# Language Error Typology

Overview

Our typology for learner errors is hierarchical, with up to three levels below the top-level root category of "language errors." We define this "Level 0" to contrast our typology with those that address issues with other dimensions of writing, which would start from different "Level 0" roots.

The below figure illustrates the hierarchical nature of the typology:
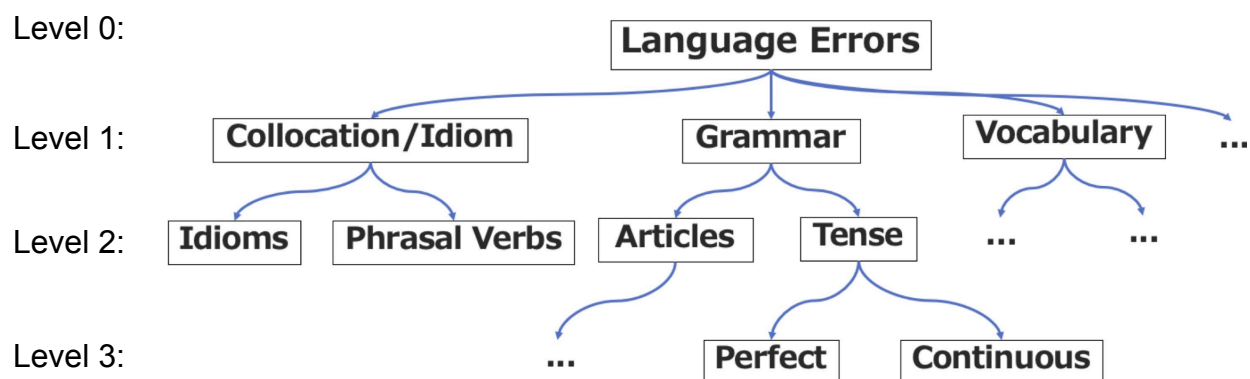
Annotators select error tags in a top-down fashion, starting with one of the six tag "collections" such as "Grammar" and "Vocabulary" (Level 1). These contain individual tags (Level 2), which may then have subtags (Level 3). In total, there are 81 tags in the typology at this time.

Design Principles

The typology is designed to reflect different kinds of "topics" of English learner language errors. Essentially, a tag should help identify the **knowledge gap** that is responsible for the error.

Most tags should be able to be connected with a "grammar point" that can be referred to. In a downstream automated system, there could be a distinct link for each tag, which could be set to a public website or a page in a given electronic textbook.

Note that the typology is not designed to be a perfect model of English grammar. Items in the same grammatical form can be divided across multiple tags (e.g. gerunds are found in both `Verb Nominalization` and `Go + ing`). In other cases, several phenomena are combined into one tag (both infinitives and gerunds are in `Verb Nominalization`; a large variety of "multi-word expressions" are sorted into only a few `Collocation and Idiom` categories).

The granularity of types was defined as a compromise between specificity and ease of use for annotators. We also prioritized alignment with concepts familiar to teachers of English as a second or additional language, such as terms commonly in textbook indices.

Tag names are written with each word capitalized, and are surrounded by backtick (``) characters. **Tag names refer to the attempted or correct structure, not the erroneous version the learner has written.** Tags that do not fit this model, such as `Vocabulary Choice`, are named as if to complete the sentence "This is an error with…"

List of Tags

Text in parentheses are notes/examples, not part of tag names.

Starting at Level 2, tags are alphabetized within their group, except for "Other" tags, which are always placed at the bottom of their respective lists.

Language

- **Vocabulary:**
    - Part-of-Speech Confusion (e.g. "surprised" vs. "surprising")
    - Vocabulary Choice (i.e. Wrong Word)
    - Other Vocabulary Issue

- **Collocation and Idiom:**
    - Fixed Expressions (e.g., "spill the beans," "as a matter of fact")
    - General Collocations (e.g., "do" vs. "make" a mistake; "flock of horses")
    - Phrasal Verbs
    - Verb + Preposition + Argument (e.g., "smile to" vs "smile at")
    - Other Collocation Issue

- **Grammar:**
    - Adverbs of Degree (e.g. "very," "too," "so")
    - Articles/Determiners
        - Articles/Determiners: Articles with Exceptional Words (e.g., "home," "school," "dinner")
        - Articles/Determiners: Definite vs. Indefinite Article
        - Articles/Determiners: Determiner-Noun Agreement

- ■ Articles/Determiners: Indefinite Article Choice
- ■ Articles/Determiners: Missing/Unnecessary Article
- ■ Articles/Determiners: Other Determiner Error
  - ○ Comparison
    - ■ Comparison: Comparative
    - ■ Comparison: Equivalence
    - ■ Comparison: Superlative
  - ○ Causative (e.g., make + person + action)
  - ○ Conditional
  - ○ Conjunction ("and," "but," etc.)
  - ○ Ditransitive Verbs (e.g., "give a gift to him")
  - ○ Dummy Subject (e.g., "it is important that…")
  - ○ Expressions of Place (e.g., "at" vs. "on" the beach)
  - ○ Expressions of Time (e.g., "while," "during," "since")
  - ○ Go + ing
  - ○ Imperative/Let's
  - ○ Infinitive
  - ○ Modal (e.g., "should," "would")
  - ○ Negative Formation
  - ○ Participle (e.g., "bored" vs. "boring")
  - ○ Passive vs. Active Voice
  - ○ Plural vs. Singular
    - ■ Plural vs. Singular: Noun Countability
    - ■ Plural vs. Singular: Noun Number
  - ○ Possessive
  - ○ Prepositions:
    - ■ Prepositions: Cause (e.g., "due to," "because of")
    - ■ Prepositions: Means/Agent
    - ■ Prepositions: Status (e.g., "like," "as")
    - ■ Prepositions: Other
  - ○ Pronoun Antecedent
  - ○ Purpose Clause
  - ○ Quantifier
  - ○ Question Formation
  - ○ Relative Clause
  - ○ Result Clause
  - ○ Subject-Verb Agreement
  - ○ Tense:
    - ■ Tense: Continuous Aspect (e.g., "be... -ing")
    - ■ Tense: Future Formation
    - ■ Tense: Past Formation
    - ■ Tense: Perfect (e.g., "have never")
    - ■ Tense: Tense Choice (present/past/future)
  - ○ That Clause

- ○ Verb Nominalization
- ○ Word Order
- ○ Other Grammar Issue

- ● **Punctuation, Spelling, and Mechanics:**
  - ○ Capitalization (proper nouns, etc.)
  - ○ Colons
  - ○ Commas
  - ○ Contractions
  - ○ Hyphenation
  - ○ Parentheses
  - ○ Quotation
  - ○ Run-on Sentence
  - ○ Semicolon
  - ○ Spacing (e.g., "police woman" → "policewoman")
  - ○ Spelling
  - ○ Terminal Punctuation
  - ○ Other Punct/Mechanics Issue

- ● **Coherence and Cohesion:**
  - ○ Fragment
    - ■ Fragment: Incomplete Thought
    - ■ Fragment: Missing Object
    - ■ Fragment: Missing Subject
    - ■ Fragment: Missing Verb
  - ○ Grammatical Redundancy
  - ○ Incomprehensible
  - ○ Transition (e.g., "However," "Furthermore," etc.)
  - ○ Other Coherence/Cohesion Issue

- ● **Style and Register:**
  - ○ Archaic or Formal Language
  - ○ Casual or Informal Language
  - ○ Potentially Rude/Insensitive
  - ○ Stylistic Redundancy
  - ○ Other Style and Register Issue