NOTE: Salient changes from the previous version of the guidelines are highlighted in yellow.

# Annotation Guidelines for Social DETERMINANTS OF HEALTH

Social history sections from clinical notes will be annotated for a range of social determinants of health (SDOH), including socio-demographic, behavioral, and environmental exposure factors. The annotation involves the identification of SDOH *events*, where each SDOH *event* is represented by a *trigger* and set of entities. The trigger consists of a multi-word span (word or phrase) and a label indicating the type of SDOH (e.g. employment or tobacco use). Figure 1 is an example of an annotated tobacco *event* from the BRAT annotation tool, which will be used in this annotation project. In Figure 1, the trigger span is “smokes” and the trigger label is *tobacco*. This tobacco event is characterized by a *status* argument, consisting of a multi-word span and label indicating the tense of the event (i.e. *current*), as well as *amount*, *type*, *frequency*, and *exposure history* arguments. All annotated phenomena are defined in terms of the span (words associated with phenomena) and the span type (e.g. *amount*, *status*, etc.), and some annotated phenomena, like *status*, will also include a span label (e.g. *current* or *past*). If an annotated span includes a noun, the selected span should include the entire applicable noun phrase. If an annotated span includes a verb, the selected span should include applicable adverbs. The inclusion of entire noun phrases and relevant adverbs is applicable to all annotated phenomena.

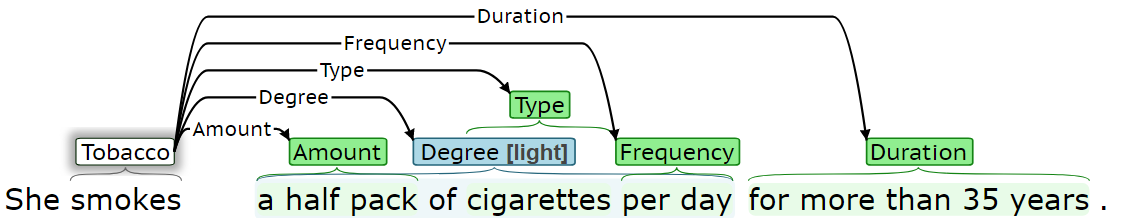


Figure 1. Annotation example

The annotated factors include:

* Socio-Demographic:

1. Employment
2. Insurance
3. Living status
4. Sexual orientation
5. Gender identity
6. Country of origin
7. Race

* Behavioral:

1. Substance abuse (tobacco, alcohol, drug)
2. Physical activity

* Environmental exposure:

1. Environmental exposure

The following sections describe how each determinant will be annotated. For each annotated determinant, some entities should always be present and are required. However, some entities will not always be present. Required entities for each determinant are indicated for each determinant.

# A. Socio-Demographic:

## (1) Employment

*Employment* will be annotated with the following information:

**(1) Trigger (required):** The trigger is the span that most clearly indicates an employment event is present. Table 1 presents several trigger span annotation examples. The trigger span should be a work-related key phrase (e.g. “works,” “occupation,” “employed,” “unemployed,” “on disability”, “retired,” “homemaker,” etc.), if present. Otherwise, it should be a specific profession. In notes with employment-related subheadings, like “Occupation:…” or “Employment:…,” the subheading without the colon (i.e. “Occupation” or “Employment”) should be the trigger.

Table 1. Employment - trigger examples. Underlined words indicate the identified span.

|  |  |
| --- | --- |
| **Priority** | **Examples** |
| 1. Work-related key phrase | Occupation: Retired  She used to be employed as a registered nurse.  … she works at a nursing home.  She worked as a travel agent.  … she is retired.  … the patient is widowed and a retired engineer.  He is unemployed.  …the patient is on disability.  … the patient is a homemaker.  Currently works in construction. |
| 2. Profession | She is a full-time student at University.  The patient is a registered nurse by profession.  She is an office manager for a gravel company. |

**(2) Status (required):** *Status* annotation consists of a span selection and a label (*employed*, *unemployed*, *retired*, *on disability*, *student*, or *homemaker*). Table 2 presents *status* examples. Where a patient’s previous employment is described without a description of current employment, the status label should be *unemployed*, unless the note explicitly describes retirement, being on disability, etc. Figure 2 presents an annotated *employment* event, where it can be inferred that the patient is of retirement age; however, the *status* label should *unemployed*, as there is no explicit reference to retirement. In some cases, a patient’s employment status may be discussed in multiple sentences. The annotated notes will likely be processed sentence-by-sentence, and individual events should be confined to a single sentence, where possible (see “Defining Events” section at the end of this document for details). If there is an *employment* trigger in each sentence, then separate *employment* events should be created in each sentence. Figure 3 presents annotated *employment* events across multiple sentences, where the first sentence describes retirement and the second sentence describes previous work. The previous employment in the second sentence in Figure 3 should be labeled as *unemployed*, as there is insufficient information within the sentence to assign a different label. The conflicting *employment* *status* labels in the first and second sentences will be resolved in downstream postprocessing, separate from the annotation process.

Table 2. Employment - type. Underlined words indicate the identified span.

|  |  |
| --- | --- |
| **Type value** | **Examples** |
| employed | Occupation: works  … she works at a nursing home.  The patient is a registered nurse by profession.  Currently works in construction. |
| unemployed | He is unemployed. |
| retired | … the patient is widowed and a retired engineer. |
| on disability | …the patient is on disability.  He is married, on medical leave from his job… |
| student | She is a full-time student at University. |
| homemaker | … the patient is a homemaker. |

