Appendix C. Annotation Guidelines

Appendix C.1. General principles

Appendix C.1.1. Coordination

In annotating coordination we follow the “standard” LLM practice of exploding the coordination when conjuncts can be easily detected. for instance `renal and cardiac complications` could become

```
renal_complications_33050943_4 renal complications condition
cardiac_complications_33050943_4 cardiac complications condition
```

Appendix C.1.2. DUMMY element

In general, triples are filled with entities which are actually present in the sentence. However, the use of an non-existing DUMMY element, is allowed in cases where the triple could not be appropriately filled. For instance:

```
Gradual baclofen and gabapentin administration was prescribed,..
DUMMY_33964989_2 baclofen_33964989_2 treated_by 33964989_2_0
DUMMY_33964989_2 gabapentin_33964989_2 treated_by 33964989_2_1
```

A dummy NE is present in the NE table with its own id and its type (mostly `person`)

Appendix C.1.3. No inference

Just describe facts which are explicit in the sentence, not logically inferrable. For instance, the sentence

```
Clinical findings and work-up including cardiac magnetic resonance
imaging (MRI) were highly suggestive of ARVC.
```

is represented as:

```
work-up_38336791_1 cardiac_magnetic_resonance_imaging_(MRI)_38336791_1 consists_of 38336791_1_0
work-up_38336791_1 ARVC_38336791_1 has_result 38336791_1_1
Clinical_findings_38336791_1 ARVC_38336791_1 has_result 38336791_1_2
```

without inferring that cardiac imaging revealed ARVC.

Appendix C.2. Predicates

The meaning of the predicates we use is described, for instance on the NIH site . Here we focuses on special uses for this dataset or newly injected predicates.

Appendix C.2.1. `sameAs`

It is not a UMLS-SN relation. It is used to denote synonym, as in `fluorescein angiography (FANG)`

Appendix C.2.2. `has_features UMLS-SN`

It is a relation which signals the attribution of a property. It could be seen as a fuzzy represantation of a copular construction. In many cases it represents noun phrase internal properties.

Appendix C.2.3. `has_location UMLS-SN(inverse)`

It denotes a spatial relation between an entity and a place. It is typically used to denote the presence of a patient in a medical structure or the location of an intervention or a desease.

Appendix C.2.4. `causesUMLS-SN`

The predicate is used if a treatment or an event is the explicit cause of something else. If the cause-effect link is not explicitely established, rather use `temporally_related_to`

Appendix C.2.5. `has_resultUMLS-SN (inverse)`

The subject is a test, the object is the result. The results of a treatment should be under causes

Appendix C.2.6. `treated_by UMLS-SN`

It refers to any treatment receive by a patient. The subject is generally a human. However we accept the synecdoche useage when the subject is a body part, and also cases when the subject is a desease.

Appendix C.2.7. `measurement_of UMLS-SN`

We follow the tendency of most LLM to consider a measurement as an independent NE, even though it is inserted into a larger Noun Phrase (e.g `elevated lactate of 2.5 mmol/L`). The subject is always the measure (quantity), the object the measured entity. The measure might be just an imprecise string, such as `elevated`.

Appendix C.2.8. `carries_out` *UMLS-SN*

It is our ‘jolly’ predicate type. It is meant to represent any event, state or activity which cannot be represented otherwise. The first argument is normally the surface syntactic subject and the second one (object) an `event`, usually incarnated by a verb. Normally an event has only one `carries_out` relation, linking the subject to the event. In the case in which multiple arguments are *necessary* the relation `carries_out` is repeated as in:

```
Her respiratory status improved and she was weaned off of NPPV after 3 days.

NPPV_33743806_5 weaned_off_33743806_5 carries_out 33743806_5_4
Her_33743806_5 weaned_off_33743806_5 carries_out 33743806_5_0
```

A better formulation (but outside UMLS) could be `has_some_role_in`

Appendix C.2.9. `produces` *UMLS-SN*

It denotes all cases where an actor causes something to come to existence. It is usually reserved to dynamic development of symptoms, conditions, diseases etc.

