Guidelines for onset mention annotation

Written by Natalia Viani, Sumithra Velupillai

Data

In this annotation task, we will consider patients with a diagnosis of schizophrenia. The documents have been filtered by requiring at least 1 symptom keyword and at least 5 time expressions.

Annotation task

The dataset consists of batches of 10 patients each. For each patient, multiple first referral documents are provided.

Your task will be to identify psychosis onset information in these documents (**ONSET_MENTION** annotation). Specifically, we are interested in identifying the *earliest* reference to psychosis, for the ultimate goal of being able to calculate the duration of untreated psychosis (DUP).

We will consider three types of ONSET_MENTION annotations, captured by the **type** attribute (in decreasing order of specificity):
1. *psychotic_symptoms*
2. *diagnosis* (specific to psychosis)
3. *non_specific* (e.g. first admission to mental health services, not specific to psychosis)

For each ONSET_MENTION annotation, you should fill in the **time_information** attribute with one of the following three values:
- *Past_anchored*: when the mention is clearly linked to a time in the past (e.g., "she started hearing voices at the age of 8")
- *Past_not_anchored*: when the mention is not clearly linked to a time (e.g., "history of delusional beliefs")
- *Current*: when the mention is current (e.g., "Pat. Not known to this service presenting with auditory hallucinations. This is clearly a first psychosis episode").

Important: If the onset is within a month or the current presentation, then we can consider it to be CURRENT. As a general rule, "current" annotations are likely to be made when a document doesn't include any other reference to past onset (i.e. as an evidence of first episode psychosis).
Examples:
- "in the last two days" --> Current
- "Friday evening" --> Current
- "13 year old patient with first psychosis episode" --> Current
- "worsening over the past year"--> Past_anchored

When the annotation is Past_anchored and the reference date is clearly written in the text, you should write this in the **value** attribute (if the reference to the date is relative, e.g., "two days before", please calculate the actual value):

- "at the age of 8", value = AGE8Y.
- "in 2012", value = 2012.
- "two years ago", value = 2010 (if the document date has year 2012)
- "in October 2011", value = 2011-10
- "between 2009 and 2012", value = [2009,2012]

Note 1: Symptom onset information could be found in sentences ("he first experienced hallucinations in 1999") as well as in longer text spans (e.g., a whole paragraph). It is important that you mark the entire textual portion that is needed to recover the onset information, i.e. the symptom and the associated date/time (if present).

Note 2: For each document, only one onset mention per type (the *earliest*) should be annotated, if present in the text. It is important to note that there might not be onset information at all.

Note 3: To determine what constitutes symptom onset information of type *psychotic_symptoms*, the following non-exhaustive list of symptoms can be considered as a guide:

1. hallucinations
2. delusions
3. delusional
4. thought disorder
5. paranoia
6. formal thought disorder
7. tangential
8. thought interference
9. flight of ideas
10. persecutory ideas
11. thought block
12. circumstantial
13. tactile hallucinations
14. loosening of associations
15. running commentary
16. floridly psychotic
17. somatic passivity
18. thought alienation
19. somatic hallucinations
20. derailment
21. tangential speech
22. circumstantial speech
23. auditory command hallucinations
24. systematised delusions
25. knights move thinking

26. clang associations
27. visual hallucinations
28. hallucinatory
29. delusions of grandeur
30. thought disordered
31. paranoid thoughts
32. paranoid ideation
33. paranoid thinking
34. paranoid ideas
35. persecutory thoughts
36. persecutory voices
37. persecutory delusions
38. persecutory comments
39. persecutory symptoms
40. thought blockage
41. thought blockages
42. circumstantiality
43. passivity
44. passivity phenomena
45. passivity of thought
46. tangentiality
47. hearing voices
48. auditory hallucinations
49. command hallucinations
50. thought insertion
51. thought withdrawal
52. first rank symptoms
53. psychotic beliefs
54. psychotic symptoms
55. psychotic thought content
56. thought broadcast

Document with no onset mentioned

You can use the **NO_ONSET_MENTIONED** label for those documents where no ONSET_MENTION was found. In these cases, you can mark the first sentence in the document as NO_ONSET_MENTIONED, adding one of the following "types":
- not_psychosis_patient: if you think this is not a psychotic patient
- no_mention_found: if there is not reference at all to the onset, but you think this is a psychotic patient