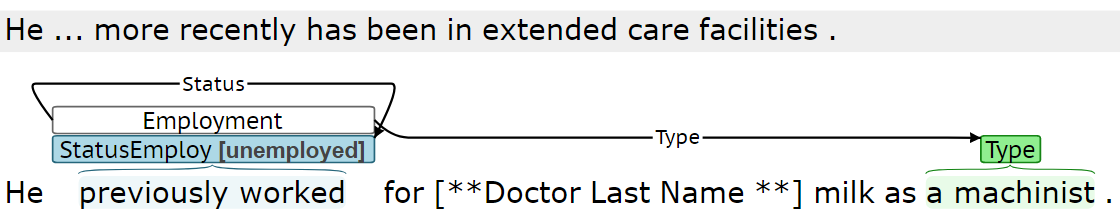


Figure 2. Employment - status example where retirement not explicitly referenced

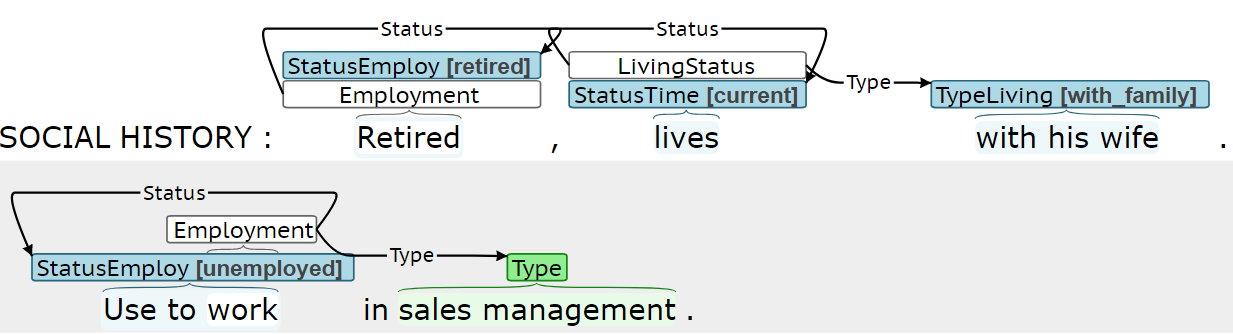


Figure 3. Employment - status example with multiple employment events across multiple sentences

**(3) Duration:** *Duration* spans describe how long the event persisted. Table 3 presents *duration* span examples. *Duration* spans will often be a prepositional phrase or adverb phrase, in which case the entire phrase should be selected. **This is applicable to all determinants for which *duration* is a valid argument.**

Table 3. Employment – duration examples. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| He works as a plumber over the last 5 years.  He has worked as a tax attorney for the past 48 years… |

**(4) History:** *History* spans describe how long ago an event occurred. Table 4 presents *history* span examples. *History* spans will often be a prepositional phrase or adverb phrase, in which case the entire phrase should be selected. **This is applicable to all determinants for which *history* is a valid argument.**

Table 4. Employment - history examples. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| She stopped working more than 20 years ago due to…  She retired 7 years ago because of low back pain.  She lost her job in 2018. |

**(5) Type:** *Type* indicates the occupation or profession of the patient. Table 5 presents *type* examples. *Type* span should be a specific profession, if present. Otherwise, the *type* should be a span that conveys the general type of work performed or location of work.

Table 5. Employment - type. Underlined words indicate the identified span.

|  |  |
| --- | --- |
| **Priority** | **Examples** |
| 1. Specific profession | He works as a geologist at Petrogas.  She is currently a registered nurse, circulator, and scrub technician in…  … He is employed as a United States Marine officer, artillery repair specialist.  She is retired from running a restaurant. |
| 2. Type of work or location of work | … he previously worked with pipeline work…  Works in a business office.  She is married long-term, retired from Avon. |

Figure 4 and Figure 5 are example BRAT *employment* annotations. Figure 4 includes a single employment event; however, the employment description in Figure 5 must be represented through two separate events, to both represent the previous work in the Air Force and current work at Lockheed Martin.

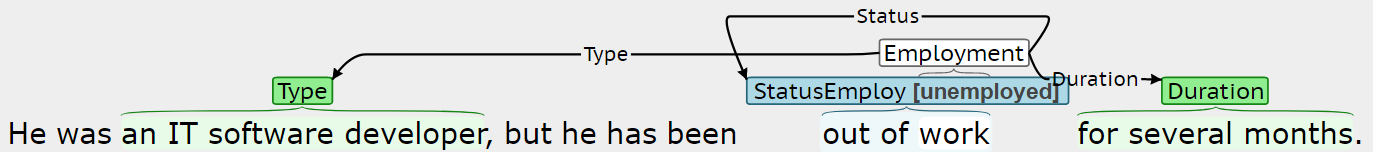


Figure 4. Employment example 1

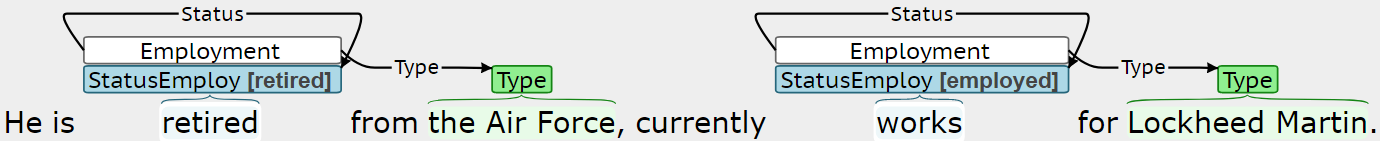


Figure 5. Employment example 2

## (2) Insurance

*Insurance* will be annotated with the following information:

**(1) Trigger (required):** The trigger is the span that most clearly indicates an insurance event is present. Table 8 presents several trigger span annotation examples. The trigger span will likely include some form of “insurance,” “insured,” “coverage,” etc.

Table 6. Insurance - trigger examples. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| … her insurance would not cover the medication.  He has been off insurance for over a year.  The patient will have continued pain medication coverage with Darvocet and Flexeril.  …his inability to have adequate health insurance. |

**(2) Status (required):** *Status* is a binary label indicating whether or not the patient is adequately insured: *yes* (has insurance) or *no* (does not have insurance or is underinsured).

Table 7. Insurance - status examples. Underlined words indicate the identified span.

|  |  |
| --- | --- |
| **Status label** | **Examples** |
| yes | The patient will have continued pain medication coverage… |
| no | … her insurance would not cover the medication.  He has been off insurance for over a year.  …transferred to UIHC for a lack of insurance…  …his inability to have adequate health insurance. |

## (3) Living Status

*Living status* will be annotated with the following information:

**(1) Trigger (required):** The trigger is the span that most clearly indicates a *living status* event is present. Table 8 presents several trigger span annotation examples. The trigger span will likely be a form of “lives” or “resides.”

Table 8. Living status - trigger examples. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| …she lives with her husband.  The patient does live with his family.  … the patient has been living in his father’s home…  He currently lives at home with his parents. |

**(2) Status (required):** *Status* indicates whether the *living status* event is *current*, *past*, or *future*. The highlighted span should be the shortest span that distinguishes between *status* labels. *Status* will often include a verb.

