

# Semantic Textual Similarity Evaluation Guidelines

## I. Goal

We would like to assess the **degree of meaning correspondence/equivalence** between a pair of sentences. The two sentences could be in the same language or in two different languages.

## II. General Guidelines

### Task Description

1. You will be provided with a pair of sentences and/or audio snippets.
2. The pair could be in the same language or in two different languages.
  - a. Please assess the sentence pairs consistently using the same scoring criteria provided below.
3. Your task is to assess:
  - a. If sentence1 is coherent
  - b. If sentence2 is coherent
  - c. How well the pair of texts correspond to each other on a scale from 1-5
4. When rating semantic similarity, please ignore minor typos, grammatical errors, and pronunciation errors if they do not affect your understanding of the texts/audio segments. Please ignore capitalization and punctuation differences.
  - a. When evaluating texts, you may see some terms or phrases being written and spelled incorrectly, e.g. "Beijing" -> "bayjing" or "genome" -> "jenoum". If you are able to confidently guess the meaning of the term based on the sound, then you can ignore the spelling error. However, if the error is too severe to understand, the pair of sentences should at most get a score of 3 (or lower, depending on the remaining parts of the sentences).

### Rating Guidelines

#### Sentence Incoherence:

- **You must still evaluate STS even if the sentences are incoherent as long as it is possible to compare the sentences by word (examples in the rating categories).**
- Incoherence of each sentence should be assessed **independently**. Do not compare incoherence between the sentences.
- **Please mark a sentence as Incoherent ONLY for the following reasons:**
  - **At least 50% of the sentence** is in a language other than the one indicated in the task
  - The text is not displaying correctly or not displaying at all
  - The sentence is **MOSTLY** incomprehensible, not coherent or intelligible
    - The wording of the sentence is completely distorted

- You are unable to identify any normal syntax in the majority of the sentence to have even a gist of the intended meaning
- If you are able to assess Semantic Textual Similarity between the two sentences, but the sentence meets any of the above criteria for incoherence (e.g., both sentences are incoherent but STS scores 2 or above), please still mark the sentence as “Incoherent” as well as give the sentence pair the appropriate STS score
- **Please do NOT mark a sentence as Incoherent if:**
  - >50% of the sentence is in the indicated language and you are able to grasp the full intent of the sentence
  - You should still NOT mark a sentence as incoherent if:
    - There are simple grammatical errors and typos
    - There are small abnormalities in syntax, phrasing or terminology but you are still able to have a gist of the intended meaning of the sentence
    - Sentence requires additional context for better understanding
    - You should NOT mark “Incoherent” when you do not agree with the content of the post (eg. the post is expressing an opposing political/religious view/value system to your own).

### **Semantic Textual Similarity Scale:**

#### **NOTE: If evaluating text:**

- Please ignore minor typos and grammatical errors if they do not affect your understanding of the texts.
- Please ignore capitalization and punctuation differences.

**[1] The two sentences are not equivalent, share very little details, and may be about different topics. If the two sentences are about similar topics, but less than half of the core concepts mentioned are the same, then 1 is still the appropriate score.**

#### **Example A (different topics):**

Text1: John went horseback riding at dawn with a whole group of friends

Text2: Sunrise at dawn is a magnificent view to take in if you wake up early enough for it

#### **Example B (similar/related topics):**

Text1: The **woman** is **playing** the **violin**.

Text2: The **young lady enjoys listening** to the **guitar**.

#### **Example C (very little overlap):**

Text1: On Friday we all ate a lot.

Text2: They went to a concert on Friday.

#### **Example D (indecipherable on one side or another):**

Text1: lobo lorbo lorl Friday.

Text2: They went to a concert on Friday.

**[2] The two sentences share some details, but are **not equivalent**. Some important information related to the primary subject/verb/object differs or is missing, which alters the intent or meaning of the sentence.**

**Example A (opposite polarity):**

Text1: They **flew out** of the nest in groups.

Text2: They **flew into** the nest together.

**Example B (word order that results in a different meaning)**

Text1: James voted for Biden.

Text2: Biden voted for James.

**Example C (missing salient information):**

Text1: In May 2010, the US troops **attempted to** invade Kabul, **but failed**.

Text2: The US army invaded Kabul in May, 2010.

Text1: "He is not a suspect anymore." John said.

Text2: John said he is **considered a witness** but not a suspect.

**Example D (substitution/change in named entity)**

Text1: I bought the book at **Amazon**.

Text2: The book was purchased at **Barnes and Noble** by me.

**[3] The two sentences are **mostly equivalent**, but some unimportant details can differ. There cannot be any significant conflicts in intent or meaning between the sentences, no matter how long the sentences are.**

**Example A (minor details that are not salient to the meaning):**

Text1: In May 2010, US troops invaded Kabul.

Text2: The US army invaded Kabul on May **7th last year**, 2010.

**Example B (minor verb tense and/or unit of measurement differences):**

Text1: He **bought 2 LBs** of rice at Whole Foods.

Text2: He **buy 1 KG** of rice at WholeFoods.

**Example C (small, non-conflicting differences in meaning):**

Text1: She **loves** eating ripe apples in the fall.

Text2: She **usually** eats ripened apple in autumn.

**Example D (minor difference):**

Text1: Seven years ago and scored **my** history 🕶️

Text2: Seven years ago and recorded history 🕶️

**Example E (omitted non-critical information, but no contradictory info introduced):**

Text1: Several of the sailors set out on a **rainy** Tuesday morning.

Text2: Several of the sailors set out on a Tuesday morning.

**[4] The two sentences are **paraphrases** of each other. Their meanings are **near-equivalent**, with no major differences or information missing. There can only be minor differences in meaning due to differences in expression (e.g., formality level, style, **emphasis**, potential implication, idioms, common metaphors).**

**Examples:**

Text1: This is Europe the so-called human rights country

Text2: This is Europe, the country of alleged human rights

- **Explanation: "So-called" and "alleged" have a different level of formality**

Text1: Special bike for more info call 0925279927

Text2: Special bike for more information call now 0925279927

- **Explanation:**
  - **"Now" implies urgency, "call now" suggests advertising or commercial use**
  - **"Info" less formal than "information"**

Text1: He did a really great job.

Text2: He hit it out of the park.

- **Explanation: "Hit it out of the park" is an idiom, "really great job" is not.**

Text1: Now available right before the amount runs out

Text2: Available now before stock runs out

- **Explanation: Differences in phrasing, text 1 is oddly phrased and more general than text2**

Text1: The **bird** is bathing in the sink.

Text2: **Birdie** is washing itself in the water **basin**.

- **Explanation: Different level of formality ("Birdie" vs "bird").**

**[5] The two sentences are exactly and **completely equivalent** in meaning and usage expression (e.g., formality level, style, **emphasis**, potential implication, idioms, common metaphors)**

In other words, nuance is completely preserved and there is a faithful correspondence. Fidelity is also preserved.

**Examples:**

Text1: A control should be brought to fake ads like this.

Text2: A restriction should be brought to fake advertisements like this.

Text1: What's up y'all?

Text2: Howdy guys!

Text1: Prayer of reference for the first time in the second sermon we must pay attention

Text2: Prayer of reference for the first time in the second sermon, we must pay attention

Text1: One two three apples oranges green Norway

Text2: One two three apples oranges green Norway

***Please note that in this example, both sentences may be marked incoherent as the sentences were incapable of rendering any meaning but can still be evaluated as [5] for STS due to the correspondence/equivalence between the two sentences.***