Annotation guidelines for pathogenetic mechanism corpus for lung disease, IPF

Entities and cues

Entities and cues will be described below.

General rules: basic rules

For normalization of entities (and event triggers), IDs of the Unified Medical Language System (UMLS) will be adopted. Each UMLS ID has some semantic types, such as "dsyn: Disease and Sundrome" and "gngm: Gene or Genome". Based on such semantic types, entity types, such as Disorder and GGPs, can be determined.
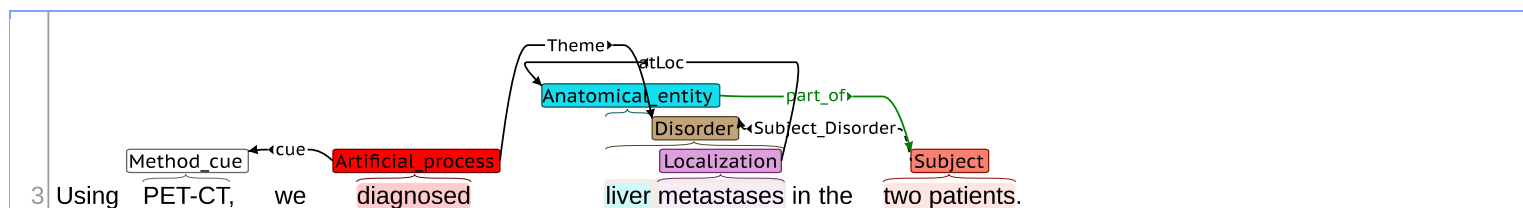
In order to search appropriate IDs, the NCI Metathesaurus (NCIm) can be used.

There are general rules for entities, as follows:

- Articles, such as "a", "an" and "the", are not included in the entities.
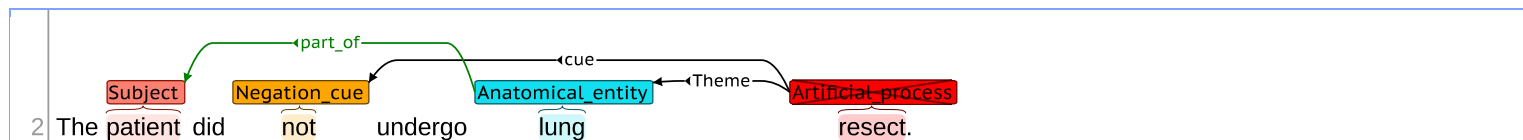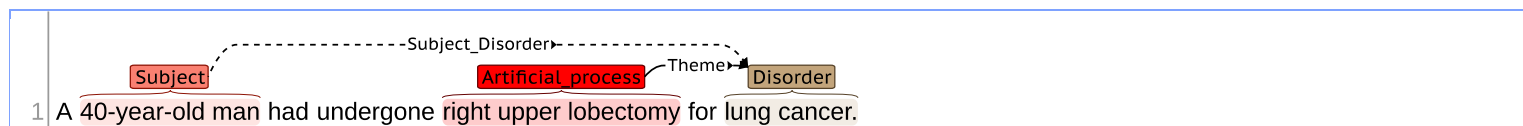
- Numerals should be included in the entities, as they can suggest groups.

Disorder: Disorders

The definition of 'Disorder' is based on the Semantic types of

- 'dsyn' (Disease or Syndrome),
- 'neop' (Neoplastic Process),
- 'patf' (Pathologic Function), and
- 'sosy' (Sign or Symptom)

in the UMLS database.

Diseases and symptoms, which can be a group in phenotype, will be annotated in this category.

4	[Disorder] [Disorder] [Disorder] Idiopathic pulmonary fibrosis (IPF) is a fatal condition.
5	Subject_Disorder [Subject] [Artificial_process] —Theme→ [Disorder] A 40-year-old man had undergone right upper lobectomy for lung cancer.

The following words/phrases should be annotated as 'Disorder'.

- lung cancer
- lung adenocarcinoma
- non-small cell lung cancer
- multiple myeloma
- plasma cell myeloma
- chronic obstructive pulmonary disease
- COPD
- acute exacerbation of chronic obstructive pulmonary disease
- AECOPD
- bronchiolitis obliterans (syndrome)
- idiopathic pulmonary fibrosis
- IPF
- idiopathic interstitial pneumonia
- interstitial lung disease
- pulmonary hypertension
- sarcoidosis
- pulmonary sarcoidosis
- pleural effusion
- lung lesions
- allodynia

Some disease names start with 'AE (acute exacerbation)'. As these words suggest the change in disorder entities, 'Disorder', they may be annotated as event entity, Biological_process:
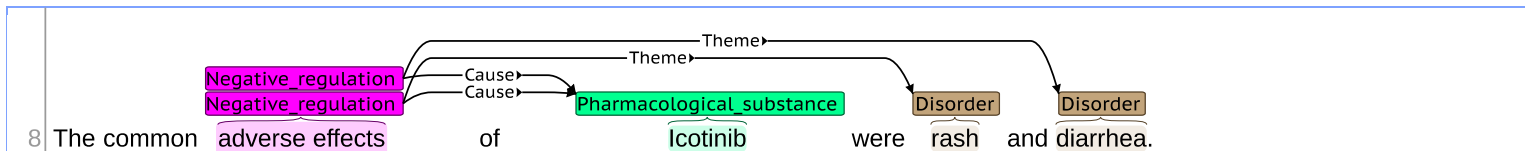- acute exacerbation of idiopathic pulmonary fibrosis
- AE-IPF

6	[Disorder] [Biological_process] [Disorder] [Biological_process] The pathogenesis of acute exacerbations of idiopathic pulmonary fibrosis (AE-IPF)

Symptoms, which can be a group in phenotype, will be annotated in this category. The following words/phrases should be annotated as 'Disorder'.

- breathing disorder
  - sleep breathing disorder
- rash
- diarrhea
- hematologic toxicity

7 The [Regulation **impact**] of [Disorder **IPF-related**] [Disorder **sleep breathing disorders (SBDs)**] on [Biological_process **survival**].
(Regulation, Cause, Cause, Theme, Theme)

8 The common [Negative_regulation **adverse effects**] of [Pharmacological_substance **Icotinib**] were [Disorder **rash**] and [Disorder **diarrhea**].
(Negative_regulation, Cause, Cause, Theme, Theme)

- fatal condition

9 [Disorder **Idiopathic pulmonary fibrosis**] [Disorder **(IPF)**] is a [Disorder **fatal condition**].

