Prompt 1. Formula generation

Role: You are an expert NLP researcher specializing in text-complexity modeling.
Context: I am designing an empirical study on text complexity. My corpus is mixed-domain, and I have already extracted six feature families:

- **Lexical diversity** (e.g., type–token ratio variants)
- **Lexical density** (content-word ratio, information density)
- **Syntactic complexity** (e.g., mean clause length, subordination index)
- **Text coherence** (discourse-connective density, entity-grid scores)
- **Named-entity load** (NER counts, % of tokens that are NE)
- **Readability metrics** (Flesch, SMOG, etc.)

Task 1 – Feature grouping

- Logically cluster these six families into higher-level dimensions **(max 3 groups)** and justify each cluster in 1-2 sentences.

Task 2 – Formula design

- Propose **10 distinct, mathematically explicit formulas** for a composite Text-Complexity Score (TCS).
- For each formula, list:
  - the normalized feature terms it uses,
  - the weighting scheme (constant weights, learned weights, log-scaling, etc.),
  - a one-line rationale (e.g., "emphasises syntactic difficulty for academic prose").

Output constraints

- Return your answer in **Markdown** with two top-level sections: `## Feature Groups` and `## TCS Formulas`.
- In tables, use the column order: **Formula ID | Expression | Weighting approach | Intended use case**.
- Think step-by-step internally, but **do not reveal chain-of-thought**; present only the final structured answer.
- Keep the entire response under **700 words**.

Parahonco Alexandr