Table 9. Living status - status examples. Underlined words indicate the identified span.

|  |  |
| --- | --- |
| **Status label** | **Examples** |
| current | He lives in a dorm there.  … patient is a resident of Lake Harris…  She is homeless.  He currently lives at home with his parents. |
| past | She was living with her parents… |
| future | She will be moving in with her parents for the summer. |

**(3) Type (required):** *Type* annotation consists of a span selection and a label (*alone*, *with family*, *with others*, or *homeless)*. Table 2 presents *type* examples. *Type* will often be a prepositional phrase.

Table 10. Living status - type. Underlined words indicate the identified span.

|  |  |
| --- | --- |
| **Type label** | **Examples** |
| alone | … she lives by herself.  She lives alone at home. |
| with family | … he lives with parents and siblings.  Lives with her husband.  He…lives with his partner. |
| with others | … she…is now living in assisted living. |
| homeless | … is currently homeless and in need of housing…  She lives on the streets for the past 6 months. |

**(4) Duration:** *Duration* spans describe how long the event persisted, for example: “for seven years” or “for the past year.”

**(5) History:** *History* spans describe how long ago an event occurred, for example: “until about a month ago” or “five years ago.”

Figure 6 and Figure 7 are example BRAT *living status* examples.

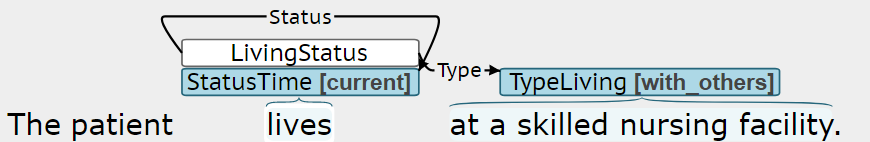


Figure 6. Living status example 1

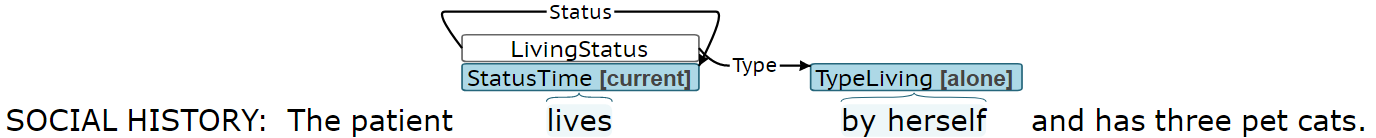


Figure 7. Living status example 2

## (4) Sexual Orientation

Explicit descriptions of patient *sexual orientation* will be annotated, for example “Patient is heterosexual…” or “His partners are male.” However, implicit descriptions of patient sexual orientation, like “He is here with his wife.” or “Her wife…,” should not be annotated.

**(1) Trigger (required):** The trigger is the span that most clearly indicates a *sexual orientation* event is present. Table 11 presents several trigger span annotation examples. Trigger spans should be explicit references to sexual orientation (i.e. homosexual, heterosexual, or bisexual), if present.

Table 11. Sexual orientation - trigger examples. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| He is heterosexual and…  He is homosexual.  She is a lesbian.  His partners are male… |

**(2) Status (required):** Status indicates whether the *sexual orientation* event is *current* or *past*. Table 12 presents *status* span annotation examples. *Status* will often include a verb, indicating tense (past vs. present).

Table 12. Sexual orientation - status examples. Underlined words indicate the identified span.

|  |  |
| --- | --- |
| **Status label** | **Examples** |
| current | He is heterosexual and…  His partners are male… |
| past | The patient participated in homosexual activity in Haiti during 1982… |

**(3) Type (required):** Type indicates the patient's sexual orientation and consists of a span selection and a label (*heterosexual*, *homosexual*, or *bisexual)*.

## (5) Gender identity

*Gender identity* will be annotated with the following information:

**(1) Trigger (required):** The trigger is the span that most clearly indicates a *gender identity* event is present. Table 13 presents several trigger span annotation examples. Trigger spans should be explicit references to gender identities (i.e. cisgender or transgender), if present. Otherwise, the trigger spans should be phrases, like “identifies as,” “uses pronouns,” or “gender non-conforming.”

Table 13. Gender identity - trigger examples. Underlined words indicate the identified span.

|  |  |
| --- | --- |
| **Priority** | **Examples** |
| 1. Gender identity category | He is transgender. |
| 2. Description of gender identity | She is biologically male but identifies as female.  He is biologically female but uses pronouns he/him.  She is gender non-conforming. |

**(2) Status (required):** *Status* indicates whether the *gender identity* event is *current* or *past*. Table 14 presents *status* examples. The span should be the shortest span that distinguishes between *status* labels and will typically include a verb.

Table 14. Gender identity - status examples. Underlined words indicate the identified span.

|  |  |
| --- | --- |
| **Status label** | **Examples** |
| current | He is transgender.  She is biologically male but identifies as female. |
| past | Previously identified as female. |

**(3) Type (required):** *Type* annotation consists of a span selection and a label (*cisgender, transgender)*.

## (6) Country of Origin

*Country of origin* will be annotated. Note that *country of origin* events should not be associated with international travel.

**(1) Trigger (required):** The trigger is the span that most clearly indicates the patient’s country of origin is being described. Table 15 presents trigger examples. The trigger span should not be the name of a country. The trigger will frequently be similar to “is from.”

Table 15. Country of origin - trigger examples. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| The patient is from Ukraine.  He is originally from Venice, Italy… |

**(2) Type (required):** *Type* is the span that indicates the specific country from which the patient originates. Table 16 presents *type* examples. *Type* will typically be a proper noun, and the entire noun phrase should be selected. For example, the city should be included in the span selection, if provided.

Table 16. Country of origin - type examples. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| The patient is from Ukraine.  He is originally from Venice, Italy… |

## (7) Race

*Race* will be annotated with the following information:

**(1) Trigger (required):** The trigger is the span that most clearly indicates race or ethnicity is being described. Table 17 presents trigger examples. If present, the trigger should be a term like “race” or “ethnicity.” Otherwise, the trigger should be a specific race or ethnicity (e.g. African-American).

Table 17. Race - trigger examples. Underlined words indicate the identified span.

|  |  |
| --- | --- |
| **Priority** | **Examples** |
| 1. Term for race/ethnicity | Race: Caucasian. |
| 2. Specific race/ethnicity | She is African American. |

**(1) Type (required):** *Type* is the span that indicates the patient’s racial background with the most specificity. Table 18 presents type examples.