Regarding the image observation, the following phrases can be 'Disorder':
- ground glass
- ground-glass
- ground glass opacity
- ground-glass opacity
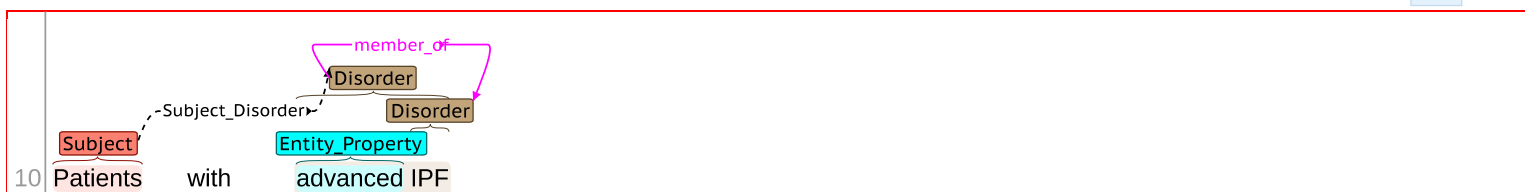- ground glass opacification
- ground-glass opacification

The following phrases also can be 'Disorder':
- chronic lung allograft dysfunction
- restrictive allograft syndrome

If the words, which indicate ambiguous degree/state, are not included in the UMLS data, the UMLS IDs should be selected for the disease themselves.

- IPF (UMLS; C1800706)
- advanced "IPF" (UMLS; C1800706)

edit

10 [Subject **Patients**] with [Entity_Property **advanced**] [Disorder **IPF**]
(member_of, Subject_Disorder, Disorder, Disorder)

The following words/phrases for stage/degree of disease will be annotated as Entity Property.

- stage I
- stage II
- stage IIIB
- stage IV
- stage IIIB/IV
- advanced (UMLS; C0205179)
- severe (UMLS; C0205082)

Measurement: Measurement for lung diseases

The definition of 'Measurement' for lung diseases may be mostly based on spirometry, and oxygen saturation.

'Measurement', which can be a group in phenotype, will be annotated in this category. The following words/phrases should be annotated as 'Measurement'.

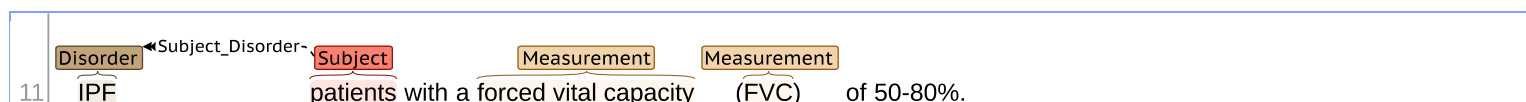The definition of 'Measurement' is based on the Semantic types of

- 'lbpr' (Laboratory Procedure)
- 'lbtr' (Laboratory or Test Result)
- 'clna' (Clinical Attribute)
- 'diap' (Diagnostic Procedure)

in the UMLS database.

These words/phrases may suggest the state of Disorder, and usually appear along with the corresponding values.

Based on spirometry:

- forced vital capacity (UMLS; C3714541)
- FVC (UMLS; C3714541)
- % predicted forced vital capacity (UMLS; C2919678)
- %FVC (UMLS; C2919678)
- total lung capacity (UMLS; C0040509)
- TLC (UMLS; C0040509)
- Diffusion Capacity of the Lung for Carbon Monoxide (UMLS; C1516251)
- DLco (UMLS; C1516251)
- % Predicted Diffusion Capacity of the Lung for Carbon Monoxide (UMLS; C4054207)
- %DLco (UMLS; C4054207)
- forced expiratory volume in 1 second (UMLS; C0849974)
- FEV1 (UMLS; C0849974)
- FEV1% (UMLS; C0849974)
- FEV1/FVC ratio (UMLS; C3815113)
- Forced expiratory flow (UMLS; C3804964)
- FEF (UMLS; C3804964)
  - maximal (mid-)expiratory flow [Rate] (UMLS; C0024966)
  - FEF 25%-75% (UMLS; C0024966)
  - mean mid-expiratory flow rate (UMLS; C1306020)
  - MMFR (UMLS; C0024966 or C1306020)
- peak expiratory flow (UMLS; C1518922)
- PEF (UMLS; C1518922)
- tidal volume (UMLS; C0040210)
- TV (UMLS; C0040210)

11 Disorder ◄Subject_Disorder– Subject   Measurement   Measurement
    IPF   patients with a forced vital capacity   (FVC)   of 50-80%.

Based on oxygen saturation:

- oxygen saturation (UNLS; C0523807)

- arterial oxygen saturation (UMLS; C0428175)
- SaO2 (UMLS; C0428175)
- venous oxygen saturation (UMLS; C0428176)
- SvO2 (UMLS; C0428176)
- peripheral oxygen saturation (UMLS; C2317096)
- SpO2 (UMLS; C2317096)

The following words/phrases are not from Spirometry, but definitely 'Measurement', as they can suggest the state of Disorder along with the corresponding values:
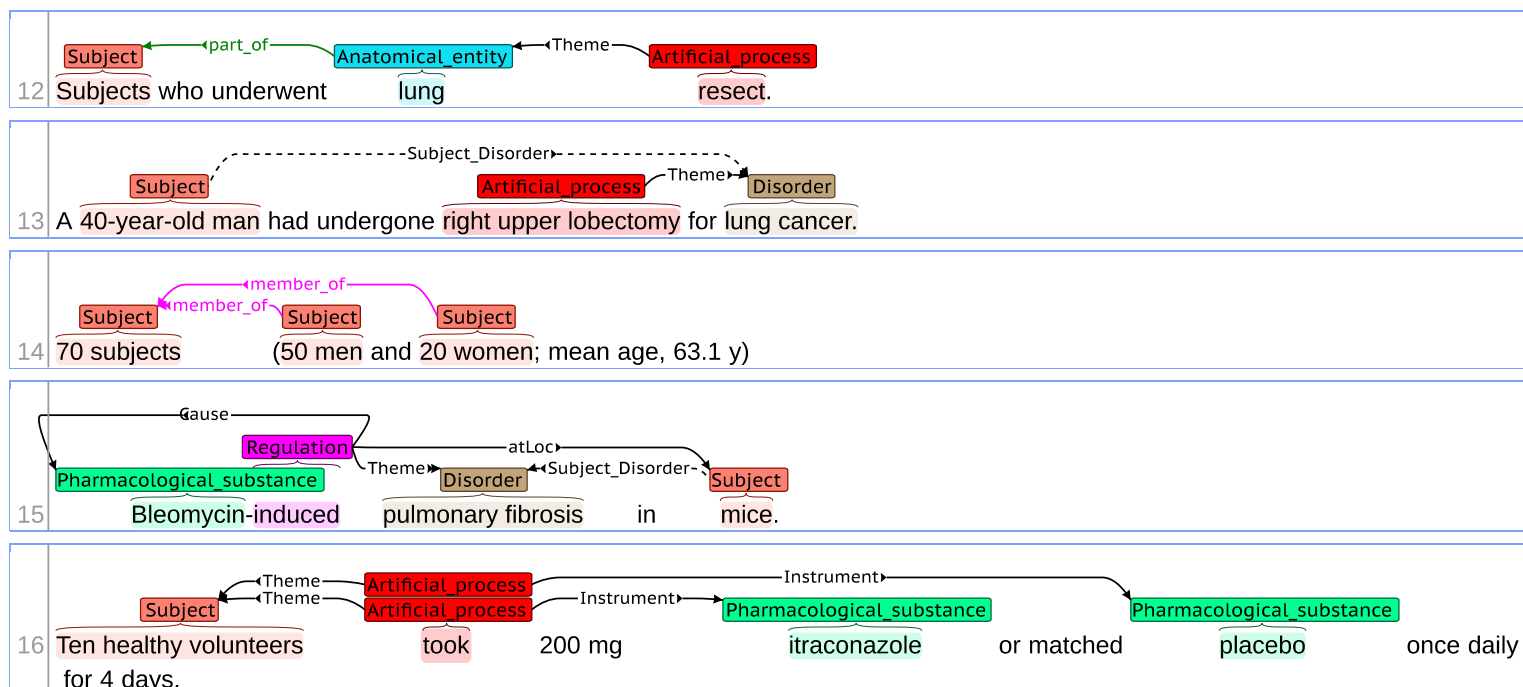
- 6-min walk distance (UMLS; C4055399)
- 6MWD (UMLS; C4055399)
- 6-min walk test (UMLS; C4055399)
- 6MWT (UMLS; C4055399)
- cardiopulmonary exercise test
- cardiac stress test
- cardiac diagnostic test
- CPX test
- maximal oxygen consumption
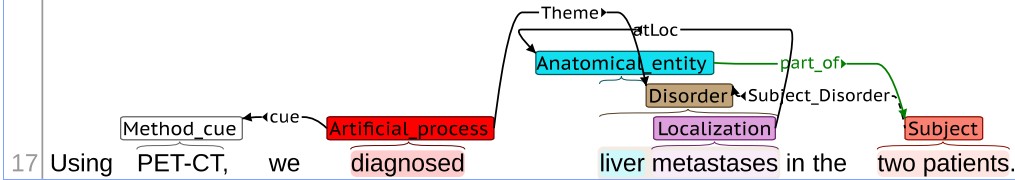- peak oxygen consumption
- VO2 peak
- VO2 max