Appendix C.2.10. `negates`

This predicate is not present in UMLS-SN and it is used to represent negation. It should be noticed that it is a *semantic* negation, not a syntactic one, so it can occur also in cases where, for instance, no explicit negative particle is found. Nevertheless the subject of the predicate is always some negating element. Here are some examples:

```
Hematology workup including genetic testing showed no evidence of coagulopathy.

+----+------------------------------+----------------------------+-------------+--------------+
|    | subjet                       | object                     | type        |      rels_id |
|----+------------------------------+----------------------------+-------------+--------------|
|  0 | Hematology_workup_36357911_6 | coagulopathy_36357911_6    | has_result  | 36357911_6_0 |
|  1 | no_evidence_36357911_6       | coagulopathy_36357911_6    | negates     | 36357911_6_1 |
|  2 | Hematology_workup_36357911_6 | genetic_testing_36357911_6 | consists_of | 36357911_6_2 |
|  3 | genetic_testing_36357911_6   | coagulopathy_36357911_6    | has_result  | 36357911_6_4 |
+----+------------------------------+----------------------------+-------------+--------------+

She reported abdominal pain, pruritus, and boils on her back preventing her from standing upright.
+----+--------------------------------+-----------------------------+--------------+--------------+
|    | subjet                         | object                      | type         |      rels_id |
|----+--------------------------------+-----------------------------+--------------+--------------|
|  0 | She_36447286_1                 | abdominal_pain_36447286_1   | exhibits     | 36447286_1_0 |
|  1 | She_36447286_1                 | pruritus_36447286_1         | exhibits     | 36447286_1_1 |
|  2 | She_36447286_1                 | boils_36447286_1            | exhibits     | 36447286_1_2 |
|  3 | boils_36447286_1               | her_back_36447286_1         | has_location | 36447286_1_3 |
|  4 | She_36447286_1                 | standing_upright_36447286_1 | carries_out  | 36447286_1_4 |
|  5 | preventing_negation_36447286_1 | standing_upright_36447286_1 | negates      | 36447286_1_5 |
+----+--------------------------------+-----------------------------+--------------+--------------+
```

Appendix C.2.11. ***indicatesUMLS-SN***

It is a quite rare predicate where something (analysis, symptom, etc) show the possibility of something else. Not to be confused with the results of a test. For instance:

```
Echocardiograms revealed new aortic regurgitation, indicating "possible endocarditis" per the Modified Duke Criteria.
+----+----------------------------------+-----------------------------------+-------------+--------------+
|    | subjet                           | object                            | type        |      rels_id |
|----+----------------------------------+-----------------------------------+-------------+--------------|
|  0 | Echocardiograms_37194080_3       | aortic_regurgitation_37194080_3   | has_result  | 37194080_3_0 |
|  1 | aortic_regurgitation_37194080_3  | possible_endocarditis_37194080_3  | indicates   | 37194080_3_1 |
|  2 | possible_endocarditis_37194080_3 | Modified_Duke_Criteria_37194080_3 | has_feature | 37194080_3_2 |
+----+----------------------------------+-----------------------------------+-------------+--------------+
```

Appendix C.2.12. ***precedesUMLS-SN***

We deviate from the temporal meaning stated in UMLS and we restrict the predicate to denote everything in the patient history. So it is *not* used to mean temporal precedence in the case of a running case, which is generically annotated as `is_temporally_related`. Example:

```
The granddaughter had a history of recurrent fevers, significant weight loss, and a suppurative skin condition.
+----+--------------------------+---------------------------------------+----------+--------------+
|    | subjet                   | object                                | type     |      rels_id |
|----+--------------------------+---------------------------------------+----------+--------------|
|  0 | granddaughter_34243803_8 | fevers_34243803_8                     | precedes | 34243803_8_0 |
|  1 | granddaughter_34243803_8 | weight_loss_34243803_8                | precedes | 34243803_8_1 |
|  2 | granddaughter_34243803_8 | suppurative_skin_condition_34243803_8 | precedes | 34243803_8_2 |
+----+--------------------------+---------------------------------------+----------+--------------+
```