Table 18. Race - type examples. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| Race: Caucasian.  She is African American. |

# B. Behavioral:

## (1) Substance Abuse - Alcohol/drug/tobacco

Alcohol, drug, and tobacco use will be annotated. Drug events should involve marijuana, illegal drugs, or the abuse of prescription drugs. Drug events should not include the medical use (intended use) of pharmaceuticals. Only first-hand smoking should be annotated as tobacco events. Secondhand smoking should be annotated as environmental exposure. Keep in mind that alcohol may be abbreviated ETOH in chart notes.

**(1) Trigger (required):** The trigger is the span that most clearly indicates an alcohol, drug, or tobacco event is present. Table 19 presents several trigger span annotation examples. The trigger span should be a noun phrase describing a general substance type, like “alcohol,” “ETOH,” “tobacco history,” or “illicit drug use,” if present. If general substance types are not present, the span should be a relevant verb (e.g. “drink” or “smokes”) or a noun phrase describing a specific substance type (e.g. “cocaine” or “beer”).

Table 19. Substance trigger examples. Underlined words indicate the identified span.

|  |  |
| --- | --- |
| **Priority** | **Examples** |
| 1. Substance words | Alcohol none.  Does not consume alcohol.  He does not drink alcohol.  He denies tobacco use.  Tobacco history: smokes two packs…  He has a history of illicit drug use…  He used to drink alcohol rarely…  …no ETOH. |
| 2. Verb or substance type | Admits to heroin use…  …she does not drink…  He does drink six beers a day.  He has not smoked for the past 27 years.  He does not smoke.  He has about 10 drinks a year.  …patient experimented with amphetamines, cocaine, marijuana… |

**(2) Status (required):** Status indicates whether the substance consumption is *none*, *current*, or *past*. The *none* label indicates the patient does not use the substance currently and has not used it in the past. The highlighted span should be the shortest span that distinguishes between the *status* labels and will often include a verb. Table 20 presents *status* examples.

Table 20. Substance abuse- status examples. Underlined words indicate the identified span.

|  |  |
| --- | --- |
| **Status label** | **Examples** |
| none | The patient denies any alcohol.  He currently denies any drinking.  … she does not drink…  does not consume alcohol.  No alcohol abuse.  No history of drug or alcohol abuse.  She is a nonsmoker and nondrinker.  She uses no illicit drugs.  No use of illicit drugs.  He never used…drugs  Alcohol none. |
| current | She occasionally has alcohol.  He does have chronic alcohol use.  She drinks 2 to 3 glasses of wine…  Occasionally drinks.  Patient currently has 2-3 drinks per day.  He does drink six beers a day.  He is a social drinker.  She does have a history of chronic marijuana use.  Also admits today using cocaine.  He does smoke.  Social history: active smoker.  History of drinking for approximately 25 years.  Occasional alcohol use. |
| past | He previously was an alcoholic…  She used to smoke…  She quit smoking in 1972.  He used cocaine in the past…  The patient experimented with amphetamines…  He admits to having smoked marijuana…  Personal history: He is an ex-smoker.  He has a history of tobacco use in the past. |

**(3) Duration**: *Duration* indicates the length of time the patient used the substance(s) described in the substance *event*. Table 21 presents *duration* examples.In clinical shorthand, “x” can be interpreted as “times” or “for” and should be included in *duration* spans where applicable (e.g. “x 10 years”).

Table 21. Substance abuse - duration. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| She smokes 5 cigarettes per day, has done so for 10 years.  He also quit smoking cigarettes in 1984 , after 16 years of smoking.  The use of alcohol, drugs , or tobacco during the pregnancy were denied.  ...the patient had a long history of smoking but quit many years ago.  She quit smoking 3 years ago after 18 years of heavy smoking.  Tobacco : smokes 4 packs / day x ~40 years . |

**(4) History**: *History* indicates when a patient quit using a given substance. Table 22 presents *history* examples.

Table 22. Substance abuse - history. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| …used to smoke a pack a day and quit in 1985.  …the patient had a long history of smoking but quit many years ago.  …he stopped using drugs when he was 25.  She quit smoking 3 years ago after 18 years of heavy smoking. |

**(5) Method:** *Method* indicates how the substance is used by the patient. Table 23 presents *method* examples. *Method* does not need to be annotated for alcohol, as *method* is always oral (drinking). The *method* for tobacco use is assumed to be smoke, and only non-smoking tobacco use methods, like chewing or vaping, need to be annotated.

Table 23. Substance abuse - method. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| He smokes pipe.  He chews tobacco… |

**(6) Type:** *Type* indicates the specific type of substance used by the patient, beyond the general substance categories of alcohol, drug, and tobacco. Table 24 presents *type* examples.

Table 24. Substance abuse - type. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| He smokes pipe.  She drinks wine a few drinks a month.  He occasionally has a beer.  …he is a non-cigarette smoker.  …some minor secondhand tobacco exposure.  … but dips (smokeless) tobacco.  He used cocaine in the past…  No history of recreational drug use.  He denies alcohol or illicit drug use.  The patient experimented with amphetamines, cocaine, marijuana approximately… |

**(7) Amount:** *Amount* indicates the quantity of the substance used by the patient. Table 25 presents *amount* examples. *Amount* spans are quantitative description of the substance quantity, (e.g. “three drinks” or “2 packs”) and do not include qualitative descriptions (e.g. "significant" or “heavy”). If *amount* includes a noun, the entire noun phrase should be selected. In general, *amount* spans should not include *frequency* information, and *frequency* and *amount* spans should not overlap.

Tobacco use is often quantified by packs per day (ppd) and pack-years. Although ppd and pack-years describe both the amount and frequency of tobacco use, ppd and pack-year references should be labeled as *amount*. With the exception of smoking *ppd* and *pack-year* nomenclature, *amount* spans should not include the *frequency* span information (e.g. “per day”).

Table 25. Substance abuse – amount. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| She drinks one glass of wine per day.  He drinks alcohol occasionally with up to three drinks a week.  Smoking history: smoked 1 ppd, quit…  …he has cut down to half a pack per day.  Tobacco : smokes 4 packs / day x ~40 years .  Forty pack year smoking history. |

**(8) Frequency:** *Frequency* indicates how often the given substance is used by the patient. Typically, the *frequency* span describes how often the identified *amount* is used. Table 26 presents *frequency* examples. *Frequency* spans may be prepositional phrases starting with prepositions like “per” or “every.” In clinical shorthand, forward slashes, “/”, can be interpreted as “every” or per” and should be included in *frequency* spans where applicable (e.g. “2 drinks / day”). Qualitative descriptions like “occasional” or “rarely” will typically be annotated as part of *degree*, as opposed to *frequency*.

