**ExECT V2.1**

**What and How of annotating with Markup**

These instructions assume some familiarity with Markup and are aimed to aid the annotating of epilepsy clinic letters as part of establishing an example annotation set (Gold standard) for information extraction from epilepsy clinic letters using ExECT. The appendix and lists are linked to help in looking up terms and concept features.

General points

Usually, when a term of interest is highlighted, all possible UMLS options from the loaded dictionary will be shown in the UMLS box, beginning with the best match. The UMLS list includes terms for all concepts: epilepsy, seizure types, AEDs, comorbidities, so care must be taken to select the correct entity from the Markup list. For some concepts, e.g. investigations or birth history, the attribute must be typed in the UMLS search box to extract the appropriate term and CUI.

When a selected concept does not appear in the UMLS drop down list, it can be searched for in the search box using a similar term e.g., Autistic Spectrum Disorder written as an acronym ASD is not on the UMLS list but searching for “autism” will produce the appropriate phrase.

Current UMLS list is based on the UMLS derived gazetteers used in ExECT v2.1, which apart from the epilepsy terms contain a list of other disorders, symptoms, or life events to be captured under the general term, Patient History. This list does not contain all the possible conditions or events that may be important, however, during the validation stage we should only annotate terms with the UMLS match, whilst collecting the terms we feel are important and should be added to the next version of ExECT / Markup.

**Certainty levels** should be given to Diagnosis, When Diagnosed, Onset, Epilepsy Cause, Birth History, and Patient History, in relation to the concept itself not its attributes. We are not allocating Certainty to Seizure Frequency, Prescription, and Investigations.

Polarity (**Negation**) should be assigned to all concepts except Seizure Frequency, Investigations, and Prescription. Missing attributes can be ignored. At present, apart from Febrile Seizures we are annotating only affirmed statements.

**Dates**, when given as attributes to the concepts, are recorded as DayDate, MonthDate, YearDate (in full), using the Markup dropdown list.

**Appendix A** lists the features to be assigned to all annotated concepts (entities).

Diagnosis

Includes Epilepsy, Epilepsy type and syndromes, seizure types. No generic “seizure/absence/myoclonic jerk” should be included. Past and present tense is accepted, but hypothetical statements are not.

The pattern to follow is: Diagnosis trigger or Person term followed by Epilepsy Term or Specific Seizures.

**List 1** gives the terms used as Diagnosis Triggers.

Level of certainty depends on the context, terms affecting the certainty level are provided in **List 2.** If a term expressing doubt is given but it is not on the list, please annotate as it seems appropriate and take a note of the term so it can be added to the list.

Seizures in combined seizure phrases, such as partial seizures with secondary generalisation should be annotated separately as partial seizures and secondary generalisation. But a focal to bilaterally convulsive seizure is just another term for secondary generalised seizure and has just one CUI.

Similarly, in combined epilepsy phrases such as refractory focal epilepsy, refractory epilepsy and focal epilepsy should be annotated separately. Symptomatic focal epilepsy/localisation related symptomatic epilepsy has its own CUI so it should be annotated as one.

Example 1: *He has been diagnosed with focal epilepsy;*Diagnosis = focal epilepsy, Certainty = 5, DiagClass = Epilepsy

Example 2: *She is having possible complex partial seizures;*Diagnosis:complex partial seizures, Certainty = 3, DiagClass = MultipleSeizures

Example 3: *Should her focal seizures continue, we would increase the dose further.*

Diagnosis:focal seizures, Certainty = 5, DiagClass = MultipleSeizures

We know that seizures are happening so it has to be 5 although “should” is a hypothetical trigger and GATE may have a problem with this. It would be different if instead of “continue” there were “return”, this would suggest that seizures are not happening now, and we would not annotate as the statement is hypothetical.

Example 4: *He has not had a generalised tonic clonic seizure for a while.*

Diagnosis: generalised tonic clonic seizure, DiagCategory = single seizure, Certainty = 5, Negation = Affirmed

Although he is not having seizures now, he “usually” does, and even if the sentence seems to be a negation it states that he has had gtcs. (We would annotate the same sentence for seizure frequency as 0 with no time period or point in time)

**Common errors:**

Example 1: *We discussed driving regulations relating to epilepsy*.  
Epilepsy is “hypothetical” in this context, but also there are no diagnosis triggers for this concept to be annotated.

