdest Day Ever.* shows some correlation of concreteness with essay score (0.33).

Next, we grouped essays by three types of story that their prompts were classified into. This move allows to use data from all essays. Results are presented in Table 4B. The *Fictional* story type has the highest correlation of concreteness and essay score ($r=0.413$), and it also has the highest correlation for nouns, for adjectives and for adverbs (as compared to other story types). Stories of the *Hypothetical* type show weak (yet significant) correlation of concreteness with scores, for nouns, verbs, adjectives and overall. Interestingly, the *Personal* story type shows the least relation of concreteness to scores, 0.138 overall; the adjectives there have correlation of 0.237, adverbs 0.209, and the nouns barely reach 0.2.

The results above suggest that the relation of concreteness to essay score varies for different story types. We checked whether the essays from three story types differ in concreteness or quality. An analysis of variance of narrative scores for three groups, F(2,937)=1.427, p=0.241, reveals that they did not differ in the average quality of stories. We also compared the average per-essay concreteness for the three groups. Mean concreteness for *Fiction* essays is 2.91, for *Hypothetical* essays it is 2.99, and 2.90 for *Personal*. An analysis of variance, F(2,937)=19.774, p<0.0001, shows that average concreteness is not equal in those groups. Post hoc comparisons (Tukey HSD test) indicated that only the *Hypothetical* group differed significantly from the two other groups. Those results indicate that the different strength of correlation between lexical concreteness and essay score that we observe in the three groups is not due to between-group differences in either narrative scores or lexical concreteness.

5   Conclusions

We presented a novel methodology for computing per-text lexical concreteness scores. For student-written stories, lexical concreteness is weakly to moderately positively correlated with narrative quality. Better essays score higher on lexical concreteness and use relatively less abstract words. While those results support the old adage 'prefer the concrete to the abstract', we also found that this relation varies for different story-types. It is prominent for *Fictional* stories, less pronounced for *Hypothetical* stories, and rather weak for *Personal* stories. Nouns and adjectives carry this relation most prominently, but it is also found for adverbs and verbs. This study lays the groundwork towards automatic assessment of lexical concreteness. In future work we will extend its application for text evaluation and feedback to writers.

References

Ruth A. Berman and Bracha Nir-Sagiv. 2007. Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse processes*, 43(2):79–120.

Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2013. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46:904–911.

Max Coltheart. 1981. The MRC psycholinguistic database. *Journal of Experimental Psychology*, 33:497–505.

James D. Evans. 1996. *Straightforward Statistics for the Behavioral Sciences*. Brooks/Cole Pub. Co., Pacific Grove, CA, USA.

Julius Goth, Alok Baikadi, Eun Ha, Jonathan Rowe, Bradford Mott, and James Lester. 2010. Exploring individual differences in student writing with a narrative composition support environment. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing: Writing processes and authoring aids*, pages 56–64. Association for Computational Linguistics.

Arthur Graesser, Danielle McNamara, and Max Louwerse. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods Instruments and Computers*, 36(2):193–202.

Diana Hacker and Nancy Sommers. 2014. *The Bedford Handbook*, 9 edition. Bedford/St. Martins, Boston, MA, USA.

William Strunk Jr. and Edward A. Tenney. 1999. *The Elements of Style*, 4 edition. Pearson, Harlow, UK.

Junyi Jessy Li and Ani Nenkova. 2015. Fast and Accurate Prediction of Sentence Specificity. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 2381–2387. Association for the Advancement of Artificial Intelligence.

Emiddia Longobardi, Pietro Spataro, Maria-Luisa Renna, and Clelia Rossi-Arnaud. 2013. Comparing fictional, personal, and hypothetical narratives in primary school: Story grammar and mental state language. *European Journal of Psychology of Education*, 29:257–275.

Annie Louis and Ani Nenkova. 2011. Automatic identification of general and specific sentences by leveraging discourse annotations. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 605–613, Chiang Mai, Thailand. Asian Federation of Natural Language Processing.

Annie Louis and Ani Nenkova. 2013. What makes writing great? First experiments on article quality prediction in the science journalism domain. *Transactions of the Association for Computational Linguistics*, 1:341–352.

Luca Lugini and Diane Litman. 2017. Predicting specificity in classroom discussion. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 52–61, Copenhagen, Denmark. Association for Computational Linguistics.

John Maguire. 2012. The Secret to Good Writing: It's About Objects, Not Ideas. *The Atlantic*.

Owen Matson. 2017. Why Concrete Language is Essential to Engaging Content.

Allan Pavio. 1971. *Imagery and verbal processes*. Holt, Rinehart & Winston, Oxford, England.

Allan Pavio. 2013. Dual Coding Theory, Word Abstractness, and Emotion: A Critical Review of Kousta et al. (2011). *Journal of Experimental Psychology: General*, 142(1):282–287.

Dorit Ravid. 2005. Emergence of linguistic complexity in later language development: Evidence from expository text construction. In *Perspectives on language and language development*, pages 337–355. Springer.

Mark Sadoski. 2001. Resolving the effects of concreteness on interest, comprehension, and learning important ideas from text. *Educational Psychology Review*, 13(3):263–281.

Mark Sadoski, Ernest T. Goetz, and Joyce B. Fritz. 1993. Impact of concreteness on comprehensibility, interest, and memory for text: Implications for dual coding theory and text design. *Journal of Educational Psychology*, 85(2):291–304.

Mark Sadoski, Ernest T. Goetz, and Maximo Rodriguez. 2000. Engaging Texts: Effects of Concreteness on Comprehensibility, Interest, and Recall in Four Text Types. *Journal of Educational Psychology*, 92(1):85–95.

Smarter Balanced. 2014a. English Language Arts/Literacy Item Specifications - Narrative.

Smarter Balanced. 2014b. Scoring Guide For Grades 3, 6, and 11, English/Language Arts Performance Task full-write baseline sets.

Swapna Somasundaran, Michael Flor, Martin Chodorow, Hillary Molloy, Binod Gyawali, and Laura McCulla. 2018. Towards evaluating narrative quality in student writing. *Transactions of the Association for Computational Linguistics*, 6:91–106.

Lei Sun and Marilyn A. Nippold. 2012. Narrative writing in children and adolescents: Examining the literate lexicon. *Language, speech, and hearing services in schools*, 43(1):2–13.

A   Appendix

Example prompts for three types of text styles:

Personal Experience: "Proudest Moment" - *There are moments in everyones lives when they feel pride and accomplishment after completing a challenging task. Write a story about your proudest moment.*

Hypothetical Situation: "You are the Teacher" - *Pretend that one morning you wake up and find out that you've become your teacher for a day! What happened? What do you do? Do you learn anything? Write a story about what happens. Use your imagination!*

Fictional Story: "Message in a Bottle" - *Throughout the years, many have placed messages in sealed bottles and dropped the bottles into the ocean where they eventually washed up on foreign shores. Occasionally the finder has even contacted the sender. Write a story about finding your own message in a bottle.*