Table 26. Substance abuse - frequency. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| She drinks one glass of wine per day.  He drinks alcohol occasionally with up to three drinks a week.  …he has cut down to half a pack per day.  … with smoking a pack every 6 days.  The patient reports smoking marijuana a few times in his life…  Tobacco : smokes 4 packs / day x ~40 years . |

**(9) Degree**: *Degree* describes alcohol, drug or tobacco use through a set of fixed classes: *none*, *light*, *moderate*, or *heavy*. The *degree* will typically correspond with the *amount* and *frequency* spans; however, the *type*, *exposure history*, and *quit history* spans may be relevant. The *degree* span should be the shortest span that contains all of the information required to determine the *degree* label. Table 27 describes the *degree* labels for each substance. Regarding tobacco use, 1 pack year is equivalent to smoking 1 pack per day (ppd) for 1 year. The *degree* will always be *none* when the *status* is *none*, so *degree* *none* labels do not need to be annotated.

Table 27. Substance Degree levels (UNDER CONSTRUCTION)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Substance** | **Degree** | | | |
| **none** | **light** | **moderate** | **heavy** |
| Any substance | No usage | Described as “mild,” “occasional,” or “social” | Described as “moderate” | Described as “abuse,” “heavy,” “significant,” or “chronic” |
| Alcohol | No usage | < X drinks/day | X-Z drinks/day | > Z drinks/day |
| Drug | No usage | ? | ? | ? |
| Tobacco | No usage | < X packs/day  < C pack years  <K pods/day  < A cans of chew/day | X-Z packs/day  C-D pack years  K-L pods/day  A-B cans of chew/day | > Z packs/day  > D pack years  > L pods/day  > B cans of chew/day |

Figure 8 is an example BRAT annotation for substance use.

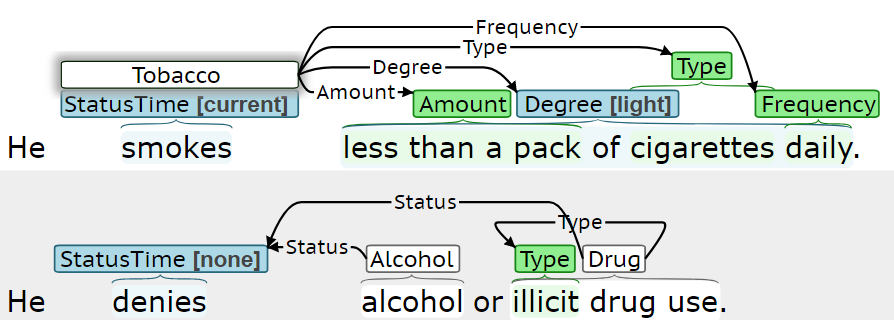


Figure 8. Substance use example

## (2) Physical Activity

*Physical activity* will be annotated with the following information:

**Trigger (required):** *Trigger* is the span that most clearly indicates physical activity. Table 11 presents several *trigger* span annotation examples. The trigger span should be a general phrase indicating activity (e.g. “active,” “sedentary,” “exercises,” etc.), if present. Otherwise, the span should be a specific activity (e.g. “walks” or “runs”).

Table 28. Physical activity trigger examples. Underlined words indicate the identified span.

|  |  |
| --- | --- |
| **Priority** | **Examples** |
| 1. General | He travels a lot and is extremely active.  She used to exercise vigorously…  She exercises weekly, mainly tennis and swimming. |
| 2. Specific | They walk every evening for one hour.  She likes to swim in the summertime.  She will be involved in cheerleading…  Currently walks 30 minutes a day. |

**(1) Status (required):** *Status* indicates whether the physical activity event is *current* or *past*. Table 29 presents status span annotation examples. The span should be the shortest span that distinguishes between status labels and will typically include a verb.

Table 29. Physical activity - status examples. Underlined words indicate the identified span.

|  |  |
| --- | --- |
| **Status label** | **Examples** |
| current | Work, diet, and exercise patterns are within normal limits.  He jogs or does cross-country track five times a week…  She also likes to play softball.  …, but he enjoys lifting weights…  Currently walks 30 minutes a day. |
| past | She used to exercise vigorously…  He has not been participating in regular exercise now because… |

**(3) Duration:** *Duration* spans describe how long the event persisted. Table 30 presents *duration* span examples.

Table 30. Employment – duration examples. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| She has not been getting regular exercise for approximately two years…  He runs since 1980. |

**(4) History:** *History* spans describe how long ago an event occurred. Table 31 presents *history* span examples.

Table 31. Employment - history examples. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| She quit running 2 years ago due to… |

**(5) Type:** *Type* indicates the specific type of physical activity. Table 32 presents *type* examples.

Table 32. Physical activity - type. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| He is very active and walks about four miles every few days.  She likes to swim in the summertime.  She will be involved in cheerleading, track , volleyball , and basketball. |

**(6) Amount:** *Amount* indicates the quantity physical activity. Table 33 presents *amount* examples.

Table 33. Physical activity – amount. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| He is very active and walks about four miles every few days.  They walk together every evening for one hour. |

**(7) Frequency:** *Frequency* indicates how often the given substance is used by the patient. Table 34 presents *frequency* examples.

Table 34. Physical activity - frequency. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| He is very active and walks about four miles every few days.  They walk together every evening for one hour. |

**(8) Degree:** *Degree* categorizes physical activity as *none*, *light*, *moderate*, or *heavy*. *Degree* will typically correspond with the *type*, *amount*, and *frequency* spans. The span should be the shortest span that contains all the information required to assess *degree*. Table 35 describes how to determine the *degree* label.

Table 35. Physical activity degree (UNDER CONSTRUCTION)

|  |  |
| --- | --- |
| **Degree** | **Description** |
| none | Inactive or sedentary lifestyle. |
| light | Lower-degree exercise X-Y times per week, for example:  “walking 2 times per week” |
| moderate | TBD |
| heavy | Vigorous exercise or physically demanding labor multiple times per week. Higher-degree exercise includes activities like running, swimming, biking, lifting, etc. Higher-degree labor includes working in construction, etc. |

Figure 9 and Figure 10 are example BRAT *physical activity* annotations.

