CLAREO

Methodology & Metrics Guide

Introduction

Clareo analyzes writing using quantitative metrics drawn from computational linguistics research. This document explains each metric: what it measures, how it is calculated, and how to interpret the results.

These metrics describe linguistic features of text. They do not determine authorship or intent. A high score in any metric simply indicates a pattern is present, not why it is there.

1. Grade Level (Flesch-Kincaid)

Grade Level estimates the U.S. school grade needed to understand a text. It is based on two factors: sentence length and word complexity (measured by syllables).

The Formula

```
Grade = 0.39 x (words/sentences) + 11.8 x (syllables/words) -
15.59
```

What Each Part Means

Words per sentence: Longer sentences are harder to follow. A 25-word sentence is more demanding than a 10-word sentence.

Syllables per word: Multi-syllable words (like "approximately") are typically harder than short ones (like "about").

Example Calculation

Consider this 50-word passage with 4 sentences and 78 syllables:

Average sentence length: 50 / 4 = 12.5 words

Average syllables per word: 78 / 50 = 1.56 syllables

Grade Level: (0.39 x 12.5) + (11.8 x 1.56) - 15.59 = 4.88 + 18.41 - 15.59 = **7.7**

This text is appropriate for a 7th-8th grade reader.

Interpretation

Grade Level	Typical Audience	Example Publications
5-6	General public	USA Today, most fiction
7-9	High school	Time Magazine, YA novels
10-12	College	Academic journals, legal documents
13+	Graduate/Professional	Scientific papers, technical manuals

2. Sentence Variance (CV)

Sentence Variance measures how much sentence lengths differ from each other. It uses the Coefficient of Variation (CV), which expresses the standard deviation as a percentage of the mean.

The Formula

```
CV = (Standard Deviation / Mean) x 100
```

Step-by-Step Calculation

1. Count words in each sentence. Example: 8, 15, 12, 22, 10 words

2. Calculate the mean (average). (8 + 15 + 12 + 22 + 10) / 5 = 13.4 words

3. Find how far each sentence is from the mean. Differences: -5.4, +1.6, -1.4, +8.6, -3.4

4. Square each difference. 29.16, 2.56, 1.96, 73.96, 11.56

5. Average the squared differences. (29.16 + 2.56 + 1.96 + 73.96 + 11.56) / 5 = 23.84

6. Take the square root (this is the standard deviation). Square root of 23.84 = 4.88

7. Divide by the mean and multiply by 100. (4.88 / 13.4) x 100 = **36.4%**

Why CV Matters

Human writing naturally varies. We use short punchy sentences for emphasis. Then we might elaborate with longer, more complex structures that develop an idea. This creates rhythm.

Formulaic writing often has uniform sentence lengths. Every sentence hovers around the same word count, creating a monotonous rhythm.

Interpretation

CV Value	Interpretation
< 25%	Low variation: sentences are very uniform in length
25-40%	Moderate variation: typical of natural writing
40-60%	High variation: mix of short and long sentences

| > 60% | Very high variation: may include fragments or very long sentences |

3. Vocabulary Variety (sTTR)

Vocabulary Variety measures how diverse a writer's word choices are. Clareo uses the Standardized Type-Token Ratio (sTTR) to account for text length.

Basic Type-Token Ratio

The simple Type-Token Ratio (TTR) divides unique words by total words:

```
TTR = Unique Words / Total Words
```

Problem: TTR naturally decreases as texts get longer because we repeat common words ("the," "is," "and"). A 100-word text will almost always have a higher TTR than a 1,000-word text by the same author.

Standardized TTR (sTTR)

sTTR solves this by calculating TTR on consistent chunk sizes and averaging them:

1. Divide the text into 50-word chunks

2. Calculate TTR for each chunk

3. Average all chunk TTRs

Example

A 200-word text produces 4 chunks of 50 words each:

Chunk	Unique Words	TTR
1	38	76%
2	35	70%
3	37	74%
4	36	72%

sTTR = (76 + 70 + 74 + 72) / 4 = 73%

Interpretation

sTTR Value	Interpretation
< 60%	Limited vocabulary: significant word repetition
60-70%	Average vocabulary diversity

70-80%	Good vocabulary diversity
> 80%	Very high diversity: few repeated words

4. Formulaic Language (Weighted Score)

This metric counts stock phrases and cliches that appear frequently in template-driven or overly formal writing. Each phrase is assigned a weight based on how distinctive it is.

Weighting System

Weight	Description	Examples
3 (High)	Distinctive stock phrases rarely used naturally	delve into, rich tapestry, myriad, plethora, beacon
2 (Medium)	Common in formal/template writing	leverage, harness, journey, foster, transformative
1 (Low)	Can appear naturally but suspicious in clusters	crucial, vital, furthermore, moreover, utilize

Calculation

The weighted score multiplies each phrase's count by its weight:

`Score = Sum of (count x weight)` *for each phrase found*

Example

A text contains:

- "delve" appears 2 times (weight 3): 2 x 3 = 6

- "leverage" appears 1 time (weight 2): 1 x 2 = 2

- "furthermore" appears 3 times (weight 1): 3 x 1 = 3

Total weighted score: 6 + 2 + 3 = **11**

Interpretation

Score	Interpretation
0-5	Minimal formulaic language
6-10	Some stock phrases present
11-20	Notable formulaic patterns
> 20	Heavy reliance on stock phrases

Context matters: academic writing legitimately uses more formal connectors. Compare against the student's baseline rather than absolute thresholds.