Example 2: *She presents with myoclonic jerks and absences.*Because myoclonic jerks and absences are too generic, we would not annotate them in Diagnosis but under Patient History. Myoclonic seizures and absence seizures, however, would be annotated in Diagnosis. Missing CUIs, partial annotation of combined seizures, mistakes in certainty levels – please remember.

Onset

Only specific seizure types or epilepsy should be annotated. There are clear onset triggers to look out for: began, started, first occurred, onset was, **followed or preceded by** age / time since / date or other point in time.

**List 3** gives the terms used to capture onset, but they all need to be associated with age or point in time.

**List 4** gives points in time that are often used in clinic letters instead of dates.

Features to be assigned are age, years since, date, or point in time i.e., last year, whichever is given in the text.

Example 1: *Julie has been suffering from epilepsy since 1998;* Onset = Epilepsy, YearDate = 1998

Example 2: *Mary had her first tonic clonic seizure while on holiday in Spain last year;*  
Onset = tonic clonic seizure, PointInTime = last year

Example 3: *John’s complex partial seizures started when he was a teenager;*Onset: complex partial seizures, AgeUnit = Year, AgeLower = 12, AgeUpper = 19

Example 4: *She has been having frequent complex partial seizures for the last year*.

There is no onset information here only frequency, we do not know when her seizures started only that they were frequent for the last year.

**List 5** contains age/age groups to be used when age is not expressed in numbers. These groups also apply to WhenDiagnosed and PatientHistory. When age range is given in years and months, it should be converted to months.

Example 4: *he started having generalised seizures between the age of 1 year and 18 months;*Onset: generalised seizures, AgeUnit = Month, AgeLower = 12, AgeUpper = 18

Words such as increased, continue, changed, returned, are a clear indication of continuation rather than an onset, so events in that context **should not** be annotated.

Common mistakes

Onset of generic seizures such as absences, myoclonic jerks, convulsions will be captured by Patient History so they should not be annotated in Onset.

Onset should not be confused with When Diagnosed, as this clearly illustrates:  
 *He was diagnosed with epilepsy in 2017 but he had his first tonic clonic seizure in his teens.*Tonic clonic seizures should be annotated as onset, with age as in Example 2 above.

When Diagnosed

Only specific seizure types, and epilepsy / epilepsy syndromes should be annotated in phrases that clearly state that the diagnosis was made. Triggers are: Diagnosed / Diagnosis **followed or preceded by** age / time since / date.

Features to be assigned are the same as for onset.

The example given in Common Mistakes under Onset is valid here, there is a clear distinction between When Diagnosed and Onset.

Patient History

Any other significant diagnoses, comorbidities, accidents, non-specific seizures / seizure-like events, and specific abnormalities identified on neuroimaging should be annotated here. At present we are limiting the concepts to those that can be matched with the UMLS based dictionary, as described in the General Points above.

Diagnosis Triggers (**List 1**), Onset phrases (**List 3**), but also Medical History (**List 6**) and Opinion phrases (**List 7**) are all used in the ExECT rules, so they are included here as a guide.

Concepts to be annotated may be listed as other diagnoses at the top of a clinic letter, as background history, may be mentioned as a list of past events and disorders, or could be presented within the concluding comments / opinion.

When date, age, or time since onset of non-specific seizures, events or other diagnoses are provided these should be annotated as features. These terms should not be used in seizure frequency.

There may be some overlap with seizure frequency when annotating generic seizures within Patient History. The general rule is when a person trigger is present i.e., *her seizures,* this should be annotated; but *seizure frequency* on its own should be ignored i.e., *he had no seizures since, or his seizure frequency is,*

At present we are not annotating concepts reported from examination and those given in seizure description – seizure semiology.

Attributes (features) to be assigned, if present, are Age, Time Since, Date, or PointInTime.

Example 1: *John suffered a severe head injury due to an RTA in 2010.*Two concepts should be annotated separately here, *Head Injury* and *RTA,* with Certainty of 5 and YearDate of 2010

Example 2: *For the last 2 months he has been having episodes of myoclonic jerks during the day and at night in addition to some absences. Videotelemetry did not show any EEG correlate. Our impression is that these episodes are not epileptic in nature.*

Myoclonic jerks and absences should be annotated separately with TimePeriod = Month, NumberOfTimePeriods = 2;

“not epileptic in nature” should not be annotated as non-epileptic seizures as they are too general, so they must be ignored.

For febrile seizures we want to extract both, Affirmed and Negated statements. Negated statements are assigned a Certainty of 1 and are Negated. Febrile seizures may be given with age, and here we try to use a range if a number of ages are given.