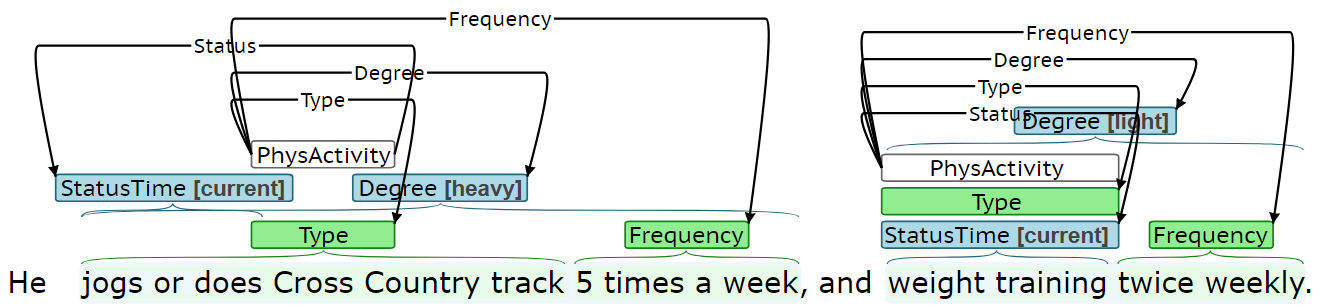


Figure 9. Physical activity example 1

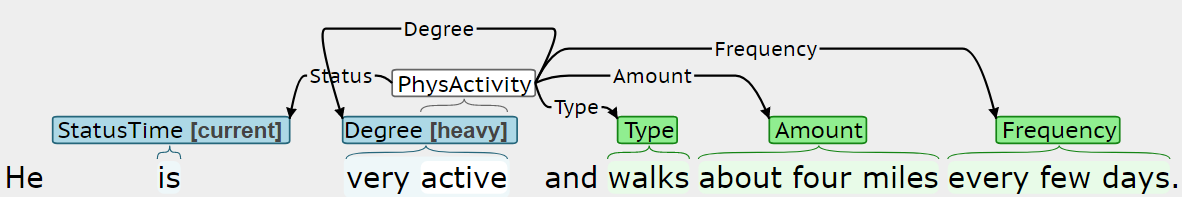


Figure 10. Physical activity example 2

# C. Environmental exposures:

## (1) Environmental exposure:

Environmental exposure includes exposure to things like secondhand smoke, asbestos, chemicals, industrial gases, or pets. Environmental exposure also includes the absence of any relevant exposures (see status none examples in Table 37).

**(1) Trigger (required):** *Trigger* is the span that most clearly indicates an *environmental exposure* event is present. Table 36 presents several *trigger* span annotation examples.

Table 36. Environmental exposure - trigger examples. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| There are no pets in the home.  No known exposures.  There is no smoking in the home.  Denies exposures to asbestos, chemicals, or industrial gases.  At home also live 2 dogs and 2 outside cats.  There is second-hand smoke exposure.  …previously did a lot of work in jewelry business, working he states with chemicals. |

**(2) Status (required):** *Status* indicates whether the environmental exposure is *none*, *current*, or *past*. The *none* *status* label indicates the patient does not have relevant environmental exposure, currently or in the past. The highlighted span should be the shortest span that distinguishes between the *status* labels. The *status* span will often include a verb. Table 32 presents *status* examples.

Table 37. Environmental exposure - status examples. Underlined words indicate the identified span.

|  |  |
| --- | --- |
| **Status label** | **Examples** |
| none | There are no pets in the home.  No known exposures.  Exposure: negative to asbestos.  There is no smoking in the home.  Denies exposures to asbestos, chemicals, or industrial gases. |
| current | At home also live 2 dogs and 2 outside cats.  There is second-hand smoke exposure.  Currently exposed to chemicals and industrial gases… |
| past | …previously did a lot of work in jewelry business, working he states with chemicals. |

**(3) Duration:** *Duration* spans describe how long the event persisted, for example: “for seven years” or “past year.”

**(4) History:** *History* spans describe how long ago an event occurred, for example: “until about a month ago” or “five years ago.”

**(5) Method:** *Method* indicates how the patient was exposed to the environment. *Method* spans will describe contact, inhalation, indigestion, consumption, respiratory, hearing, etc. Table 38 presents *method* examples.

Table 38. Environmental exposure - method. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| Loss of hearing, secondary to profession.  Respiratory exposure from profession. |

**(6) Type:** *Type* indicates the specific type of environmental exposure. Table 39 presents *type* examples.

Table 39. Environmental exposure - type. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| At home also live 2 dogs and 2 outside cats.  Denies exposures to asbestos, chemicals, or industrial gases.  There is no smoking in the home.  …previously did a lot of work in jewelry business, working he states with chemicals. |

**(7) Amount:** *Amount* indicates the quantity of the exposure. Table 40 presents *amount* examples.

Table 40. Environmental exposure - amount. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| Her father smokes a pack a day. |

**(8) Frequency:** *Frequency* indicates how often the patient was exposed. Table 41 *frequency* examples.

Table 41. Environmental exposure - frequency. Underlined words indicate the identified span.

|  |
| --- |
| **Examples** |
| Her father smokes a pack a day. |

Figure 11 is an example BRAT *environmental exposure* annotation.

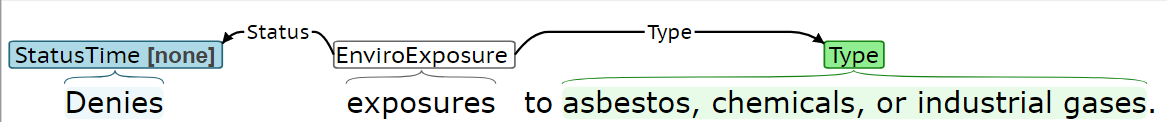


Figure 11. Environmental exposure example

# summary

This section contains a summary of all of the annotated phenomena. The social determinant information will be annotated using BRAT, which is a web-based tool. To reduce the length of some annotated phenomena, longer phenomena names have been truncated/abbreviated within the BRAT environment. Additionally, the names of some phenomena have been expanded to differentiate between different sets of labels within the BRAT environment. Table 42, Table 43, and Table 44 summarize all of the annotated phenomena. These tables include parenthetical references the specific naming convention used within BRAT. The required entities for each determinant type are highlighted in this table with a light blue background.

Table 42 presents a summary of the socio-demographic factors that will be annotated.