Subject: Subjects for clinical trials, patients, and experimental animals (PHAEDRA)

The name of this entity group is based on the PHAEDRA corpus at NaCTeM.

Patients, and subjects for clinical trials, and experimental animals, such as mice, are categorized into this entity group.

This entity indicates 'individual level' above the Anatomical_entity.

17	Using PET-CT, we **diagnosed** liver metastases in the two patients.

(Annotated sentence 17: Method_cue — cue → Artificial_process "diagnosed"; Theme, atLoc → Anatomical_entity; Anatomical_entity — part_of → Subject; Disorder ← Subject_Disorder; Localization "liver metastases"; Subject "two patients".)
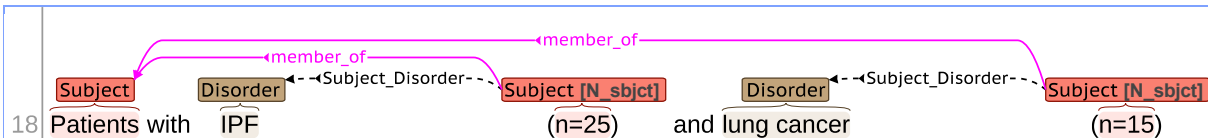
Regarding UMLS IDs for this entity, the IDs will be assigned as follows:

- C0030705; patients
- C2986479; Healthy Control
- C0009932; Control Groups (Controls who are not healthy.)

Regarding the attribute for this entity, Count_sbjct has been defined.

18	Patients with IPF (n=25) and lung cancer (n=15)

(Annotated sentence 18: Subject, Disorder "IPF" — Subject_Disorder; Subject [N_sbjct] "(n=25)" — member_of; Disorder "lung cancer"; Subject [N_sbjct] "(n=15)" — member_of.)

Cell: Cell types and cell lines

This entity group is based on Cell_natural and Cell_cultured in the GENIA-Meta-knowledge corpus at NaCTeM.

This entity is based on Cell ontology or Cell line ontology .

The definition of 'Cell' is based on the Semantic types of

- 'cell' (Cell)

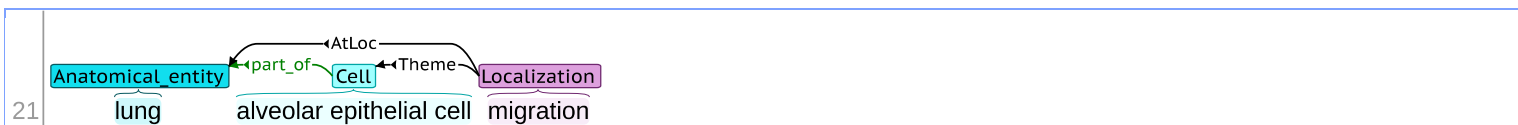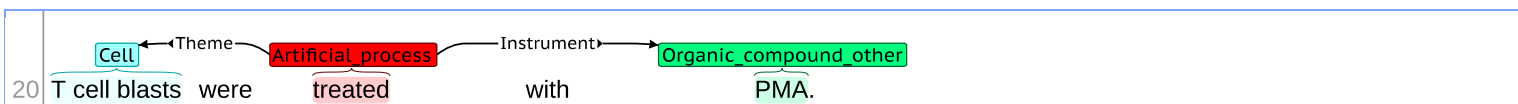in the UMLS database.

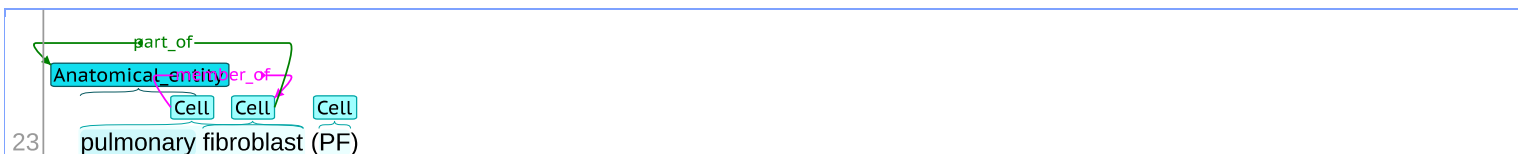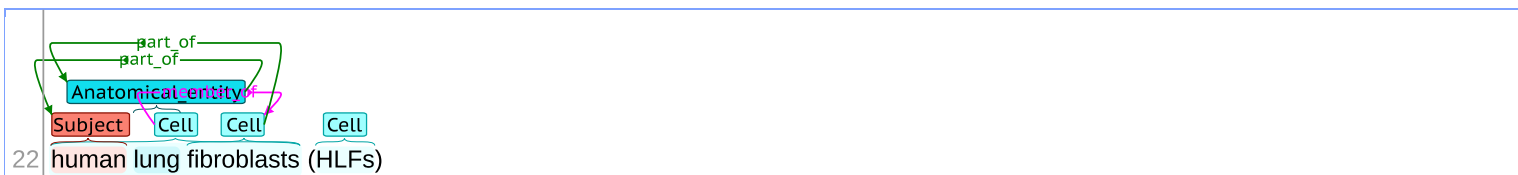'Cell' is categorized into this entity group.

The following words/phrases are examples of this entity.