Example 3: *There is no history of febrile convulsions, head injury or meningitis.*  
All three concepts are negated, so the only annotation here would be:  
Febrile convulsions, Negation = Negated, Certainty = 1

Example 4: *She had febrile seizure at the age of 3 and 5.* Febrile seizure, LowerAge = 3, UpperAge = 5, AgeUnit = Year – it is assumed that when no time unit is given, “year” is implied.

Common mistakes

Annotating concepts that belong to diagnosis (specific seizures), missing generic seizures such as absences. Annotating phrases that refer to seizure description, or neurological examination.

Not annotating abnormalities reported in neuroimaging.

Epilepsy Cause

These are events in the patient’s history that are stated to be a cause of epileptic seizures or epilepsy. They are identified by “Causality phrases” (**List 8**) such as: related to, due to, caused by… followed by a specific event, disease, or a brain abnormality.

UMLS concepts and CUIs relate to the CAUSE and this must be highlighted by the annotator, not the Epilepsy Term.

Events that may trigger seizures, such as alcohol intake, drug abuse, or medication should not be annotated here.

Example 1: *Her epilepsy clearly relates to the severe head injury suffered in 2005;*Epilepsy Cause: severe head injury, the year does not need to be annotated as this should be done in Patient History

Example 2: *Her epilepsy started following a fall she sustained on holiday last year.*The fall is not stated to be a cause of epilepsy here, it preceded the onset, but the association is not stated strongly enough to be read as a cause, so it should not be annotated.

Birth History

Injuries sustained during or before the time of birth should be annotated here. These have their own CUIs, so it is important to highlight a whole phrase while annotating for the UMLS match to appear.

Normal birth should be annotated, but normal delivery at this stage should not. We are revising the CUI list and may add terms relating to delivery.

Premature Birth – gestational age (grouped) in weeks corresponds to specific UMLS terms / CUIs. A Drop-down in MarkUp would give the appropriate term to search for in the UMLS search box. Levels of prematurity expressed in the number of weeks from full term should be converted to gestational age in weeks.

Example 1: *John was born 8 weeks prematurely;*BirthHistory: Premature Birth = 32to<37\_ModerateToLatePreterm, Certainty = 5, …  
The term “moderate to late preterm” should be searched for in the UMLS search box.

Investigations

EEG/CT/MRI followed by an abnormal/normal result. A list of phrases indicating investigation results is given in **List 9,**  separately for EEG and CT/MRI. The list of abnormalities may not be complete and, as with other diagnoses (Patient History), we would like the annotators to annotate all terms that can be considered as abnormalities and collect a list of terms so that they can be added to ExECT. However, during **the validation process** only the terms that can be matched with an item on the UMLS drop down list should be annotated.

To find the UMLS match, a test type, and normal/abnormal must be entered in the search box.

For EEG, the type of test needs to be annotated, **if stated** (it should not be assumed), otherwise it can be ignored.

When the results are **stated to be unknown** – annotate with a CUI for the test itself i.e., *EEG, MRI, or a CT* searching for CT/MRI/EEG unknown in the UMLS search box.

Investigations without any mentions of result should be ignored.

Example 1: *I reviewed this patient’s EEG along with our Chief EEG technician. This was a routine EEG examination prior to videotelemetry. EEG is normal during the attacks…*

Only the last mention of EEG should be annotated as: EEG normal, EEG type is not mentioned in the phrase (it is in a previous sentence), so it must be ignored. No other EEG mentions should be annotated.

Example 2: *A CT scan of the brain showed bleeds on both sides of the brain.* CT abnormal.

Common mistakes

Annotating Investigations without any mentions of results and assuming the result. Missing UMLS CUIs, assigning certainty and negation.

Prescription

We are only annotating current prescriptions for antiepileptic drugs (AED) as: Drug name, Dose (quantity), Dose Unit (measurement in mg or g), and frequency. In Markup drugs are listed under their generic names; a list of AEDs as generic and brand names is given in **List 10**

In the UMLS dropdown list they are shown under their generic and brand names, it is important to match the names precisely, (without substituting the brand for a generic name as it is done in the Markup attributes’ dropdown).

Drugs without a dose, should not be annotated, except for rescue medications such as midazolam or diazepam, for which frequency may be annotated with “As required”.

If frequency is NOT stated use once a day, or ‘As Required’ for Clobazam (and the rescue drugs).