Table 42. Summary of socio-demographic annotation. A light blue background indicates a required entity.

|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **Entity** | **Labels** | **Span** |
| (1) Employment | trigger | Employment  (*Employment*) | Employment related phrases, like “occupation,” “employed,” “works,” “retired,” “a student,” “a registered nurse,” etc. |
| status  (*StatusEmploy*) | • employed  • unemployed  • retired  • on disability  • student  • homemaker | “currently works,” “a registered nurse,” “unemployed,” “medical leave,” “full-time student” |
| duration  (*Duration*) | -- | Descriptions of how long event/condition persisted, e.g. “for the last five years” or patterns like:  • for <#> <time unit>  • since <date> |
| history  (*History*) | -- | Descriptions of how long ago event occurred, e.g. “15 years ago” or patterns like:  • in <date>  • <#> <timeunit> ago |
| type  (*Type*) | -- | Specific occupations, like “a geologist” or “registered nurse,” general types of work, like “office work.” |
| (2) Insurance | trigger | Insurance  (*Insurance*) | Insurance related phrases, like “health insurance” or “coverage.” |
| status  (*StatusInsure*) | • yes  • no | Descriptions of insurance status, like “will have continued” or “has been off.” |
| (3) Living status | trigger | Living status  (*LivingStatus*) | Phrases like “lives” or “resides.” |
| status  (*StatusTime*) | • current  • past  • future | Phrases conveying tense, like “lives,” “was living,” “will be moving,” or “currently lives.” |
| type  (*TypeLiving*) | • alone  • with family  •with others  • homeless | Phrases like “by herself,” “alone,” “with her husband,” “in assisted living,” “on the streets,” “couch hopping” |
| duration  (*Duration*) | -- | Descriptions of how long event/condition persisted. See *employment* *duration* for examples. |
| history  (*History*) | -- | Descriptions of how long ago event occurred. See *employment* *history* for examples. |

Table 42 continued.

|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **Entity** | **Labels** | **Span** |
| (4) Sexual orientation | trigger | Sexual orientation  (*SexualOrient*) | Phrases conveying sexual orientation, like “heterosexual,” “lesbian,” “his husband,” etc. |
| status  (*StatusTime*) | • current  • past | Phrases conveying tense, like “is” or “participated in.” |
| type  (*TypeSexualOrient*) | • heterosexual  • homosexual  • bisexual | Phrases conveying sexual orientation, like “heterosexual,” “lesbian,” etc. |
| (5) Gender identity | trigger | Gender identity | Phrases conveying gender identity, like “transgender,” “identifies as,” or “uses pronouns.” |
| status  (*StatusTime*) | • current  • past | Phrases conveying tense, like “is” or “identified.” |
| type  (*TypeGenderID*) | • cisgender  • transgender | Phrases conveying gender identity, like “transgender,” “identifies as male,” or “uses pronouns she/her.” |
| (6) Country of origin | trigger | Country of origin  (*Country*) | Phrase indicating country of origin is discussed, like “is from” or “originally from” |
| type  (*Type*) | **--** | USA, England, … |
| (7) Race | trigger | Race  (*Race*) | Phrase indicating race or ethnicity is being discussed (e.g. “race,” “ethnicity,”, “African-American”) |
| type  (*Type*) | -- | Racial or ethnic group, like Caucasian, African American, Pacific Islander … |

Table 43 presents a summary of the socio-demographic factors that will be annotated. The annotated factors are described in more detail in the subsequent subsections.

Table 43. Summary of behavioral factor annotation.

| **Type** | **Entity** | **Labels** | **Span** |
| --- | --- | --- | --- |
| (1) Substance abuse (alcohol, drug, and tobacco) | trigger | • Alcohol (*Alcohol*)  • Drug (*Drug*)  • Tobacco (*Tobacco*) | Phrase describing general substance (i.e. alcohol, tobacco, or drug), substance-related verb (e.g. “drinks” or “smokes”), or a specific substance (e.g. “wine,” “heroin,” or “cigarettes”) |
| status  (*StatusTime*) | • none  • current  • past | • *none*: “currently denies,” “does not,” “no history,” “never used,” etc.  • *current*: “drinks,” “occasionally drinks,” “smokes,” “smoked,” etc.  • *past*: “quit,” “used,” etc. |
| duration  (*Duration*) | -- | Descriptions of how long substance used, e.g. “for the past eight years” or patterns like:  • for <#> <time unit>  • since <date> |
| history  (*History*) | -- | Descriptions of how long ago substance was used, e.g. “seven years ago” or patterns like:  • <#> <time unit> ago  • in <date> |
| method  (*Method*) | -- | • Alcohol: oral/drink implied, so no annotation needed  • Drug: smoke, snort, inhale, inject, etc.  • Tobacco: smoke implied, only annotate if chew or vape |
| type  (*Type*) | -- | • Alcohol: “beer,” “wine,” “hard liquor,” etc.  • Drug: “cocaine,” “heroin,” “marijuana,” “narcotics,” etc.  • Tobacco: “cigarettes,” “cigar,” “tobacco,” “vape” etc. |
| amount  (*Amount*) | -- | Substance-specific quantification of amount:  • Alcohol: # [of glasses/drinks/bottles/times/cans]  • Drug: <#> of times  • Tobacco: <#> [cigarettes/packs/times] |
| frequency  (*Frequency*) | -- | Phrases like: “daily,” “monthly,” “yearly,” etc. |
| degree  (*Degree*) | • none  • light  • moderate  • heavy | Typically, will include *amount* and/or *frequency* spans, like:  • “one glass of wine per day,” “significant,” “half a pack per day” |

Table 43 continued.

| **Type** | **Entity** | **Labels** | **Span** |
| --- | --- | --- | --- |
| (2) Physical Activity | trigger | Physical activity | Phrase describing activity/inactivity, like: “active,” “sedentary,” “runs,” etc. |
| status  (*StatusTime*) | • current  • past | Phrase conveying tense, like: “currently jogs,” “enjoys,” “use to,” “has not been,” etc. |
| duration  (*Duration*) | -- | Descriptions of how long event/condition persisted. See *employment* *duration* for examples. |
| history  (*History*) | -- | Descriptions of how long ago event occurred. See *employment* *history* for examples. |
| type  (*Type*) | -- | Specific type of activity, like: “walks,” “swims,” “cheerleading, track, volleyball,” etc. |
| amount  (*Amount*) | -- | Phrase conveying amount, like: “2 times,” “> 10 mins,” “4 miles,” etc. |
| frequency  (*Frequency*) | -- | Phrase describing frequency, like: “every few days,” “every evening,” “daily,” etc. |
| degree  (*Degree*) | • none  • light  • moderate  • heavy | Typically, will include *amount* and/or *frequency* spans, like: “walking 2 times per week” or “runs daily” |

Table 44 presents a summary of the environmental exposure factors that will be annotated.