5. Predictability (N-gram Score)

N-grams are sequences of consecutive words. Bigrams are two-word sequences; trigrams are three-word sequences. Highly predictable writing overuses common sequences.

What We Count

Bigrams (2 words): "it is," "in order," "based on," "can be," "important to"

Trigrams (3 words): "it is important," "in order to," "one of the," "plays a crucial"

Calculation

1. Count how many times each tracked n-gram appears in the text

2. Calculate the rate per 100 words

3. Compare to a baseline (approximately 15 bigrams and 3 trigrams per 100 words is typical)

4. Calculate excess over baseline

> `Predictability = (bigram excess x 2) + (trigram excess x 5)` *capped at 100*

Example

In a 200-word text:

- Found 42 tracked bigrams: 42 / 200 x 100 = 21 per 100 words (baseline: 15)

- Found 10 tracked trigrams: 10 / 200 x 100 = 5 per 100 words (baseline: 3)

- Bigram excess: 21 - 15 = 6

- Trigram excess: 5 - 3 = 2

Predictability: (6 x 2) + (2 x 5) = 12 + 10 = **22%**

Interpretation

Score	Interpretation
0-15%	Natural phrase patterns
16-30%	Somewhat predictable

| 31-50% | Notably formulaic phrasing |
| > 50% | Highly predictable patterns |

6. Paragraph Coherence

Coherence measures how well ideas flow between paragraphs. Clareo evaluates two factors: shared vocabulary and transition usage.

Shared Vocabulary

Content words are nouns, verbs, adjectives, and adverbs that carry meaning (excluding common words like "the," "is," "very"). When adjacent paragraphs share content words, ideas are connected.

For each pair of adjacent paragraphs:

1. Extract content words from paragraph 1

2. Extract content words from paragraph 2

3. Count how many words appear in both

Transition Rate

Transitions are words that signal relationships between ideas: "however," "therefore," "additionally," "meanwhile."

```
Transition Rate = (paragraphs starting with transitions / total
paragraphs) x 100
```

Coherence Score Calculation

```
Coherence = (Average Shared Words x 20) + (Transition Rate x 0.5)
```
capped at 100

Example

A 5-paragraph essay:

- Paragraphs 1 to 2 share 3 content words

- Paragraphs 2 to 3 share 2 content words

- Paragraphs 3 to 4 share 4 content words

- Paragraphs 4 to 5 share 1 content word

Average shared: (3 + 2 + 4 + 1) / 4 = 2.5 words

Transitions: 3 of 5 paragraphs start with transitions = 60%

Coherence: (2.5 x 20) + (60 x 0.5) = 50 + 30 = **80%**

Interpretation

Score	Interpretation
< 30%	Weak connections between paragraphs
30-50%	Some coherence, room for improvement
51-70%	Good paragraph flow
> 70%	Strong coherence and idea development

7. Passive Voice

Passive voice occurs when the subject receives the action rather than performing it.

Active: "The researcher conducted the experiment."

Passive: "The experiment was conducted by the researcher."

Detection Pattern

Clareo looks for: [form of "to be"] + [past participle]

Forms of "to be": is, are, was, were, been, being, be

Past participles: words ending in -ed ("conducted") or irregular forms ("written," "taken," "done")

Calculation

```
Passive Ratio = (passive constructions / total sentences) x 100
```

Interpretation

Passive voice is not inherently bad. It is appropriate in scientific writing and when the actor is unknown or unimportant. However, overuse can make writing feel impersonal.

Ratio	Interpretation
< 10%	Very active voice
10-20%	Balanced use
21-30%	Moderate passive use (common in academic writing)
> 30%	Heavy passive use

8. Comparing Against a Baseline

The most powerful feature of Clareo is comparing new work against a student's established writing profile. This tells you whether the new submission matches their typical patterns.

Building a Reliable Baseline

1. Collect 3+ writing samples you are confident represent the student's authentic work

2. Use samples of similar length and genre when possible

3. The baseline averages all metrics across the samples

Reading the Comparison

Clareo shows the difference between the baseline and new text. Look for:

Sudden jumps in Formulaic or Predictability: May indicate a shift toward template-driven language

Drops in Sentence Variance: Writing has become more uniform

Large changes in Grade Level: Complexity has shifted significantly

Combined patterns: Multiple metrics shifting together is more significant than one metric in isolation

What the Numbers Do Not Tell You

These metrics measure linguistic features, not authorship. Legitimate reasons for metric shifts include:

- The student researched the topic and adopted domain vocabulary

- They received writing instruction that changed their style

- The assignment type differs from baseline samples

- They are writing about an unfamiliar topic

Clareo provides data for conversation, not verdicts. Use unusual patterns as a starting point for discussing the work with the student.

References

Flesch, R. (1948). A new readability yardstick. Journal of Applied Psychology, 32(3), 221-233.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas for Navy enlisted personnel. Research Branch Report 8-75.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. Behavior Research Methods, 36(2), 193-202.

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type-token ratio (MATTR). Journal of Quantitative Linguistics, 17(2), 94-100.