General Annotation Guidelines

- Do not annotate whitespace outside of the entity (on either side of the beginning/end).
- Don't include punctuation unless obviously part of the entity.
- Do not annotate entities inside of other entities.
- Be greedy; always take the biggest possible entity you can.
  - In "Starfleet Academy", tag "Starfleet Academy" and not "Starfleet".
- Annotate titles along with names whenever they occur (E.g. in the sentence "Captain Picard walked onto the bridge", annotate "Captain Picard" as an entity).
- Do not annotate articles/determiners as parts of entities ("The Enterprise" should be annotated as "Enterprise", for example.)
  - "A Romulan warbird crash landed on Earth" --Tag "Romulan warbird" without "A"
- Annotate stage directions as normal.
- Don't annotate dialogue tags as PER (e.g. "KIRK: Bones, I'm a busy man." --Don't tag "KIRK"
- Do not annotate DESCRIPTIONS or other pointers. For instance:
  - Generic location stage settings like [Forest] or [Camp] are not LOC
  - Multiword descriptions pointing to specific entities, which may or may not contain named entities, are not entities
    - 'Federation colony on Beta Agnii Two'
      - This contains two entities ('Federation', GPE, and 'Beta Agnii Two', LOC) but is not itself LOC or any entity
    - 'The oldest planet in the star system'
    - 'My only yeoman'
  - Pronouns -- these are not PER
  - Nicknames or forms of address are *not* descriptions -- they are treated as names and *are* annotated
    - I.e., 'The Captain' when referred to a specific individual who happens to be captain in third person, or 'Captain' when addressing said individual
    - Contrast this with a sentence such as "The captain of the Romulan ship beamed aboard", in which 'the captain' (referring to an as-yet unspecified individual) is a mere descriptor

Entity Types

- PER
  - Tag names, nicknames, and titles of people when they refer to a specific individual
  - A note on titles:

- ■ Only tag individuals who are referred to by name or title. Do not tag descriptions of / pointers to individuals.
  - ● "The captain_PER was here. He told me there were only Klingons on the ship."
  - ● "Captain_PER, there's a message for you!"
  - ● "My only yeoman and two others dead, seven injured." --Do not tag anything here.
- ■ Do NOT tag titles when referring generically to a position, such as "Oh, you'll make captain one day!!" or "This ship really needs a captain"
- ○ A note on extent:
  - ■ Do include titles when titles appear near a person's name
    - ● "Ensign Ro" > [Ensign Ro]_PER
    - ● "Lieutenant Commander Data" > [Lieutenant Commander Data]_PER
  - ■ Do not tag descriptive content that is part of the NP but isn't a title/name
    - ● "The highly logical Spock" > [Spock]_PER"
    - ● "The half-human Worf" > half-[human]_SPEC, [Worf]_PER
    - ● "The Captain of the Enterprise, James Kirk" > [James Kirk]_PER
  - ■ "Captain Christopher Pike, United Space Ship Enterprise." > [Captain Christopher Pike]_PER, [United Space Ship Enterprise]_SHIP
- ○ Tag nicknames (like "Bones").
- ○ Do NOT tag pronouns
- ● LOC
  - ○ Specific places, planets, rooms in the ship, areas of space, solar system,
  - ○ Starbases are included here and not in SHIP since they are *basically* stationary and operate more like planets than ships.
  - ○ Examples: Starbase Eleven, Talos Four, Neutral Zone, holodeck, Beta Renner system
  - ○ Includes ship locations: Transporter Control, Medical, engine room, Ten-Forward,
  - ○ In "the Talos star group", annotate "Talos star group" since "star group" is a crucial part of the definition of the LOC.
  - ○ When "Engineering" refers to an area of the ship (e.g. "There's a hull breech in Engineering!") annotate as LOC. When "Engineering" refers to the group of people working there (e.g. "I just got transferred to Engineering", "She's the most accomplished officer in Engineering", "Engineering to Sick Bay"), annotate as ORG.
  - ○ If a location is described generically, such as "a very crowded asteroid belt" or "the planet", do not annotate this (only annotate specific references to specific locations)
- ● SHIP
  - ○ Manmade space-faring vehicles/places to live which can move through space. Shuttlecraft are included here, as well as war ships.

- ○ Starbases are not included here since they are essentially unmoving and operate more as planets than ships. (Starbases are considered LOC.)
  - ○ Examples: Enterprise, Romulan warbird
  - ○ When the name of a ship is preceded by a generic descriptor of the craft type, tag only the (Proper) name of the ship -- i.e., in 'the starship Voyager' -- tag only 'Voyager'.
- GPE
  - ○ A composite entity defined by having a population and a government.
  - ○ The definition of GPE is complicated slightly in this domain since many groups we would normally consider "governing bodies" tend to have a less direct relationship with a certain geographic location (as compared with an Earth-centric domain, like real life).
  - ○ "Klingon Empire" seems like a GPE since it is focused on a certain planet. However, "Federation" also has a certain area of space which it occupies, and although it is not explicitly aligned with any particular race, it is overwhelmingly Human and headquartered on Earth.
  - ○ Examples: Nazi Germany, Klingon Empire, Federation, Dominion
  - ○ Disambiguation: GPE vs. LOC:
    - ■ My actions were taken in the best interest of Vulcan_GPE and the High Command_ORG.
    - ■ Annotate planets, star systems, and other geographical entities as LOC by default UNLESS it is clear from context that the referent is functioning explicitly as a people with a government, as in the above example. If it's unclear whether the entity is functioning as a GPE, ask yourself: in the context in which it is found, can the entity of focus be said to possess agency or animacy? If the entity would otherwise be a LOC and the answer is 'no', then it is still a LOC.
- ORG
  - ○ To distinguish it from GPE, ORG may have a 'government' of sorts, but no population whose members are not also members of the 'government'.
  - ○ Examples: Galactic Council, Starfleet, launch bay crew
  - ○ Ship-group: e.g. Engineering, Medical
    - ■ "Engineering to Sick Bay" and the like are pretty common when someone from one group is calling the other group on the intercom.
- SPEC (SPECIES)
  - ○ Tag all explicitly named animal species, humanoid or otherwise
    - ■ We taggin' cows
  - ○ This can be a noun ('the Jarada_SPEC'), or an adjective ('Vulcan_SPEC monastery')
    - ■ "the half-Vulcan_SPEC officer Spock_PER"
  - ○ Examples:

- - - 'We are about to make a brief but necessary contact with the Jarada_SPEC, a reclusive, insect-like race known for its idiosyncratic attitude towards protocol'
    - Tips: Don't highlight articles -- only the species name itself
  - When the species name embedded within the name of an ORG, such as 'the Vulcan council' or some such, do not annotate as SPEC, prefer the larger (greedy) ORG annotation
  - Even if a SPEC is acting in an apparent sociopolitical capacity (i.e., at war with another named SPEC), prefer the SPEC tag over GPE. GPE is reserved for those entities with a clear, explicit government which extends over a population of governed subjects, a level of definition not easily achieved (and which should not be read into the mere mention of any group united by common interest, culture, or genetic similarity).

Relevant outside sources

- ACE Entity annotation guidelines for English:
  - https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/english-entities-guidelines-v6.6.pdf
- An annotated dataset of literary entities
  - https://www.aclweb.org/anthology/N19-1220.pdf