Table 44. Summary of environmental exposure annotation. A light blue background indicates a required entity.

|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **Entity** | **Labels** | **Span** |
| (1) Environmental exposure | trigger | Environmental exposure | Phrase most clearly indicating exposure, like “exposure,” “contact,” etc. |
| status  (*StatusTime*) | • none  • current  • past | • *none*: “denies,” “does not,” “no history,” etc.  • *current*: “is,” “live,” etc.  • *past*: “previously did,” “was,” etc. |
| duration  (*Duration*) | -- | Descriptions of how long event/condition persisted. See *employment* *duration* for examples. |
| history  (*History*) | -- | Descriptions of how long ago event occurred. See *employment* *history* for examples. |
| method | -- | Phrase describing contact, inhalation, indigestion, etc. |
| type  (*Type*) | -- | Specific type of exposure, for example “asbestos,” “pets,” or “chemicals” |
| amount  (*Amount*) | -- | Typically, a qualitative description, like “moderate” or “significant”; however, could be a specific quantity, like the amount of a chemical ingested. |
| frequency  (*Frequency*) | -- | Phrases like: “rarely,” “socially,” “occasionally,” “daily,” “monthly,” “yearly,” etc. |

## Defining events

Each trigger span defines a new event. In most cases, the description of a determinant will be confined to a single sentence, resulting in a single trigger span and event for that determinant. A given event should only span across multiple sentences, if a sentence contains relevant event information and does not include a relevant trigger. Figure 12 presents an annotation example where a tobacco event must continue from the first sentence to second sentence, as the second sentence does not have a suitable tobacco trigger span. In other words, the second sentence in Figure 12, by itself, lacks sufficient information to know tobacco use is being described. Figure 13 presents an annotation example where alcohol use is described in two consecutive sentences; however, there is a suitable alcohol trigger in both sentences, so separate alcohol events should be created.

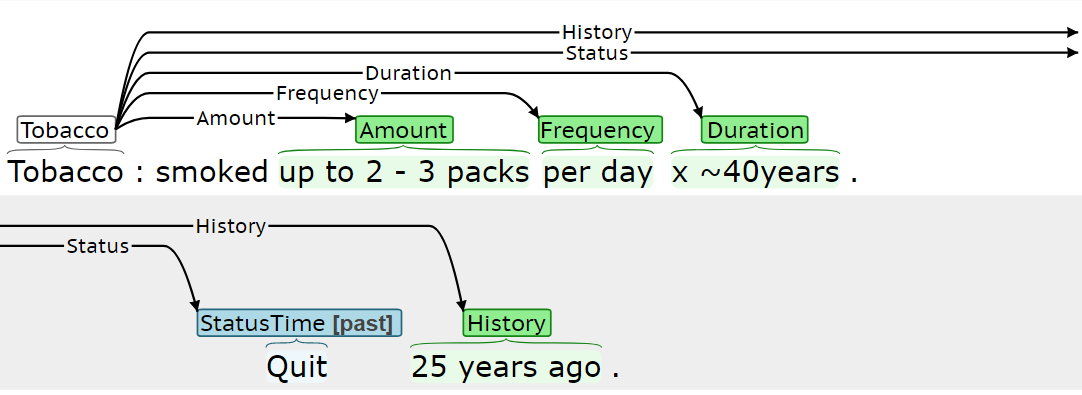


Figure 12. Example of multi-sentence Tobacco event

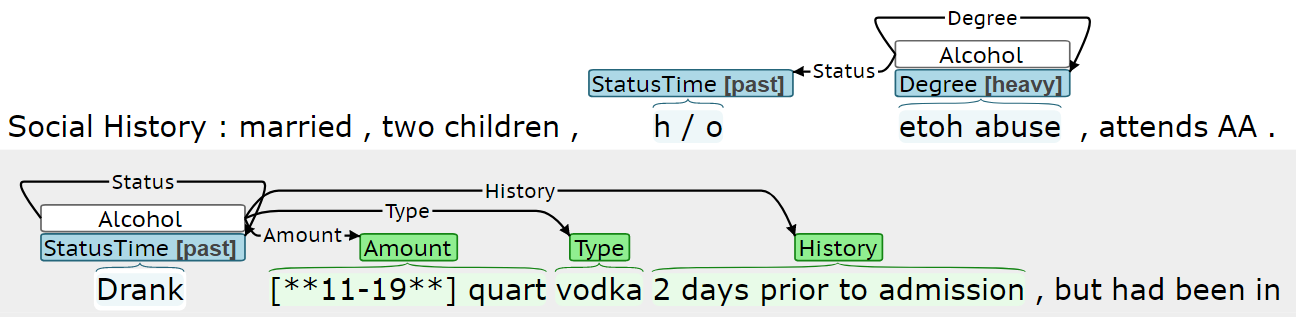


Figure 13. Example of multiple Alcohol events in consecutive sentences

In some cases, multiple events for a given determinant may need to be created within a single sentence. Figure 14 presents an annotation example where a patient’s current and previous tobacco usage is described. The *amount* is described for both the current and historical usage, and creating separate events for the current and previous usage allows this *amount* information to be better characterized.

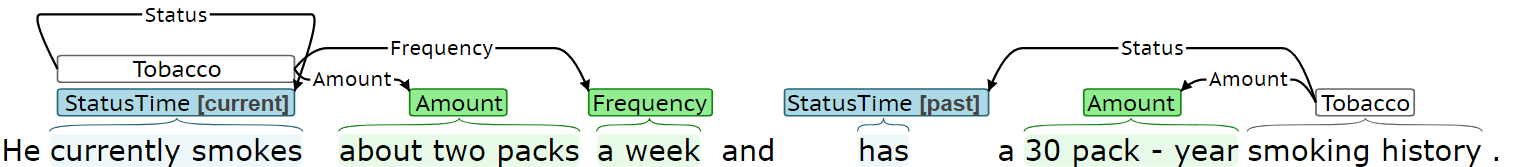


Figure 14. Example of multiple tobacco events within the same sentence