- epithelial cells (UMLS ID:C0014597)
    - alveolar epithelial cells (UMLS ID:C0225698)
- endothelial cells
    - capillary endothelial cells
- fibroblasts (UMLS ID:C0016030)
- hematopoietic stem cells (= HSCs)(造血幹細胞)
- platelets (= thrombocytes)(血小板)
- mast cells (= mastocyte = labrocyte)(肥満細胞;顆粒細胞;マスト細胞)
- red blood cells (= erythrocytes)(赤血球)
- white blood cells (= leukocytes)(白血球)
    - monocytes (= mononuclear phagocytes)(単球)
        - macrophages (= phagocytes)
        - dendritic cells (樹状細胞)
        - osteoclasts (破骨細胞)
    - lymphocytes (リンパ球)

- T lymphocytes (= T cells)
  - T lymphoblasts (= T cell blasts)
- natural killer cells (= NK cells = large granular lymphocytes)
- B lymphocytes (= B cells)
  - granulocytes (= polymorphonuclear leucocytes = blood granulocytic cells = granular leukocytes) (顆粒球)
    - myeloblasts (骨髄芽球)
      - neutrophils (= neutrophiles)(好中球)
      - eosinophils (= eosinophil granulocytes)(好酸球)
      - basophils (好塩基球)

19 Cellular crosstalk between epithelial cells and fibroblasts.

20 T cell blasts were treated with PMA.

21 lung alveolar epithelial cell migration

The following cases are complicated:

22 human lung fibroblasts (HLFs)

23 pulmonary fibroblast (PF)

Anatomical_entity: Anatomical entities, including organs and tissues

The name of this entity group is based on Anatomical terminology.

Organs, such as lung, and tissues above cells are categorized into this entity group.

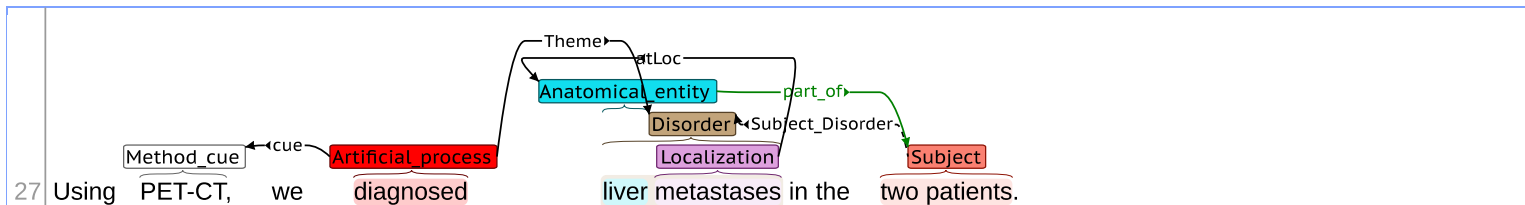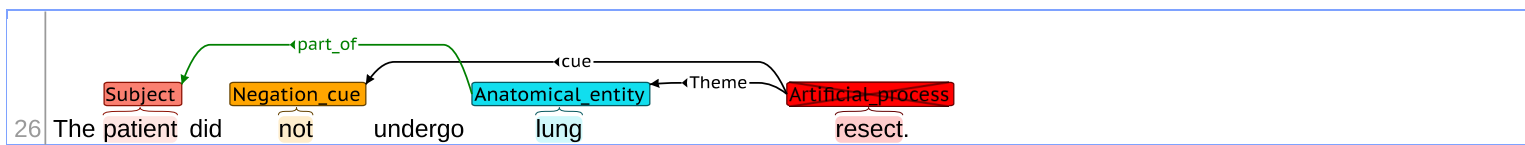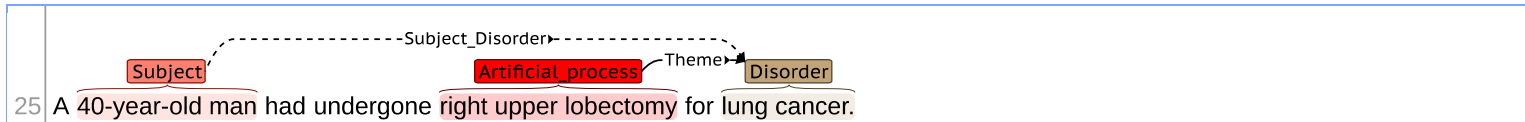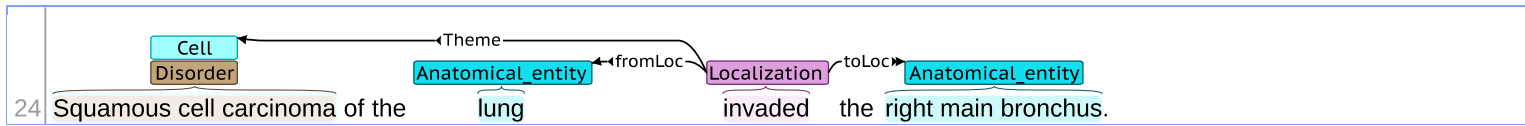The definition of 'Anatomical_entity' is based on the Semantic types of

- 'anst' (Anatomical Structure),
- 'ffas' (Fully Formed Anatomical Structure),
- 'bpoc' (Body Part, Organ, or Organ Component),
- 'bsoj' (Body Space or Junction), and
- 'tisu' (Tissue)

in the UMLS database.

Regarding the organ, such as lung, the following words/phrases can also be annotated.

- respiratory system（呼吸器系）

- respiratory apparatus
- ventilatory system（呼吸器系）
  - right lung
  - left lung
  - main bronchi
  - lobar bronchus
  - lingular division bronchi
  - trachea
  - diaphragm

24 | Squamous cell carcinoma of the lung invaded the right main bronchus.
（Cell / Disorder — Theme — Anatomical_entity — fromLoc — Localization — toLoc — Anatomical_entity）

25 | A 40-year-old man had undergone right upper lobectomy for lung cancer.
（Subject — Subject_Disorder — Artificial_process — Theme — Disorder）

26 | The patient did not undergo lung resect.
（Subject — part_of — Negation_cue — cue — Anatomical_entity — Theme — Artificial_process）

27 | Using PET-CT, we diagnosed liver metastases in the two patients.
（Method_cue — cue — Artificial_process — Theme — Anatomical_entity — atLoc — Disorder — Subject_Disorder — part_of — Subject — Localization）

Regarding the tissues in the lung, the following words/phrases can also be annotated.

- blood
- serum
- pulmonary alveolus（肺胞）
- pulmonary alveoli（肺胞）
  - alveoli
  - atrium
  - pulmonary artery
  - pulmonary vein
  - alveolar sacs
  - alveolar duct
  - connective tissue
  - capillary beds
  - mucous gland
  - mucosal lining
- pulmonary pleurae（胸膜）
  - parietal pleura
- bronchoalveolar lavage [fluid] (BAL) (UMLS; C0006279)

The following words/phrases are part of circulatory system（循環器系）.

- aorta （大動脈）
  - ascending aorta
  - descending aorta

- 'vena cava' or 'venae cavae'（大静脈）
    - inferior vena cava
    - superior vena cava
- artery（動脈）
    - pulmonary artery
    - coronary artery
- vein（静脈）
    - pulmonary vein

Cell_component: Cellular components, including organelle (GENIA)

The name of this entity group is based on the GENIA-Meta-knowledge corpus at NaCTeM.

