Ambiguity & Incompleteness:

According to Gause and Weinberg, ambiguity has two sources, **missing information** and **communication errors** [1]. In this research, we interpret missing information as **INCOMPLETNESS**. Missing information has various reasons. For instance, humans make errors in observation and recall, tend to leave out self-evident and other facts, and generalize incorrectly. Communication errors that occur between the author and the reader are due to expression inadequacies in the writing.

The view of Gause and Weinberg is shared by Harwell, Aslaksen, Hooks, Mengot, and Ptack [2] in their definition of an unambiguous requirement. "A requirement must be unambiguous in the sense that different users (with similar backgrounds) would give the same interpretation to the requirement. This has two aspects. On the one hand there is the aspect of grammatical ambiguousness, i.e. the poorly constructed sentence. On the other hand there is the aspect of ambiguousness arising from a lack of detail allowing different interpretations. The first of these can be measured or tested independently of author or user, but the second aspect can be measured only in conjunction with a set of users, since it depends on what assumptions the user makes automatically, i.e. as a result of the user's background knowledge."

General Instructions for annotation:

- Does the word/phrase/sentence create multiple plausible meanings? If the answer is yes, then we have ambiguity – if the answer is no there is no ambiguity.
- For every ambiguity, make sure to write different interpretations that can be made. For example, you can show different grouping of the words in syntactic or semantic ambiguity, or wherever possible.
- Make sure to include keywords from the definitions of the specific ambiguity in the reasoning.
- During annotation be precise as possible. We would rather have smaller sections to be annotated instead of a full sentence. For example, for lexical ambiguity, as little as one word may need to be annotated.
- For the question – Try to avoid Y/N questions. Always ask users what they want? Ask W/H a question.
- What to do if the same ambiguous word/phrase appears in different locations in the text? First, read the whole feature request and then start annotating. This will reduce the chance of annotating a piece of text that is clarified later. For example, an ambiguous word/token in the title might not be ambiguous when you read the

description. After this, if multiple annotations exist for a word/phrase, keep those. Keep all annotations - don't avoid annotating something thinking it was annotated before.

- Text copied or transcribed from another author, such as from another bug report or feature request, will be annotated normally as though the current author had written it. Identifying and excluding these excerpts from analysis at scale is deemed to be prohibitive with our current methods and is recognized in our threats to validity.
- If you suspect formatting has been lost during the scraping process, such as items in an ordered list being displayed without punctuation dividing them, refer to the original GitHub entry before considering if the text is ambiguous. We recognize a LLM may not be able to identify

Ambiguity Categories

Our categories are based on the categories outlined in the ambiguity handbook created by Berry, et al. [3]. Wherever possible, language identified as ambiguous should attempt to be matched to one of these existing categories.

1. Lexical ambiguity
   Lexical ambiguity occurs when ***a word/phrase*** has several meanings. Lexical ambiguity can be subdivided into homonymy and polysemy.

    a. *Homonymy* occurs when two different words/phrases have the same written and phonetic representation, but unrelated meanings and different etymologies, i.e., different histories of development. For example, the verb lead (to guide) and the noun lead (the metal/element) have unrelated origins.
    b. *Polysemy* occurs when a word/phrase has several related meanings but one etymology. For example, the verb "get" can mean procure, obtain, become, or understand. Other related concepts like systemic polysemy, categorical ambiguity, possessive ambiguity, and metaphor are to be annotated under the polysemy sub-class.
    c. Metaphor is only ambiguous if the intended meaning of the metaphor is itself ambiguous, or when both the literal and figurative usage of the word/phrase could fit in the given context.
    d. Technical jargon will be annotated under polysemy and homonymy as appropriate. Jargon will typically be annotated under Polysemy because the jargon's origin is often rooted in metaphor related to the term's literal

meaning (e.g. “Tweet” for a post using the Twitter/X app, which had a bird-shaped logo when the term was created). Note that jargon should only be annotated as ambiguous if the annotator is uncertain of the meaning or is unsure if the highlighted term is being used in its literal or jargon form.

e. Acronyms will be annotated under polysemy and homonymy as appropriate. Acronyms will typically be annotated under homonymy when the potential meanings are unrelated, or polysemy when they are related. For example:
    i. Polysemy occurs when the acronym for “Multimedia Messaging Service (MMS)” is often seen used to refer to MMS capability, the MMS protocol, an individual message sent using MMS, or just the non-text data attachment portion of a message sent using MMS.
    ii. Homonymy occurs when the acronym “POC” could be interpreted to mean point of contact, proof of concept, person of color, port of call, or point of care.
f. Abbreviations will be annotated under polysemy and homonymy as appropriate. For example:
    i. Polysemy occurs when the abbreviation “admin” could be interpreted to mean administration, administrator, or administrative.
    ii. Homonymy occurs when the abbreviation “temp” could be interpreted as temperature or temporary.
g. The annotator should highlight only the ambiguous word or phrase itself and note different plausible meanings of the annotated word/phrase in the reasoning section of the tool.