Example 1: “*she is also prescribed buccal midazolam”;  
AED:* Midazolam, Frequency: As required, with other attributes being ignored.

Example 2: *We suggest that he continues levetiracetam with the same dose.*  
Although the dose is known from the letter, here it is not stated so the drug should not be annotated.

When a drug highlighted in the text is not included in the Markup list it should be selected directly from the UMLS match, however the dose must be entered in the Markup attributes section.

***We plan*****to extract when drug is taken (i.e. AM/PM if mentioned)**

Seizure Frequency

Seizures, specific seizures, absences, and myoclonic jerks are to be annotated. Events, episodes, or other slang terms should not.

Seizure frequency relates to the current seizure experience described as a number of seizures or seizure frequency change (increase, decrease, same, etc.) during a defined time period or since a specific point in time.

Time\_Since or Time\_of\_Event attribute should be used only when a date or point in time are stated in order to clearly specify whether the seizures occurred since or during the stated time (date, month).

Example 1: *He had 5 seizures in May, but none since.* Two sets of annotations may be generated here.  
1. Seizure: NumberOfSeizures = 5, MonthDate = 5, TimeSince\_or\_TimeOfEvent = During.  
2. Seizure: NumberOfSeizures = 0, MonthDate = May, TimeSince\_or\_TimeOfEvent = Since.

Example 2: *Seizure free for the last 2 months. Her last episode was in August;* Here we would annotate -  
Seizure free: NumberOfSeizures = 0, TimePeriod = month, NumberOfTimePeriods =2,  
but not episodes (Slang) - 0 since Aug.

Example 3: *His last generalised seizure was 5 years ago.*  
Generalised seizure: NumberOfSeizures = 0, TimePeriod = Year, NumberOfTimePeriods = 5, TimeSince\_or\_TimeOfEvent – should be ignored, as this is used only with a date / point in time.

Example 4: *Since starting Lamotrigine his seizure frequency has improved*.  
Seizure: FrequencyChange = Decreased, TimeSince\_or\_TimeOfEvent = Since, PointInTime = DrugChange

If multiple time periods are used, as in Example 5; annotate **both**.

Example 5: *Since last being seen, she had two seizures in March*.  
1. Seizures: NumberOfSeizures = 2, PointInTime = LastClinic, TimeSince\_or\_TimeOfEvent = Since;  
2. Seizures: NumberOfSeizures = 2, MonthDate = 3, TimeSince\_or\_TimeOfEvent = During.

Example 6 – *Her last seizure was in September 2012.*  
Seizure: NumberOfSeizures = 0, MonthDate = 9, YearDate = 2012, TimeSince\_or\_TimeOfEvent = Since

Although ‘in’ would imply during, since this is an indication of no events since this date (and not that the patient was seizure free for a single month in 2012) we use Since as the TimeSince\_or\_TimeOfEvent, not During.

No seizure since = 0 seizures Since, Last seizure in / time period = 0 seizures Since Time Period

**List 11** gives word numbers that may be used in seizure frequency statements.

Common mistakes

Annotating past seizure control, change, or individual seizure event without a statement of frequency

*“…she was placed on Tegretol which in fact controlled her seizures very well”.*

*“Seizures recurred in July 2013 …She pulled over and went on to have a complex partial seizure.”*

**Lists of terms**

**List 1: Diagnosis Triggers**

|  |  |
| --- | --- |
| Diagnosis | problem, problems |
| Diagnosed | very suggestive of |
| Suffers, suffering | would be consistent with |
| in keeping with | history is suggestive of |
| seizure type, seizure types | possibility of |
| seizure type and frequency | symptoms are suggestive of |
| story is consistent with | my impression is |
| history is consistent with | we are dealing with |

