Note: This document contains the annotation guidelines used in the GUMBridge (v0.1) annotation pilot. During the adjudication of the annotations resulting from this pilot, the subtype categories of **comparison-sense-hierarchical** and **comparison-sense-non-hierarchical** were collapsed into a single **comparison-sense** category, and the category **comparison-ellipsis** was renamed to **comparison-relative**. The guidelines detailed here are exactly as they were provided to annotators during the pilot, and as such those changes are not reflected in this document.

GUMBridge: Bridging Taxonomy and Annotation Guidelines

Introduction to Bridging Anaphora:

Bridging is an anaphoric phenomena which occurs when **the reference of one entity is dependent on a previous associated, non-identical entity for interpretation**. An instance of bridging consists of an antecedent-anaphor pair, where understanding the anaphor first requires the understanding of the associated, non-identical antecedent. The anaphor in this pair is referred to as the bridging anaphor. Consider the following example:

> There is <u>a big yellow house</u> on the hill. **The door** is red.

In this example, the anaphor **The door** is understandable as a specific door (the door of the house) only when the associated antecedent, <u>a big yellow house</u>, is first interpreted. This is an instance of bridging via the part-whole lexical relation (meronomy) between the words "house" and "door". Beyond the part-whole relation, there are a variety of ways that bridging can be realized in a discourse. The different varieties we are aiming to capture in this project are outlined in the [taxonomy overview](#) and detailed in the [bridging subtypes](#) section below.

In this project, we will be using an information status informed view to identify instances of bridging. The information status (IS) of an entity refers to the extent to which an entity is accessible to the reader/hearer of a discourse. Broadly speaking, an entity is either New, Given, or Accessible. A New entity has not been mentioned before and is unfamiliar to the reader/hearer, while a Given entity has been mentioned before and is thus familiar to the reader/hearer. An **Accessible** entity has **not been mentioned before but is inferable for a reader/header in the discourse**.

A bridging anaphor is accessible to the reader/hearer from its first mention in the text, specifically because of its relationship with its associative antecedent. In other words, for the purposes of this project, **bridging occurs when an entity is Accessible via an inference from a previous non-identical entity in the discourse**.

In English, newly introduced definite entities are a common signal of an instance of bridging, but bridging anaphora are not limited to definite entities. Additionally, newly introduced definite entities with a descriptive adjunct/relative clause or prenominal modifier are common and typically not instances of bridging (e.g., "the psychology exam" or "the man who I saw" may be introduced as New entities, despite being definite).

There are also some common scenarios to point out which we **do not consider to be bridging**:
- **Coreference**: If an entity has a previous mention, it cannot be an instance of bridging, even if it has an associative relationship with a previous non-identical entity.
- **Bridging contained**: If the associative antecedent is a direct modifier in the NP of the potential bridging anaphor, e.g. "**the focus** of the story" or "**two** of them", it should not be annotated as bridging.
- **Generics/Situational bridging**: Entities that are accessible due to general world knowledge or situational context are not considered instances of bridging, i.e., if it doesn't have an associated antecedent to be bridging from, it cannot be bridging.
- **Possession with an explicit possessive**: If the potential bridging anaphor contains an explicit possessive which corefers with the associative antecedent, no bridging relation is necessary. Explicit coreference between the associative antecedent and the possessive is sufficient (e.g., [Mark]…[his] house → no bridging, coreference between "Mark" and "his").

Identifying Bridging Anaphora:

When considering whether a particular entity is a bridging anaphor:
1. Do you judge this entity to be to some degree **accessible** in the discourse?
2. Does that accessibility rely on the **understanding of a previous entity** in the discourse? If so, identify that previous entity's most recent mention.
3. Check that your entity does not fall into the "not considered to be bridging" categories described in the section above.

If your entity passes all 3 of the above criteria, it is a bridging anaphor and the previous entity is its associative antecedent.