2. Syntactic ambiguity

Syntactic ambiguity, also referred to as *structural ambiguity*, occurs when the structure of a sentence allows for multiple different interpretations or parses.

a. *Analytical ambiguity* occurs when the role of the constituents within a phrase is ambiguous. An example is a noun phrase, where the modifier can be assigned to two nouns, e.g., “The Tibetan history teacher” is a noun phrase, where Tibetan can modify history or teacher.
    i. Similar patterns can be seen in prenominal double possessives when it is unclear if the middle possessive is being used to indicate ownership or style (e.g. “Alice’s doctor’s appointment”)
    ii. The annotator should highlight and annotate the whole noun phrase and mention different interpretations.

iii. By its nature, analytical ambiguity may imply there is lexical ambiguity of one of the phrase's constituents, however in many cases resolving the structure would also resolve this lexical ambiguity. As such, lexical ambiguity should not be annotated within an analytically ambiguous phrase unless the lexical ambiguity would continue to be present within the potential parses of the phrase.

b. *Attachment ambiguity* occurs when a particular syntactic constituent of a sentence, such as a prepositional phrase or a relative clause, can be legally attached to two parts of a sentence
   i. Often involves a prepositional phrase that may modify either a verb or a noun
   ii. Example: "The police shot the rioters with guns." "With guns" is a prepositional phrase in this sentence with ambiguous attachment. There are two interpretations: (1) rioters with guns; (2) police used guns to shoot rioters.
   iii. Annotate the whole prepositional phrase whose attachment is ambiguous. List which words in the sentence the preposition can potentially target.

c. *Coordination ambiguit*y occurs when the presence of one or more conjunctions creates multiple possible parses of the sentence depending on how constituents are grouped together.
   i. Can occur when more than one conjunction, typically 'and' or 'or', is used in a sentence. For example, "I saw Mary and Paul and Peter saw me."
   ii. Can occur when one conjunction is used with a modifier. For example, "secure messages and emails" could be interpreted as "secure messages and all emails" or "secure messages and secure emails."
   iii. Can occur when one conjunction is used with multiple modifiers. For example, "flagged and unread emails" could be interpreted as "flagged emails and unread emails" or "emails that are both flagged and unread."
   iv. The annotator should highlight the entire text segment linked by the ambiguous conjunctions and list the possible parses or interpretations.
   v. Coordination ambiguity can fall under syntactic and semantic. **BUT we agree that we recognize this ambiguity as syntactic in our**

annotations.

d. Elliptical ambiguity occurs when it is unclear whether words have been omitted from a sentence, typically for brevity, or when the potential content of the omission itself is ambiguous.
    i. For example, in the question “Have you finished your book?” it is unclear if the omitted verb is “writing” or “reading”.
    ii. Annotators should highlight the whole sentence to provide the complete context. Attempting to isolate and only highlight the location of the ellipsis has been observed to be too subjective and negatively affects the agreement statistics by introducing an unintended variable. Annotating the entire sentence reflects that the addition of the omitted content is required somewhere to make the text ambiguity-free.

3. Semantic ambiguity

Semantic ambiguity covers incidents where the logical form of a sentence is ambiguous in spite of an absence of lexical or syntactic ambiguity.

a. *Referential ambiguity* occurs when anaphora (typically pronouns or ellipses) can reference more than one antecedent in the sentence. For example, in the text “I tried to send a request to get my account status, but the app failed to deliver it”, the pronoun “it” could refer to either the request or the status.
    i. If the ambiguity is **contained within ONE sentence**, we recognize it as semantic referential ambiguity. However, if the ambiguity is related to sentences appearing before and after, we consider it to be pragmatic referential ambiguity.
    ii. Annotators should only highlight the relevant anaphor and list the possible antecedents.

b. *Scope ambiguity* occurs when the relationship between quantifiers or negations and other sentence constituents is unclear.
    i. If a sentence contains two or more quantifiers, their scope can cause ambiguity. Quantifiers include every, each, all, some, several, a, one. Annotate all the quantifiers in the sentence that cause ambiguity.
    ii. A sentence that includes two or more quantifiers, AND it also contains operators such as NOT, AND, OR, etc. may contain scope ambiguity. Annotate all the quantifiers in the sentence that cause ambiguity.

iii. We include Collective/Distributive ambiguity, where the relationship between plurals is unclear, as Scope ambiguity. For example, in the text "add notes to other profiles", when translating this phrase to predicate logic, it is unclear whether we can add ONE note to all Profiles, MANY notes to ONE profile, etc. The existence of plural nouns causes multiple interpretations regarding the **quantifiers**. In this case, annotate the plural nouns. Check feature request#5 for an example. Annotators should highlight the text to include the plurals in question and give examples of the different interactions possible.