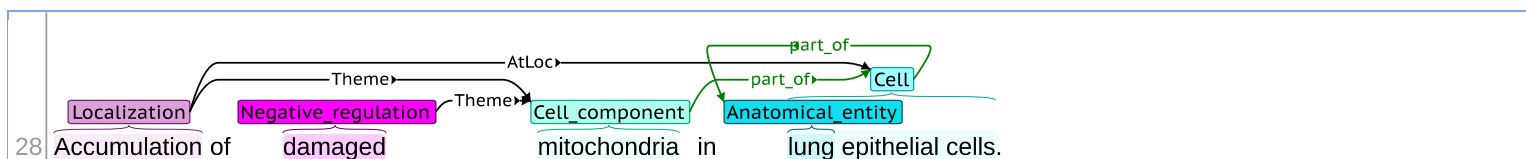The definition of 'Cell component' is based on the Semantic types of

- 'celc' (Cell Component)

in the UMLS database.

'Cell component' is categorized into this entity group.

The following words/phrases are examples of this entity.

- cytoplasm
- cytosol
- nucleus (UMLS ID:C0007610)
    - nuclear (UMLS ID:C0007610)
- transmembrane
- plasma membrane
- extracellular matrix
- mitochondria

Inorganic_compound: Inorganic compound, including metal ions (GENIA)

The name of this entity group is based on the GENIA-Meta-knowledge corpus at NaCTeM.

This entity is based on ChEMBL database.

The definition of 'Inorganic compound' is based on the Semantic types of

- 'inch' (Inorganic Chemical)
- 'elii' (Element, Ion, or Isotope)

in the UMLS database.

Inorganic compounds are typically chemicals which do not contain "carbon-hydrogen" bonds. This entity usually contains metal ions.

29	The required [Inorganic_compound: divalent metal ion] cofactor for [Protein_molecule: EcoRV]
30	[Inorganic_compound: Mg2+ ion] as a cofactor
31	Theme / Product / part_of — GGPs / Conversion / GGPs / [Inorganic_compound: Phospho-S6]

Organic_compound_other: Organic compound, excluding medicine (GENIA)

The name of this entity group is based on the GENIA-Meta-knowledge corpus at NaCTeM.

This entity is based on ChEMBL database.

Organic compounds, such as metabolites, are categorized into this entity group.

The following molecules can be categorized into this entity group.
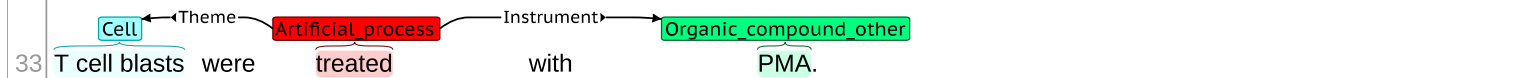
- carbohydrates
- nucleotides
- lipid molecules

The definition of 'Organic_compound_other' (Organic compounds/metabolites) is based on the Semantic types of

- 'orch' (Organic Chemical)
- 'carb' (Carbohydrate)
- 'eico' (Eicosanoid)
- 'lipd' (Lipid)
- 'strd' (Steroid)
- 'vita' (Vitamin)
- 'horm' (Hormone) (if it is not composed of peptides/proteins)
- 'nnon' (Nucleic Acid, Nucleoside, or Nucleotide) (if it is metabolite instead of part(s) of genes/gene prodcuts)
- 'bacs' (Biologically Active Substance) (if it is not peptides/proteins)

in the UMLS database.

32	The [Inorganic_compound: alpha-phosphate] —part_of→ of [Organic_compound_other: ATP].

33	Cell ◄Theme─ Artificial_process ─Instrument► Organic_compound_other T cell blasts were **treated** with **PMA**.

Pharmacological_substance: Medicine (PHAEDRA)

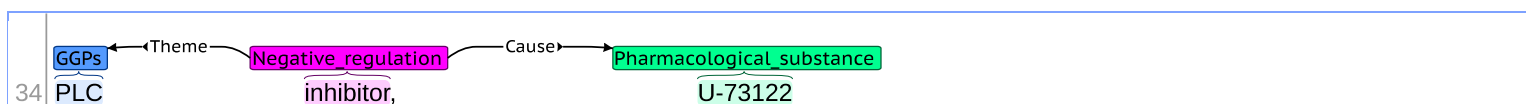The name of this entity group is based on the PHAEDRA corpus at NaCTeM.

This entity is based on ChEMBL database.

The definition of 'Pharmacological_substance' is based on the Semantic types of

- 'phsu' (Pharmacologic Substance)
- 'antb' (Antibiotic), and
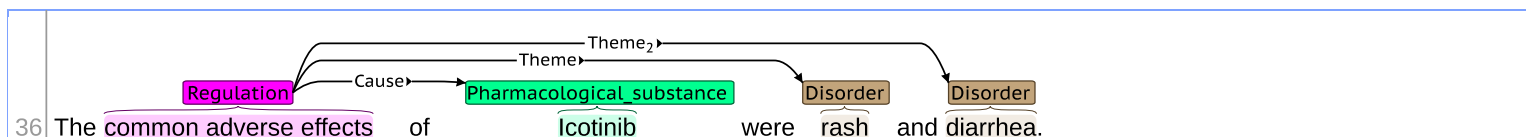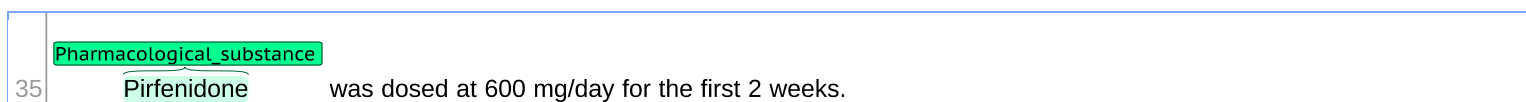- 'irda' (Indicator, Reagent, or Diagnostic Aid)

in the UMLS database.
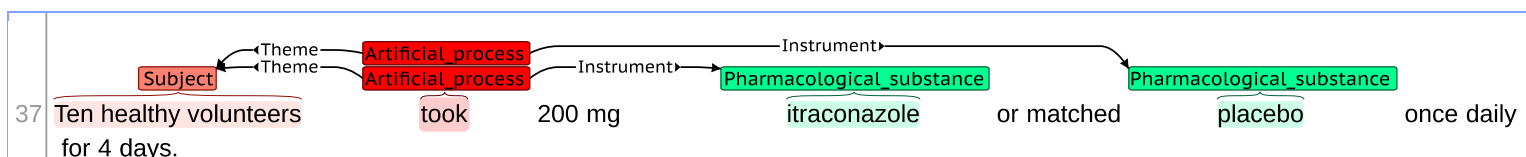
Medicines are categorized into this entity group.

34	GGPs ◄Theme─ Negative_regulation ─Cause► Pharmacological_substance PLC **inhibitor**, **U-73122**

The following protein medicines can also be categorized into this group.

