
Synesis One Data Creation Guidelines for Mind AI Campaigns

September 2022

What We Aim to Do with Synesis One	1
How to Connect Utterances	1
Specific - General	1
Specific	2
General	4
Entailment	6
How to Create Quality Utterances	10
Typicality	10
Pattern Diversity	11
Notice	12
DO	12
DONT	12

What We Aim to Do with Synesis One

Synesis One was established with the aspiration to facilitate the data collection and exchange that will contribute toward building a democratic and ethical Artificial General Intelligence (AGI). To achieve this goal, we need to find an effective way to collect massive amounts of human knowledge that is practical and applicable to educate our artificial intelligence. As the initial step, we decided to focus on collecting quality data through high-tech blockchain-based crowdsourcing to enable immediate use for Conversational Artificial Intelligence (CAI) development. Your contribution by submitting diverse utterances will enable us to create a concrete and human-serving CAI for the greater human good.

How to Connect Utterances

Specific - General

Specific

A specific utterance should refer to the theme described in the topic subject example, but with **more details**. Specific utterances will include more details of an action, a situation, and a specific intent making them longer.

Consider using semantic criteria first. If the meaning of the utterance is specific, it can be considered as specific.

- Semantic criteria: to make the creation process easier, you can start by making sentences that detail the original **meaning** of the topic subject. If the meaning of the utterance is more specific, it will be categorized as specific.
- Syntactic criteria: you can choose to create a syntactically different sentence **structure** from the topic subject. If the created utterance has equal or more content words¹ than the topic subject with a syntactic difference, that utterance will be categorized as specific.

Domain: Electronics	
Topic subject 1	Specific utterance
Are there installation services available?	✓ Do you offer installation services for <u>electronics</u> ?
	✓ I want to get an <u>on-site</u> installation service
	✓ Do you provide an installation solution?
	✓ Is there <u>electronics installation assistance</u> provided?
	✓ How can I get someone to install <u>electronics</u> from your shop?

* Underlined description: these additional words provide more specific information about the core theme (installation services available), thus making the utterances more specific.

¹ Content words, in linguistics, are words that possess semantic content and contribute to the meaning of the sentence in which they occur. In a traditional approach, nouns were said to name objects and other entities, lexical verbs to indicate actions, adjectives to refer to attributes of entities, and adverbs to attributes of actions.

One crucial thing to keep in mind is that we **DO NOT** accept duplicated sentence patterns. Take a look at these examples for **topic subject 1**:

✗ Do you offer installation services for TVs/for homewares/for microwaves?

✗ Do you offer installation services at home/at condo/at the office?

These sentences do not show a variety of sentence patterns. We encourage builders to try to submit as various sentence patterns as they can come up with. If a builder has already submitted at least one sentence pattern, other builders should try to make different sentences (both syntactically and semantically) to get validated. Also, builders must not use random words that do not specify the **core meaning** of the topic subject.

Domain: Grocery	
Topic subject 2	Specific utterance
Is there a parking lot at the branch?	✓ Does the <u>supermarket</u> provide a parking lot?
	✓ Do you offer <u>free parking in front of the store?</u>
	✓ Is there a parking area <u>for customers</u> at the branch?
	✓ Is there a <u>lady</u> parking lot at the <u>store?</u>
	✓ Does the <u>store</u> offer a parking lot <u>with a roof?</u>

The example of **topic subject 2** that will be rejected by validators:

✗ Is there a lady parking lot?

This sentence may look like a specific sentence at first glance, but the content does not cover the meaning of the topic subject ("at the branch" is missing). This sentence becomes more general to be categorized as specific.

Domain: Grocery	
Topic subject 3	Specific utterance
I would like to pay in cash on delivery	✓ Can I pay with cash when the delivery is complete?
	✓ I want to pay with <u>cash and coupon</u> on delivery
	✓ How do I pay with cash on delivery?

	✓ I want to pay when I received my order
	✓ Can I pay cash when I receive the product?