4. Pragmatic ambiguity

   Pragmatic ambiguity occurs where there is ambiguity involving relationships between a text and its context.

   a. *Referential ambiguity* occurs when anaphora (typically pronouns or ellipses) can reference more than one antecedent in the sentence. For example, in the text "Adam told me what Bob said to Charlie. I think he is being unreasonable.", the pronoun "he" could refer to Adam, Bob, Charlie, or another person appearing even earlier in the context.
      i. If the ambiguity is contained within ONE sentence, we recognize it as semantic referential ambiguity. However, if the ambiguity is related to sentences appearing before and after, we consider referential pragmatic ambiguity. Example, feature request #5: Title: Feature: Display profile notes Description: In the Web UI one can add notes to other profiles. It should be possible to view and edit them in the app too. "Them" has pragmatic referential ambiguity.

   b. *Deictic ambiguity* occurs when information about the context itself is required to disambiguate the meaning.
      i. Pronouns can be deictic if it is unclear whether the pronoun is referring to an anaphor or introducing a new subject.
      ii. Demonstratives (this, that) and indexicals, particularly those related to relative time and relative location (here, now, today, near, etc.), often cause deictic ambiguity. For example, in the text "Alice needs the report here by tomorrow", the location of the author and the time of writing are needed to resolve where "here" refers to and when "tomorrow" is.

iii. If the user refers to the location on the UI of the app, there might be deictic ambiguity. However, there needs to be multiple interpretations for the location. In the reasoning, make sure to mention GUI element and location as keywords.
iv. Expressions such as “it should be possible” or “it would be useful” do not contain deictic ambiguity for “it”.

c. *Language error ambiguity* occurs when mistakes in grammar, spelling, punctuation, diction, or other language usage issues cause ambiguity when the reader attempts to interpret the intended meaning but must consider multiple candidate meanings.
    i. Note that every mistake does not constitute ambiguity. If there is only one probable correction for a mistake that fits in the context, there is no ambiguity because there are not multiple competing interpretations. For example, the typo “piggybag” could be corrected to “piggyback” or “piggybank”, but if only one of the words fits in the context, there is no ambiguity present.
    ii. Language error can still occur when the text appears to be technically correct. For example, the text “local and federated timeline” could be interpreted literally to indicate a single timeline combining local and federated data, however a reader might suspect the intended meaning was “local and federated timelines”, which creates ambiguity between these competing interpretations.
    iii. If a section of text is ambiguous due to language error, do not annotate the individual ambiguities within the possible corrections of the text. Only annotate the language error itself.

1. Vagueness

Vagueness occurs when a quality is subjective or cannot be empirically measured or defined. For example, how tall is “tall”? How big is “big”? How old is “old”? How fast is “fast”?

d. If the requester is providing an example for the vague term (for example in feature request 817), we don’t annotate the term as vague. Shorter and easier encoding, e.g., base58-enconding, poetry encoding. In this case the requester has provided explicit examples of what qualifies as shorter or easier in their view.
e. If there is a vague term in the feature request, but it is not related to the quality of the software or the way user accomplishes a task through

software, we don't annotate the term. This has nothing to do with the defect affecting the downstream task.

Incompleteness

(Requirements engineering fundamentals: A study guide for the certified professional for Requirements Engineering Exam foundation level – Klaus Pohl, Chris Rupp)

Incompleteness occurs when key information is missing from the request that would be required to implement the feature in accordance with the author's expectations. Only a single annotation needs to be made to enumerate any missing elements of information. Examples of causes of incompleteness can be seen as follows:

1. The user can explain an issue about the app, but no clear request has been made. In this case, the feature request is incomplete and we need to ask about what feature they are requesting and how it should be implemented.

2. Nominalization: A process is converted into a singular event. All information necessary to accurately describe the process is thereby lost.

   Example: "In case of a system crash, a restart of the system shall be performed." The terms "system crash" and "restart" each describe a process that ought to be analyzed more precisely.

   Example: The process word "transmit" turns into the noun " transmission".

   Example: "Send them the finger prints securely" => send? How, define step by step …

3. Nouns without reference index:

   "The data shall be displayed to the user on the terminal"

   What data? Which user? Which terminal?

4. Incompletely specified conditions: (Trigger word: If …then, in case, whether, and depending on)

Example: “The restaurant shall offer all beverages to a registered guest over the age of 20 years.” What should be offered to the guests that are 20 years or younger?

5. Passive voice:

   Example: To log a user in, the login data is entered” => who enters the login data?

References

[1] D. C. Gause and G. M. Weinberg, *Exploring Requirements*. Dorset House Publishing Company, Incorporated, 1989.

https://www.dorsethouse.com/books/er.html

[2] R. Harwell, E. Aslaksen, R. Mengot, I. Hooks, and K. Ptack, ‘What Is A Requirement?’, *INCOSE International Symposium*, vol. 3, no. 1, pp. 17–24, 1993.

https://incose.onlinelibrary.wiley.com/doi/abs/10.1002/j.2334-5837.1993.tb01553.x

[3] E. Kamsties, P. Science, M. Krieger, P. Mathematics, and M. Berry, ‘From Contract Drafting to Software Specification: Linguistic Sources of Ambiguity’, vol. parallelism, 12 2003.

https://cs.uwaterloo.ca/~dberry/handbook/ambiguityHandbook.pdf