- Nivolumab
- Tocilizumab

35	Pharmacological_substance **Pirfenidone** was dosed at 600 mg/day for the first 2 weeks.
36	Regulation ─Cause► Pharmacological_substance ─Theme► Disorder ─Theme₂► Disorder The **common adverse effects** of **Icotinib** were **rash** and **diarrhea**.

Although 'placebo' is not any medicine, this word can be tentatively categorized into this entity.

37	Subject ◄Theme─ Artificial_process / Artificial_process ─Instrument► Pharmacological_substance ─Instrument► Pharmacological_substance **Ten healthy volunteers** **took** 200 mg **itraconazole** or matched **placebo** once daily for 4 days.

GGPs: Gene or gene products (GENIA)

The name of this entity group, 'Gene or gene products' is based on the Protein/RNA/DNA from GENIA-Meta-knowledge corpus at NaCTeM.

The definition of 'GGPs' (Gene or Gene products) is based on the Semantic types of

- 'aapp' (Amino Acid, Peptide, or Protein),
- 'enzy' (Enzyme),
- 'rcpt' (Receptor),

- 'horm' (Hormone) (if it is composed of peptides/proteins),
- 'gngm' (Gene or Genome),
- 'nnon' (Nucleic Acid, Nucleoside, or Nucleotide) (if it is part(s) of genes/gene prodcuts)
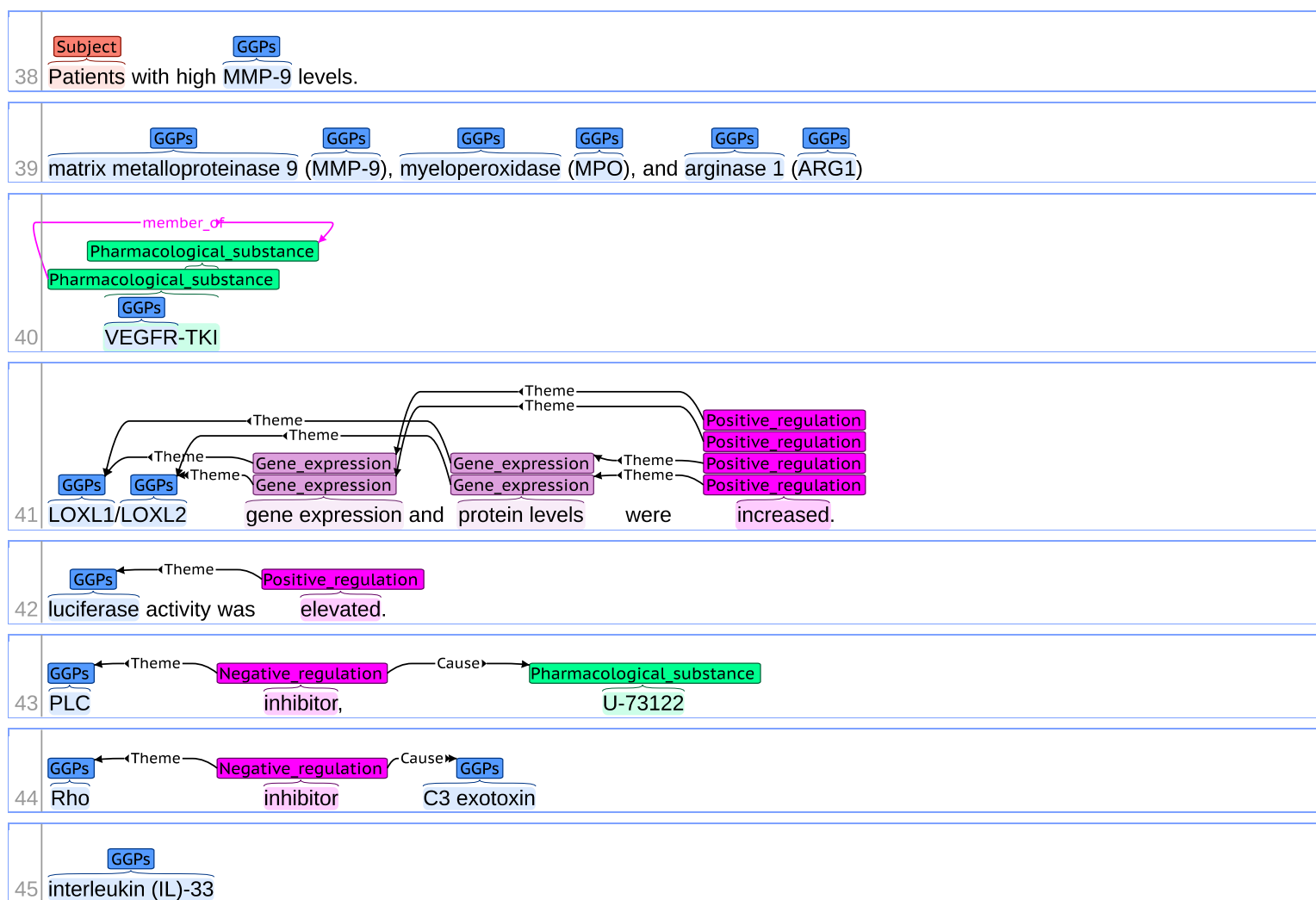
in the UMLS database.

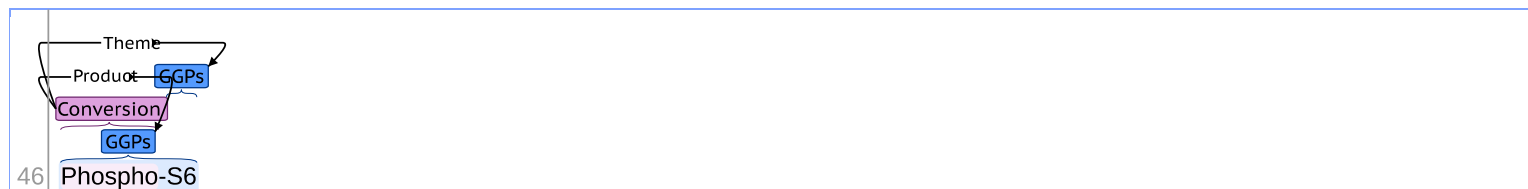However, some data of the Semantic type, 'horm', might be categorized into Organic_compound_other.

Some data of the Semantic type, 'nnon', also may be classified into Organic_compound_other, if they are metabolites, instead of parts of genes/gene products.

Protein molecules, gene products, and genes are categorized into this entity group.