The examples of **topic subject 3** that will be rejected by validators:

- ✗ I would like to pay with cash when my food/drinks/milk/snack has arrived.
- ✗ Does the shipper accept cash?

The second sentence is considered to be an entailment of the topic subject.

General

A general utterance should refer to the theme described in the topic subject example, but with **fewer details**. General utterances will include less details of action, situation, and specific intent making them shorter.

Consider using semantic criteria first. If the meaning of the utterance is general, it can be considered as general.

- Semantic criteria: to make the creation process easier, you can start by making sentences that detail the **original meaning** of the topic subject. If the meaning of the utterance is more general/simpler than the topic subject, that utterance will be categorized as general.
- Syntactic criteria: you can choose to create a syntactically different sentence **structure** from the topic subject. If the created utterance has less content words than the topic subject, that utterance will be categorized as general.

Domain: Telecommunication	
Topic subject 1	General utterance
I want to buy <u>a 3-month</u> mobile package plan	✓ Mobile package plan purchase
	✓ How to buy a mobile plan
	✓ Where can I get a mobile plan?
	✓ Mobile plan purchase recommendation
	✓ I want to buy a mobile package plan

The examples of **topic subject 1** that will be rejected by validators:

- ✗ Can I buy a package plan?
- ✗ I am interested in buying a package plan

These sentences are too broad and surpass the given domain (telecommunication). It does not describe “mobile package plan” in a more general sense, but invites different interpretations in different domains (such as travel package plan).

Domain: Grocery	
Topic subject 2	General utterance
Is there a parking lot <u>at the branch</u> ?	✓ Where can I park my car?
	✓ Is there a parking lot?
	✓ Is the parking available?
	✓ Do you have a space for parking?
	✓ Do you have a parking lot?

The examples of **topic subject 2** that will be rejected by validators:

- ✗ Can I drive to the store?
- ✗ I will go to the branch by car

These sentences do not clearly describe the speaker’s intent to know whether a parking lot is available at the branch. They can be validated as entailments, but not general sentences of **topic subject 2**.

Domain: Grocery	
Topic subject 3	General utterance
I would like to pay in <u>cash on delivery</u>	✓ How to pay in cash
	✓ Can I pay upon delivery?
	✓ Do you accept cash on delivery?
	✓ I want to pay in cash on delivery

	✓ I would like to make a <u>payment</u> on delivery
--	---

The example of **topic subject 3** that will be rejected by validators:

✗ I want to make a payment

✗ Cash on delivery

The first sentence is too broad. It does not describe “cash” or “delivery” therefore changes the speaker’s intent. The second example is too broad and ambiguous. It can refer to “I want to know about cash payment on delivery”, “I have a problem with cash payment on delivery” or anything about cash on delivery.

Entailment

NOTE: Entailment replaces “Cause-Effect”. We **will not open the “Effect” slot anymore**. Until the Synesis UI modification is done, **please regard “Cause” as “Entailment”** though there are some clear differences. Please read this Entailment section carefully.

Entailment describes a relationship between two sentences such that if the first is true, the second must also be true, as in I drive my son to school every day and I know how to drive. To create entailment utterances, builders should take into account that a given piece of **Topic Subject** entails the meaning of **Entailment Utterances**. For example, (1) entails the meaning of both (2) and (3). **Topic Subject** entails **Entailment Utterance** when a person reading **Topic Subject** would affirm that **Entailment Utterance** is most likely true. **Entailment Utterance** describes a part of information from a given **Topic Subject**.

(1) **Topic Subject:** Can you give me some water?

(2) **Entailment Utterance 1:** I am thirsty.

(3) **Entailment Utterance 2:** I need water.