**List 2: Certainty (Probability) Levels**

|  |  |
| --- | --- |
| ruled out Level=1 | to see whether Level=3 |
| doubt Level=2 | to be confirmed Level=3 |
| improbable Level=2 | to know whether Level=3 |
| not convincingly Level=2 |  |
| remote Level=2 |  |
| unclear Level=2 | not conclusive Level=4 |
| unsure Level=2 | suspicious Level=4 |
| ?? Level=2 | suspect Level=4 |
| doubtful Level=2 | suggestive Level=4 |
| not convinced Level=2 | sound like Level=4 |
| not likely Level=2 | supports Level=4 |
| remote possibility Level=2 | suspected Level=4 |
| unlikely Level=2 | suspicion Level=4 |
| unusual Level=2 | I think Level=4 |
|  | is in keeping with Level=4 |
|  | point more towards Level=4 |
|  | probable Level=4 |
|  | compatible with Level=4 |
| considered Level=3 | impression is Level=4 |
| describes himself Level=3 | likely Level=4 |
| ? Level=3 | point towards Level=4 |
| could be Level=3 | probably Level=4 |
| further clarification Level=3 | supportive of Level=4 |
| investigate her along the lines Level=3 | treated as Level=4 |
| markers Level=3 |  |
| might Level=3 |  |
| possible Level=3 | consistent with Level=5 |
| possibility Level=3 | is conclusive Level=5 |
| potentially Level=3 | are dealing with Level=5 |
| potential Level=3 | certain Level=5 |
| to be sure Level=3 | definite Level=5 |
| to see if Level=3 | in keeping with Level=5 |
| uncertain Level=3 |  |
| further investigation Level=3 |  |

**List 3: Onset Terms and Phrases**

|  |  |
| --- | --- |
| started at age / date/ time since | suffering |
| occurred | has been symptomatic for x number of years |
| appeared | suffering from for …time period |
| manifested... | history is of ... from the age ... |
| new | report him having …age / since date |
| began | background is ...age / time since |
| onset | presented with |
| suffered…first | was noted to have |
| began to experience | found to have |
| began to develop | describes…from age |
| first |  |

**List 4: Points in Time**

|  |  |
| --- | --- |
| This Year | Birthday |
| Last Year | Easter |
| Last Clinic | 1960s |
| Drug Change | 1970s |
| From Birth | 1980s |
| Surgery | 1990s |
| Discharge Date | 2000s |
| Last Christmas | 2010s |

**List 5: Person’s Age**

|  |  |  |  |
| --- | --- | --- | --- |
| **Age** | **LOWER Age** | **UPPER Age** | **Age Unit** |
| a levels | 17 | 18 | Year |
| adolescence | 10 | 19 | Year |
| adolescent | 10 | 19 | Year |
| a'levels | 17 | 18 | Year |
| baby | 0 | 12 | Month |
| child | 2 | 12 | Year |
| childhood | 2 | 12 | Year |
| early adolescent | 10 | 14 | Year |
| early childhood | 1 | 6 | Year |
| early teenage | 13 | 14 | Year |
| early teens | 13 | 14 | Year |
| early years | 1 | 6 | Year |
| gces | 17 | 18 | Year |
| gcse | 15 | 16 | Year |
| gcse's | 15 | 16 | Year |
| infant | 0 | 12 | Month |
| late teenage | 17 | 19 | Year |
| mid teenage | 15 | 16 | Year |
| mid teens | 15 | 16 | Year |
| middle age | 45 | 65 | Year |
| neonatal period | 0 | 28 | Day |
| neonate | 0 | 28 | Day |
| primary school | 5 | 11 | Year |
| puberty | 10 | 17 | Year |
| secondary school | 12 | 16 | Year |
| teenager | 13 | 19 | Year |
| teens | 13 | 19 | Year |
| toddler | 1 | 3 | Year |
| young | 13 | 19 | Year |
| young child | 1 | 6 | Year |
|  | | | |
|  |  | **Age** |  |
| one |  | 12 | Month |
| one year |  | 12 | Month |
| year and a half |  | 18 | Month |
| one and a half |  | 18 | Month |

**For ages described in terms of decades, e.g. fifties, age should be annotated as AgeRange, here 50 - 59. When a part of a decade is given, such as early, mid, or late the ranges should follow a pattern of early = 0 – 3, mid = 4 – 6, and late = 7 – 9 years added to the number of decades.**

**For the fifth decade the following age ranges will be produced:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Age** | **LOWER Age** | **UPPER Age** | **Age Unit** |
| fifties | 50 | 59 | Year |
| early fifties | 50 | 53 | Year |
| mid fifties | 54 | 56 | Year |
| late fifties | 57 | 59 | Year |

**This pattern should be used for all the decade-based ages.**

**List 6: Medical History**

|  |  |
| --- | --- |
| past medical history | was under the care |
| past medical history of | known to suffer |
| past history of | comorbidities |
| used to suffer | labelled |
| background | on the record as |

**List 7: Opinion Triggers**

|  |  |
| --- | --- |
| history suggests | point towards |
| case of | I think |
| history is suggestive | I am of the opinion |
| history is consistent | conclusion |
| impression is | likely explanation |
| suggestive | would seem |
| opinion is | seems |
| description is consistent | evidence of |
| does have |  |