- interleukin (UMLS ID:C0021764)
- IL (UMLS ID:C0021764)
- PINK1 gene (UMLS ID:C1422771)

38	**[Subject]** **[GGPs]** Patients with high MMP-9 levels.
39	**[GGPs] [GGPs] [GGPs] [GGPs] [GGPs] [GGPs]** matrix metalloproteinase 9 (MMP-9), myeloperoxidase (MPO), and arginase 1 (ARG1)
40	*member_of* **[Pharmacological_substance]** **[Pharmacological_substance]** **[GGPs]** VEGFR-TKI
41	**[GGPs] [GGPs]** Gene_expression Gene_expression Positive_regulation (×4) LOXL1/LOXL2 gene expression and protein levels were increased.
42	**[GGPs]** —Theme— **[Positive_regulation]** luciferase activity was elevated.
43	**[GGPs]** —Theme— **[Negative_regulation]** —Cause→ **[Pharmacological_substance]** PLC inhibitor, U-73122
44	**[GGPs]** —Theme— **[Negative_regulation]** —Cause→ **[GGPs]** Rho inhibitor C3 exotoxin
45	**[GGPs]** interleukin (IL)-33

In case of modified GGPs, annotation can be performed as follows:

46	Theme / Product **[GGPs]** / Conversion **[GGPs]** Phospho-S6

Regarding *mutant/variant* information, it is extremely difficult to normalize them, as there are too many possible mutants for any GGPs. Thus, such *mutant/variant* information will be included in Genetic_info:

For protein molecules, the PRO database (Protein Ontology) in the PIR database are usually used for normalization in the other corpus. On the other hand, the granularity of IDs from the UniProt database is usually too small to use. Thus, the UMLS database is adopted in this work.

An example is indicated as follows:

Protein name	UMLS ID	PRO ID	UniProt ID
Vascular endothelial growth factor A	C1823619	PR:000017284	-
VEGFA	C1823619	PR:000017284	-
human Vascular endothelial growth factor A	-	PR:P15692	P15692
hVEGFA	-	PR:P15692	P15692

Chemokines, such as IL-8, and their corresponding receptors, which are composed of proteins, are categorized into this entity.

Protein name	cells attracted by corresponding protein	cytokine/receptor	UMLS ID
CCL1	T-lymphocytes	cytokine for CCR8	
CCL2	Monocytes / macrophages / T-lymphocytes	cytokine for CCR2	
CCL3	Monocytes / macrophages	cytokine for CCR1	
CCL5 (RANTES)	Monocytes / macrophages / eosinophils	cytokine for CCR5	C0072978
CCL7	Monocytes / macrophages / eosinophils	cytokine for CCR2	
CCL8	Monocytes / macrophages	cytokine for CCR1, CCR2B, CCR5	
CCL11	Eosinophils	cytokine for CCR2, CCR3, CCR5	
CCL13	Monocytes / macrophages / eosinophils	cytokine for CCR2, CCR3, CCR5	
CCL17 (TARC)	Monocytes / macrophages / T-lymphocytes	cytokine for CCR4	
CCL22 (MDC)	Monocytes / macrophages / T-lymphocytes	cytokine for CCR4	
CCL24	Eosinophils	cytokine for CCR3	
CCL26	Eosinophils	cytokine for CCR3	
CXCL8 (IL-8)	Neutrophils	cytokine for CXCR1, CXCR2	C0079633
CCR1	Mast cells	receptor for CCL3	
CCR2	Mast cells	receptor for CCL2, CCL7, CCL12	
CCR3	Mast cells	receptor for CCL11, CCL26, CCL7, CCL13, CCL5, CCL15	
CCR4	Mast cells	receptor for CCL17, CCL22	
CCR5	Mast cells	receptor for CCL3, CCL4, CCL5	
CXCR2	Mast cells	receptor for CXCL8, CXCL2, CXCL3, CXCL5	
CXCR4	Mast cells	receptor for CXCL12	

Entity Property: various technical terms for lung diseases

Technical terms, which are not any other Entity information, can be categorized into this category, 'Entity Property'.

The following words/phrases for Disorder can be tentatively annotated for this category.

- stage I (UMLS; C0441766)
- stage II (UMLS; C0441767)
- stage III (UMLS; C0441771)
- stage IIIB (UMLS; C0456599)
- stage IV (UMLS; C0441772)
- stage IIIB/IV
- stable
- severe
- grade 1
- chronic phase (UMLS; C0457343)
- chronically progressive phase
- subacutely progressive phase
- subacute phase
- stable phase
- inflammatory phase

The following words/phrases, which are used to indicate so-called *cell marker expression* or *immunophenotyping*, for Cell and GGPs can be tentatively annotated for this category.

- null (UMLS ID:C0205160)
- (-) (UMLS ID:C0205160)
- (+) (UMLS ID:C1446409)

In the following example, 'CD4(+) T-cells' have got a UMLS ID (C0039215), whereas 'CD4(+)CD28(null) T-cells' do not have an ID, so that member_of relation is annotated to indicate their relationships.
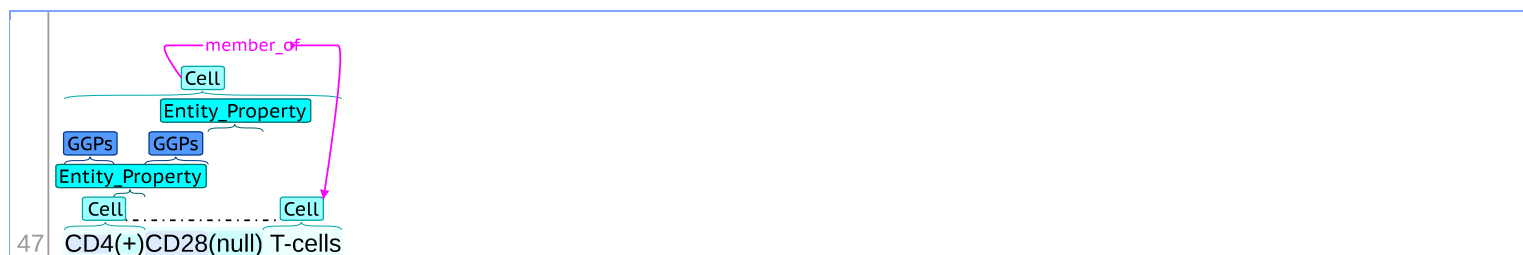
The following words related to cell cycle (UMLS; C0007586) can be annotated in this category: (see Cell cycle)

- cell cycle(UMLS; C0007586)
  - G1 phase (UMLS; C0079395)
  - G1/S [transition] checkpoint (UMLS; C1517340)
  - S phase (UMLS; C0080129)
  - G2 phase (UMLS; C0079396)
  - G2/M [transition] checkpoint (UMLS; C3549430)
  - M phase (UMLS; C0007591)

- prophase
- prometaphase
- metaphase
- anaphase
- telophase

Genetic_info: other technical terms for mutation info

Technical terms for mutation information, which cannot be included in GGPs, can be categorized into this category, 'Genetic_info'.