Appendix C.2.13. ***goal***

It is a non UMLS predicate: it denotes any purpose of an action, attained or not. Both the subject or object are not necessarily events, as the event can be implicit (e.g. administration). Examples:

```
 A 61-year-old Japanese woman with a history of stroke was hospitalized for breast cancer surgery.
+----+---------------------+--------------------------+----------+--------------+
|    | subjet              | object                   | type     |      rels_id |
|----+---------------------+--------------------------+----------+--------------|
|  0 | patient_33863374_0  | stroke_33863374_0        | precedes | 33863374_0_0 |
|  1 | hospital_33863374_0 | breast_cancer_33863374_0 | goal     | 33863374_0_1 |
+----+---------------------+--------------------------+----------+--------------+
```

```
A check-up to look for possible etiologies for coronary artery ectasia was carried out and returned normal.
+----+-----------------------+------------------------------------+------------+--------------+
|    | subjet                | object                             | type       |      rels_id |
|----+-----------------------+------------------------------------+------------+--------------|
|  0 | etiologies_37277850_6 | coronary_artery_ectasia_37277850_6 | causes     | 37277850_6_1 |
|  1 | check-up_37277850_6   | normal_37277850_6                  | has_result | 37277850_6_2 |
|  2 | check-up_37277850_6   | etiologies_37277850_6              | goal       | 37277850_6_3 |
+----+-----------------------+------------------------------------+------------+--------------+
```

Appendix C.2.14. ***diagnoses UMLS-SN***

It refers to the act of ‘assessing that someone is affected by’. On this respect the difference w.r.t. `affected_by`is just modal, in the sense that the latter denotes a medical fact, the former rather a hypothesis. It should be noticed that we use the predicate in a passive way: as the emitter of the diagnosis is usually not known, the first argument (subject) refers to the patient and the second one (object) to the disease. Examples:

```
She was diagnosed with thyroiditis due to the coronavirus disease 2019 vaccine and was treated with propranolol.
+----+---------------------------------------------+------------------------+------------+--------------+
|    | subjet                                      | object                 | type       |      rels_id |
|----+---------------------------------------------+------------------------+------------+--------------|
|  0 | She_38098118_3                              | thyroiditis_38098118_3 | diagnoses  | 38098118_3_0 |
|  1 | coronavirus_disease_2019_vaccine_38098118_3 | thyroiditis_38098118_3 | causes     | 38098118_3_1 |
|  2 | She_38098118_3                              | propranolol_38098118_3 | treated_by | 38098118_3_2 |
+----+---------------------------------------------+------------------------+------------+--------------+
```

Appendix C.2.15. `exhibitsUMLS-SN`

The object is represented by the signs that an entity shows. Typically a patient exhibits symptoms. It should not be confused with `affected_by`, which signals a disease, not a symptom. For instance upon admissions, the patient would exhibit symptoms.

Appendix C.2.16. `affected_by UMLS-SN (invers)`

It is used to indicate the relation between a patient and its disease.

Appendix C.2.17. `temporally_related_to UMLS-SN`

Given the poor coverage of UMLS for temporal relations, we use this to capture all temporal modification. Note that the subject is not necessarily an event, but it could be an impplicit event, such as [the appearance of] a symptom. The object is almost always a time, an event or a procedure. Notice that we include under the NE annotation the whole event and not only the temporal specification. So, for instance the phrase `Over three years after the initial diagnosis` is considered as a single entity.

Appendix C.2.18. `consists_of UMLS-SN`

It is a generic part_of relation which could relate phisical parts (the whole consists_of parts), but also an event with its sub-events.

References

[1] D. Xu, W. Chen, W. Peng, C. Zhang, T. Xu, X. Zhao, X. Wu, Y. Zheng, Y. Wang, E. Chen, Large language models for generative information extraction: A survey (2024). arXiv:2312.17617.
URL https://arxiv.org/abs/2312.17617

[2] M. Agrawal, S. Hegselmann, H. Lang, Y. Kim, D. Sontag, Large language models are few-shot clinical information extractors, in: Y. Goldberg, Z. Kozareva, Y. Zhang (Eds.), Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Association for