Taxonomy Overview:

Bridging Subtypes

- comparison
  - comparison-ellipsis
  - comparison-sense
    - hierarchical
    - non-hierarchical

- ○ comparison-time

- ● entity
  - ○ entity-associative
  - ○ entity-meronomy
  - ○ entity-property
  - ○ entity-resultative

- ● set
  - ○ set-member
  - ○ set-subset
  - ○ set-span-interval

- ● other

Bridging Subtypes:

comparison

The anaphor is interpretable via a comparison to the antecedent (or vice versa).

comparison-ellipsis

The anaphor is preceded by a comparative marker (other, another, same, more, ordinal modifiers, etc.) which implies a comparison to the antecedent (or vice versa).
For example:

Several women walked into the room. **Other women** soon followed.

This creates an "ellipsis of the compared" which can generally be explicitly constructed by adding a "than" or "as" phrase containing the antecedent to the anaphor (if this phrase is already present, it is a case of bridging-contained, and it does not need to be annotated as bridging). In the example above, "than those women that previously entered the room" could be added after "Other women". Notably, the lexical head is present in the anaphor ("women", in the above example).

comparison-sense

The type of the anaphor is omitted but inferable via comparison to the antecedent (or vice versa).

hierarchical

The anaphor additionally has a set-relation/portion-substance relation with the antecedent.
For example:

Several women walked into the room. **One** *left immediately.*

non-hierarchical

The anaphor and antecedent do not have a hierarchical relationship.
For example:

I've been to the Chinese restaurant. I want to go to **the Italian one***.*

The nominal version of comparative markers (others, another (by itself, without a lexical head)) are included here, rather than under comparison-ellipsis.
For example:

I went to an Italian restaurant for the first time. I want to go to **others***.*

comparison-time

The anaphor refers to a specific time/timeframe which is understandable with reference to the time/time frame expressed by the antecedent (or vice versa).
For example:
- today → next week
- this year → 10 years later
- May 22nd → Wednesday (if interpretable in the discourse as some particular Wednesday relative to the date)

entity

The anaphor is an attribute or associated entity of the antecedent (or vice versa). This associated attribute/entity may be necessary, probably, or inducible from the specific context of the discourse.

entity-associative

The anaphor is an attribute or closely associated entity of the antecedent (or vice versa).
This includes:
- Characteristic features of an entity class: e.g., wedding → venue, ceremony, etc.
- Relational nouns (whose lexical meaning expresses how it relates to another entity)
    - Frequently conveying a social role, including kinship and organizational roles: murder → victim, company → ceo, business → customer, etc.
- Implicit roles of events (participants, instruments, etc.)
- Other non-social associative relations: book → publication, teaching → class

entity-meronomy

The anaphor is a subunit of the antecedent (or vice versa) (there is some part-whole relation between the anaphor and the antecedent). This includes:

- Physical subparts
- Substance-portion relations
- Regions/subsections (the left, the edge, etc.).

entity-property

The anaphor a physical or intangible property of the antecedent (or vice versa).
For example: smell, length, style, etc.

entity-resultative

The anaphor is logically inferable from the antecedent (or vice versa). Including the anaphor as:
- The entailed result of the antecedent event (e.g., falling tomato → splatter)
- A transformed/transmuted version of the antecedent entity (e.g., batter → cake)
- The reason/cause of the antecedent

set

There is a set relation between the bridging anaphor and antecedent.

set-member

The anaphor is an element of the antecedent set (or vice versa). This includes:
- Lists and examples
- Groups and members
- Indefinite member → definite plural set
- Plural set → particular member of the set
- Class → instance

set-subset

The anaphor is a subset of the antecedent set (or vice versa).

set-span-interval

The anaphor is a sub-span of the spatial or temporal antecedent interval (or vice versa).
For example:
- Several days in a week/month
- A subsection of a day (e.g., morning, night)
- Intervals of distance

other

The anaphor and antecedent fit the criteria for identifying a bridging pair, but do not fall into any of the bridging subtypes detailed above.
This may include near identity relations which are not annotated as coreference, such as:
- Name Metonymy: a single name referring to multiple facets of a complex entity (e.g., USA first referring to the country as an organization then to the landmass as a place).

- Spatio-temporal dissimilation: when "the same" entity is split on spatial or temporal characteristics (e.g., "Europe, pre-war Europe")