**List 8: Causality Phrases – to be used when identifying epilepsy cause.**

|  |  |
| --- | --- |
| due to | associated with |
| caused by | left her/him with |
| related to | resulting from |
| subsequent to | resulted in |
| result of | effect of |
| secondary to |  |

**List 9: Investigation Results**

**EEG Results**

|  |  |
| --- | --- |
| **Phrase** | **Annotate as** |
| abnormal | Results=Abnormal |
| Abnormal | Results=Abnormal |
| abnormalities | Results=Abnormal |
| abnormality | Results=Abnormal |
| bilateral discharges | Results=Abnormal |
| both normal | Results=Normal |
| burst suppression | Results=Abnormal |
| clear | Results=Normal |
| did not capture any events | Results=Normal |
| dysrhythmic | Results=Abnormal |
| epileptic | Results=Abnormal |
| epileptic activity was not seen | Results=Normal |
| epileptiform | Results=Abnormal |
| epileptogenic | Results=Abnormal |
| failed to alter | Results=Normal |
| focal discharge | Results=Abnormal |
| focal ictal rhythms | Results=Abnormal |
| focal slowing | Results=Abnormal |
| focus | Results=Abnormal |
| generalised discharges | Results=Abnormal |
| generalised slowing | Results=Abnormal |
| hypsarrhythmia | Results=Abnormal |
| irregular | Results=Abnormal |
| left side slowing | Results=Abnormal |
| left sided changes | Results=Abnormal |
| localised discharge | Results=Abnormal |
| localised discharges | Results=Abnormal |
| localised repetitive discharges | Results=Abnormal |
| low amplitude fast activity | Results=Abnormal |
| low voltage fast activity | Results=Abnormal |
| multifocal discharges | Results=Abnormal |
| no changes | Results=Normal |
| no significant findings | Results=Normal |
| non-epileptic | Results=Normal |
| non-specific interictal changes | Results=Abnormal |
| normal | Results=Normal |
| Normal | Results=Normal |
| not available | Result=Unknown |
| not have the results | Result=Unknown |
| paroxysmal fast activity | Results=Abnormal |
| photoparoxysmal response | Results=Abnormal |
| photosensitive | Results=Abnormal |
| photosensitivity | Results=Abnormal |
| polyspike | Results=Abnormal |
| poly-spike | Results=Abnormal |
| polyspike and wave | Results=Abnormal |
| polyspike-and-wave | Results=Abnormal |
| right side slowing | Results=Abnormal |
| right sided changes | Results=Abnormal |
| sharp | Results=Abnormal |
| slow spike and wave | Results=Abnormal |
| slow spike-wave discharges | Results=Abnormal |
| slow wave | Results=Abnormal |
| spike | Results=Abnormal |
| spike and wave | Results=Abnormal |
| spike wave discharges | Results=Abnormal |
| spikes | Results=Abnormal |
| spike-wave | Results=Abnormal |
| temporal intermittent rhythmic delta activity | Results=Abnormal |
| temporal slowing | Results=Abnormal |
| unremarkable | Results=Normal |
| unstable | Results=Abnormal |

**MRI / CT Results**

|  |  |
| --- | --- |
| **Phrase** | **Annotate** |
| abnormal | Results=Abnormal |
| Abnormal | Results=Abnormal |
| abnormal signal | Results=Abnormal |
| abnormalities | Results=Abnormal |
| abnormality | Results=Abnormal |
| astrocytoma | Results=Abnormal |
| atrophy | Results=Abnormal |
| atrophic changes | Results=Abnormal |
| AVM | Results=Abnormal |
| both normal | Results=Normal |
| brain asymmetry | Results=Abnormal |
| cavernoma | Results=Abnormal |
| cerebral artery occlusion | Results=Abnormal |
| cerebral oedema | Results=Abnormal |
| cerebral ischaemia | Results=Abnormal |
| clear | Results=Normal |
| cortical dysplasia | Results=Abnormal |
| CVA | Results=Abnormal |
| degeneration | Results=Abnormal |
| degenerative brain disorder | Results=Abnormal |
| DNET | Results=Abnormal |
| encephalomalacia | Results=Abnormal |
| glioma | Results=Abnormal |
| gliosis | Results=Abnormal |
| haemangioma | Results=Abnormal |
| haemorrhage | Results=Abnormal |
| heterotopic grey matter | Results=Abnormal |
| Heterotopic grey matter | Results=Abnormal |
| high intensity signal | Results=Abnormal |
| lesion | Results=Abnormal |
| lesions | Results=Abnormal |
| malformation | Results=Abnormal |
| malformations | Results=Abnormal |
| mass effect | Results=Abnormal |
| no significant findings | Results=Normal |
| non-specific lesion | Results=Abnormal |
| normal | Results=Normal |
| Normal | Results=Normal |
| not available | Result=Unknown |
| not have the results | Result=Unknown |
| sclerosis | Results=Abnormal |
| signal abnormality | Results=Abnormal |
| signal intensity | Results=Abnormal |
| Tumour / tumor | Results=Abnormal |
| unremarkable | Results=Normal |
| white matter changes | Results=Abnormal |
| white-matter hyperintensities | Results=Abnormal |