Builders should create utterances that **can replace** the **Topic Subject** and keep the original intent intact. The following is the standards in creating utterances for entailment relation:

- Utterances should **fulfill the intent** of the Topic Subject, meaning they can replace the Topic Subject and still convey the same meaning

- Given that the Topic Subject is true, utterances must **also be true**².
- Entailment utterances should describe a **part of information** from the Topic Subject

These are examples which fulfill all of the standards for our entailment ontology creation:

Domain: Electronics	
Topic subject 1	Entailment utterance
I need help with my laptop	✓ I cannot use my laptop properly
	✓ Why is there no help provided for my laptop?
	✓ I have a problem using my laptop
	✓ There are some points I don't understand while using my laptop
	✓ I need more information about how to use my laptop

The examples of **topic subject 1** that will be rejected by validators:

- ✗ People can misunderstand how to use a laptop

This sentence does not properly fulfill the topic subject's intent though it is entailed by the topic subject.

Domain: Synesis One	
Topic subject 2	Entailment utterance
I want to read the technical document of your company	✓ I need more information about your company's technology
	✓ I don't know how to get your technical paper
	✓ I am interested in your company's technology
	✓ I am curious about your company's technology
	✓ I want to study the technology of your company

The example of **topic subject 2** that will be rejected by validators:

² This is for a statement. Request utterance topic subjects will follow the first and third conditions.

- ✗ Your company has put forward a technical white paper in the past
- ✗ I'm interested in technology

These sentences do not properly fulfill the topic subject's intent though it is entailed by the topic subject.

Domain: Home Mortgage	
Topic subject 3	Entailment utterance
How can I apply for a home mortgage?	✓ I don't know if I qualify for an apartment loan
	✓ I want to prepare for a home mortgage application
	✓ I need more information on home mortgage
	✓ I want to know about a home mortgage application's instructions
	✓ I want to apply for a home loan

The example of **topic subject 3** that will be rejected by validators:

- ✗ People can apply for a housing loan
- ✗ I don't have enough money to buy a house

These sentences do not properly fulfill the topic subject's intent though it is entailed by the topic subject.

Domain: Card Payment	
Topic subject 4	Entailment utterance
How can I activate my card?	✓ I want to start paying by card
	✓ I don't know how to use my card
	✓ The bank should notify me on how to use my card
	✓ This is my first time using the card
	✓ I don't know how to activate my card

The example of **topic subject 4** that will be rejected by validators:

✘ I have a card

This sentence does not properly fulfill the topic subject's intent though it is entailed by the topic subject.

✘ Can I use my card to withdraw cash?

This cannot be inferred by **topic subject 4**.

In the original sense of the word 'entailment', the Topic Subject "Can you give me some water?" can have the following entailments(and more): "Human-beings cannot live without water", "Water is essential in maintaining biological activities", and so forth.

However, this does not fulfill the condition Utterances should fulfill the intent of the Topic Subject, so we cannot use them for Conversational Artificial Intelligence (CAI) development. Remember that the goal is to collect as many entailment relations as possible that can fulfill the usefulness of CAI development. You can see this further in the following examples:

Domain: Electronics & Customer Service	
Topic subject	Entailment utterance
Are there installation services available?	✘ Some stores offer installation services
	✘ Customer wants installation services
	✘ Installation services should be applied for
I want to subscribe to Synesis One Newsletter	✘ Synesis One offers Newsletter
	✘ Synesis One periodically updates in-house news
	✘ Anyone can request to subscribe to the Synesis One newsletter

NOTE: Some of the entailment utterances can be confused with the **specific** category utterances because syntactic/semantic specificity sometimes contains a part of information of a topic subject. Please note that **an utterance can have multiple relations at the same time**. If it fits for all the criteria in more than one category, validators will validate them all accordingly.

How to Create Quality Utterances

Good quality utterances may have two characteristics: Typicality and Pattern Diversity.