**List 10: AED**

|  |  |
| --- | --- |
| **Generic** | **Brand** |
| Acetazolamide |  |
| Carbamazepine | Tegretol, Tegretol PR, Tegretol Retard |
| Clobazam | Frisium, Perizam |
| Clonazepam |  |
| Eslicarbazepine Acetate | Zebinix |
| Ethosuximide | Zarontin |
| Gabapentin | Neurontin |
| Lacosamide | Vimpat |
| Lamotrigine | Lamictal |
| Levetiracetam | Keppra, Desitrend |
| Nitrazepam |  |
| Oxcarbazepine | Trileptal |
| Perampanel | Fycompa |
| Phenobarbital |  |
| Phenytoin | Epanutin |
| Piracetam |  |
| Pregabalin |  |
| Primidone |  |
| Retigabine |  |
| Rufinamide | Inovelon |
| Sodium Valproate | Epilim, Epilim Chrono, Episenta, SV could also be given as Valproic Acid |
| Stiripentol |  |
| Tiagabine | Gabitril |
| Topiramate | Topamax |
| Vigabatrin | Sabril |
| Zonisamide | Zonegran |

**List 11: Word Numbers**

|  |
| --- |
| single: value=1 |
| a couple: value=2 |
| a few: value=2 |
| once: value=1 |
| none: value=0 |
| a number: value=2 |
| multiple: value=2 |

**Appendix A**

**ExECT V2.1 What and How of annotating with Markup – assigning features**

**Assigning features to concepts in Markup – general points**

For each concept of interest there is a set of features (attributes) that should to be assigned during the annotation process. All possible features are shown once a word / phrase of interest is highlighted and assigned to a concept, for example, highlighting “born at term” and clicking on BirthHistory will give a list of possible features to be assigned from the dropdown lists. A phrase may be assigned to more than one concept, depending on the context provided. All possible “contexts” should be annotated during the process. For instance, in a sentence: “This lady has been suffering from epilepsy for the last 20 years”, the term “epilepsy” should be assigned to “Diagnosis” and to “Onset”, and for each of these a different set of features will be given.

**Features for Diagnosis**

These relate to epilepsy, epilepsy syndrome, or specific seizure types, and apart from Certainty and Negation the phrases should be annotated with an UMLS concept and a diagnostic category:

**DiagCategory:** Epilepsy, SingleSeizure, MultipleSeizures - to annotate whether the statement relates to epilepsy (including epilepsy syndrome), single epileptic seizure, or multiple epileptic seizures.

**Features for Onset, When Diagnosed, Patient History**

Apart from Certainty and Negation the features are:

When time of onset, epilepsy diagnosis, or an event of interest (for Patient history) is given as a **patient’s age** (numeric value or age group).

**AgeUnit**: Week, Month, Year;

**Age**: Number;

**AgeLower**: number – when age is expressed as an age group such as “teenager” or “from 3 to 5 years”, the lower value – a list of age groups with the lower / higher value is attached;

**AgeUpper**: number, as above for the higher value;

When time of onset, epilepsy diagnosis, or an event of interest (for Patient history) is given as the **time since** the event occurred, given precisely or as a range.

**TimePeriod**: Week, Month, Year;

**NumberOfTimePeriods:** number – how many weeks, months, or years;

**LowerNumberOfTimePeriods**: number – when the time period is given as a range e.g. 4 to 5 years ago, this is the lower number;

**UpperNumberOfTimePeriods**: number, as above, but this is the higher number;

When time of onset, epilepsy diagnosis, or an event of interest (for Patient history) is given as a **date** (complete or partial.

**DayDate**: number from 1 to 31

**MonthDate**: number from 1 to 12

**YearDate**: year as 4 digits

When time of onset, epilepsy diagnosis, or an event of interest (for Patient history) is given as a **point in time** or a decade. Point in time is a specified day or period but not described as a date in the text, such as birthday, last clinic, last Christmas, which later may be linked to proper dates and allow for creation of a timeline.

**PointInTime:** This\_Year, Last\_Year, LastClinic, DrugChange, From\_Birth, Surgery, DischargeDate, LastChristmas, Birthday, Easter, 1960s, 1970s, 1980s, 1990s, 2000s, 2010s

**Features for Birth History**

Apart from Certainty and Negation the events that clearly occur during birth such as birth injuries should be annotated with an UMLS concept, whereas premature birth should have additional features of:

**PrematureBirth:** Value:37+isTerm\_Birth, under37isPreterm\_Birth, 34to37isLate\_Preterm\_Birth, 32to37isModerate\_To\_LatePreterm, 28to31isVery\_Preterm, under28isExtreamelyPreterm

The UMLS matches do not appear directly so the correct term, without the numbers, must be entered into the search box e.g., Very\_Preterm

**Features for Investigations**

Investigation results relate to MRI, CT, and EEG. Results can be annotated as normal, abnormal, and unknown, with the EEG also annotated with type, if type is not stated, leave blank.

MRI\_Performed: Yes, No, Notknown

MRI\_Results: Normal, Abnormal, Unknown

CT\_Performed: Yes, No, Notknown

CT\_Results: Normal, Abnormal, Unknown

EEG\_Performed: Yes, No, Notknown

EEG\_Results Arg: Normal, Abnormal, Unknown

EEG\_Type Arg: SleepDeprived, VideoTelemetry, Standard, Ambulatory, Prolonged

**Features for Prescriptions**

Only AEDs are to be annotated, a list of generic drugs is shown in the drop-down list and when a brand name is given in the text it should be matched with a generic term from the list.

DrugName: Value: Acetazolamide, Carbamazepine, Clobazam, Clonazepam, EslicarbazepineAcetate, Ethosuximide, Gabapentin, Lacosamide, Lamotrigine, Levetiracetam, Nitrazepam, Oxcarbazepine, Perampanel, Piracetam, Phenobarbital, Phenytoin, Pregabalin, Primidone, Retigabine, Rufinamide, Sodium Valproate, Stiripentol*,* Tiagabine, Topiramate, Vigabatrin, Zonisamide.

DrugDose: number - quantity as stated

DoseUnit: mg, g - measure

Frequency: 1, 2, 3, 4, As Required - equivalents of : od, bd, tds, qds, prn (If no frequency used default to 1 (unless midazolam or clobazam, then As Required )

UMLS matches the drug that is annotated and there is no need to search for a generic term. If a drug is not shown in the Markup attributes drop-down list, but there is an UMLS match, this should be assigned, and the quantity and dose should be selected from the attributes drop-down list.

**Features for Seizure Frequency**

**TimePeriod**: Day, Week, Month, Year – time units for which seizure frequency is reported;

**NumberOfTimePeriods**: number of periods for which seizure frequency is reported;

**LowerNumberOfTimePeriods**:number, when time period is described as a range e.g. 2 seizures every 3 to 5 months, this is the lower number;

**UpperNumberOfTimePeriods**: number, as above, this is the higher number;

**FrequencyChange**: Same, Infrequent, Increased, Frequent, Decreased – when seizure frequency is not quantified;

**NumberOfSeizures**: number;

**LowerNumberOfSeizures**: number – when the number of seizures is expressed as a range, e.g. 4 to 10 seizures per day, this in the lower number;

**UpperNumberOfSeizures**: as above, this is the higher number;

**AgeUnit**: Week, Month, Year;

**Age**: number;

**AgeLower**: number – when age is expressed as an age group such as “teenager” or “from 3 to 5 years”, the lower value – a list of age groups List 5;

**AgeUpper**: number, as above, the higher value;

**DayDate**: numbers 1 to 31;

**MonthDate**: numbers 1 to 12;

**YearDate**: 4-digit number;

**TimeSince\_or\_TimeOfEvent**:Since or During

**PointInTime**: This\_Year, Last\_Year, LastClinic, DrugChange, From\_Birth, Surgery, DischargeDate, LastChristmas, Birthday, Easter, 1960s, 1970s, 1980s, 1990s, 2000s, 2010s