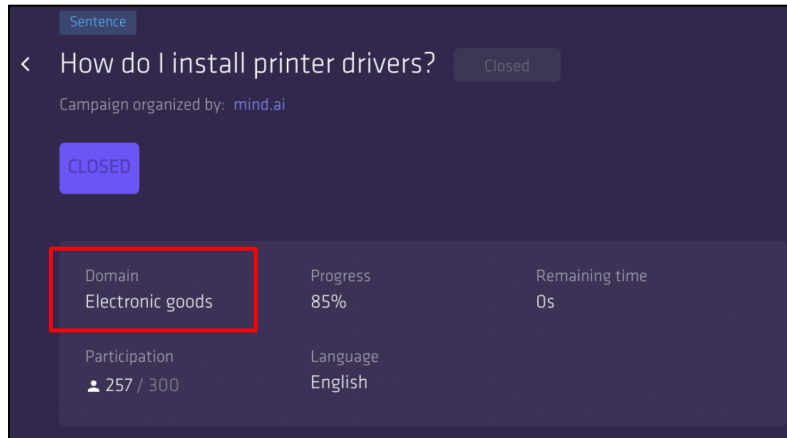
Typicality

Typicality is fulfilled when certain utterances **cannot be interpreted other than within the context of the topic subject example**. In other words, when builders create utterances they should consider whether their submissions are **relevant** and sound **natural**. An utterance is considered relevant when it clearly relates to the topic subject. Also, it is considered natural when a native speaker does not feel any awkwardness in accepting it.

Note that what is 'natural' to each language user is different. While social and regional varieties are equally valid, we focus on collecting utterances complying with the most common linguistic standards.

Domain: Food and Beverage	
Topic subject 1	Specific Utterances
Pizza order please	✓ Can I order pizza? (natural and relevant)
	✗ I need pineapples (not relevant)
	✗ I have a desire to eat pizza (unnatural)

Please note that all topic subjects are assigned to specific domains. Thus, whether an utterance is relevant and natural will be judged within the context of each related domain. Builders should create utterances taking into account the specificities of the domains and natural language expressions.



[Figure 1] Domain information section of Builder UI

Domain: Electronic goods	
Topic subject 2	General utterance
How do I install printer drivers?	✓ Instructions for printer driver installation
	✗ How to reset printer (not relevant)
	✗ Give printer driver installation instructions (unnatural)

Pattern Diversity

We say utterances have **Pattern Diversity** when they have varied syntactic and semantic patterns. Pattern Diversity is an important standard of good quality data since it means the utterances cover many different ways of expressing the same intent. Good data have pattern diversity without duplicates. That is, utterances with pattern diversity **DO NOT** include changes in one word or various types of specific entities. The example of bad pattern diversity shows no diversity in the sentence structure, but only swapping of words.

Examples of **BAD** data

Topic subject	Utterances
I want to buy clothes	✗ I want to buy (men's, girls', children's) shirts
	✗ Where can I find (men's, girls', children's) shirts

	<p>✗ I want some (men's, girls', children's) shirts for my (dad, mom, kid, girlfriend, boyfriend, wife, grandma, grandpa, grandmother...)</p>
--	---

Examples of **GOOD** data

Topic subject	Utterances
I want to buy clothes	✓ I want to get clothes
	✓ Where can I purchase clothes?
	✓ Can you suggest some good clothes?

Notice

DO

- ✓ Create natural utterances that people would type when using a chat application.
- ✓ Create utterances with good pattern diversity
- ✓ Create utterances according to General-Specific and Entailment standards

DONT

- ✗ Making utterances with use of filler words
e.g. Uh!, um, well..., that is..., you know..., I mean like...
- ✗ Making utterances with changes of brands, numbers, and types.
e.g. I need a SAMSUNG laptop, I need an ACER laptop, I need a LENOVO laptop
- ✗ Making utterances with poor pattern diversity by using an automated data generator.
e.g. I want this book, I want that book, I want these books, I want those books
- ✗ Making utterances that's a combination of two intents.
e.g. I want to return this product and get a refund
Please tell me what to do and where to go
I'm looking for canned pet food for my pet and dairy products for myself