However, as it would be difficult to assign the UMLS IDs to this type of entity, this entity will be annotated without the corresponding IDs.
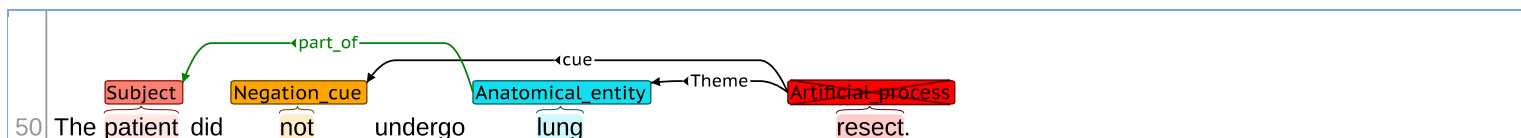
The following words must be typical mutation info:

- 1G/2G
- rs1800925(C/T)
- rs1800925 T allele
- CC genotype
- genotypes CT and TT
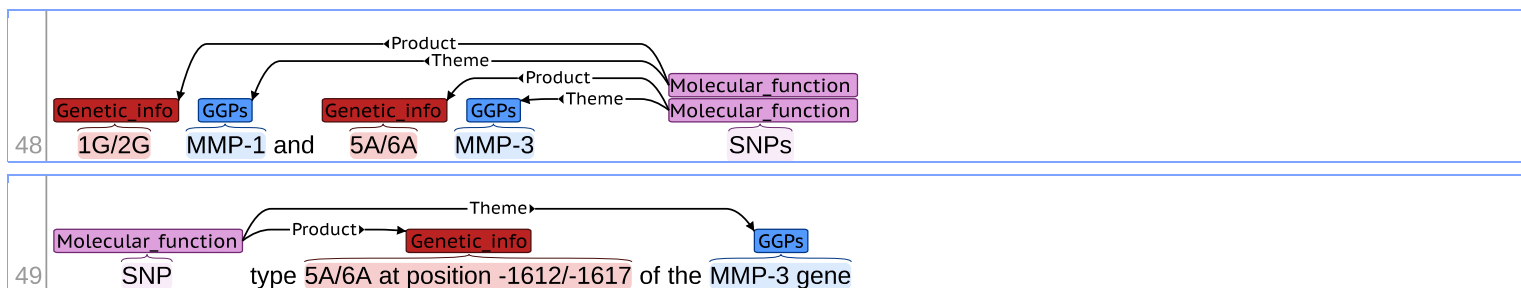
Negation cue: Negation words and phrases

Negation words, such as "not", can be combined with event expressions.

However, even if the negation words can be identified as Negation_cue, such cues are not always connected with event expressions.

In addition to typical negation words, such as "no", "not', "none" and "neither ~ nor ~", the following words/phrases can be negation cues.

- instead of
- lack(s)/lacking
- loss/lost
- rather than
- other than

- without
- absence/absent
- barely
- failed to
- inability/incapable
- (very) little
- (very) poor
- (too/very) low

Speculation cue: Speculation words and phrases

In addition to verbs, such as "suggest", "show", and "indicate", the following words/phrases can be 'Speculation cue'.

- can/could *(be) (concluded/considered/described/interpreted/rationalized)*
- may/might
- Analyze(d/s)/analyse(d/s)
- Implication
- Insight(s)
- Proposal(s)
- probability
- ~ consistent with
- ~ agree(s) (well) with
- ~ appear(s) *(likely/favorable/to involve)*
- Argue against
- apparent/apparently
- understand(ing)
- hypothesize
- suggestive (of)
- Putative
- Presumably/probably
- presume(d)
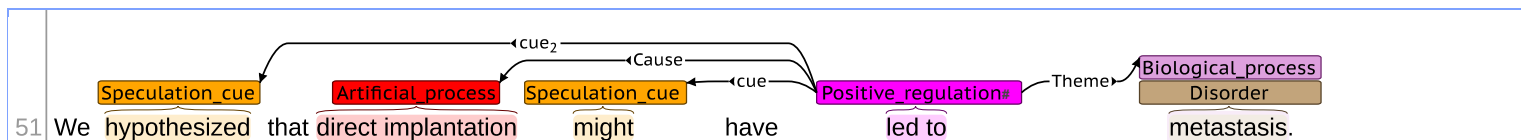
The following words indicate that the events involved are not clear/known.

- unclear
- unknown

The following words indicate that the events involved are tested/investigated. Thus, the events involved have not been confirmed yet.

- test(ed)
- investigate(d)/investigation
- evaluate(d)/evaluation
- measure(d)/measurement

- assess(ed)/assessment

52 Markers of fibrogenesis, including collagen and CTGF were evaluated by measuring mRNA level using RT-PCR.
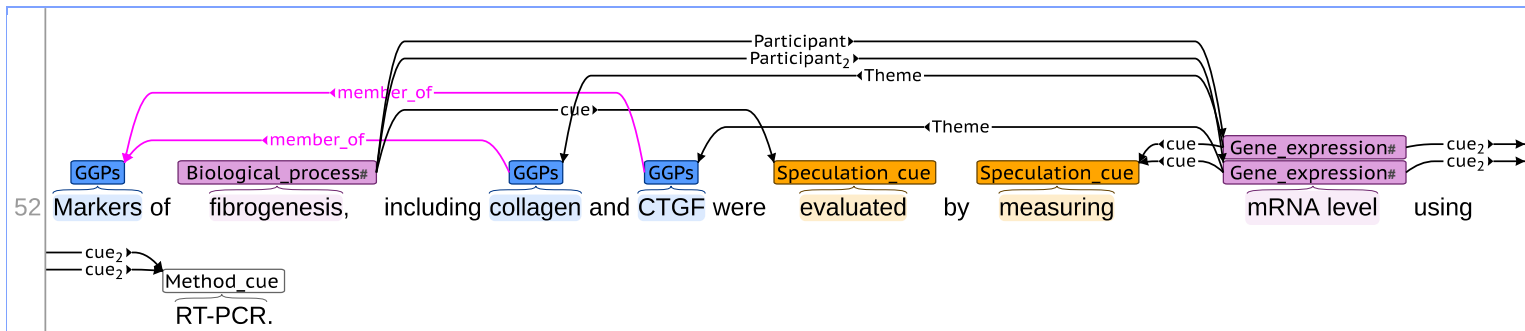
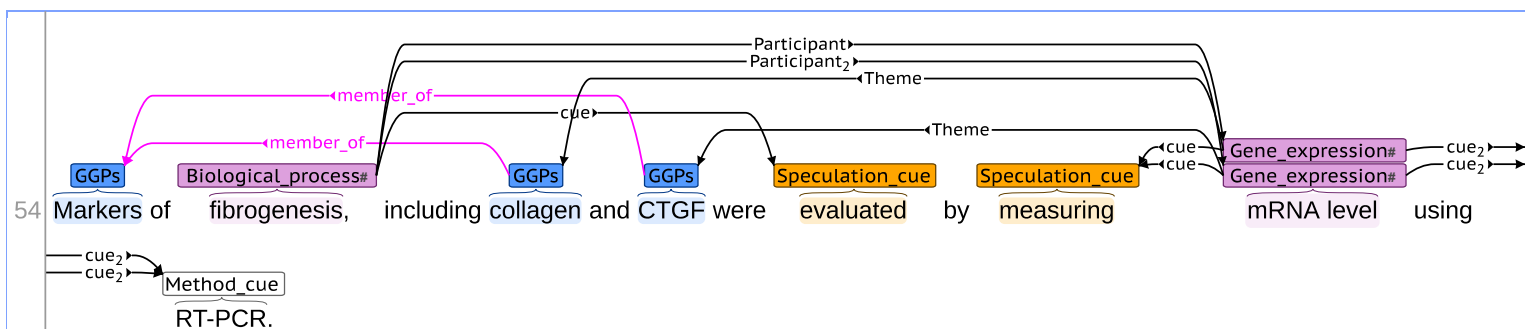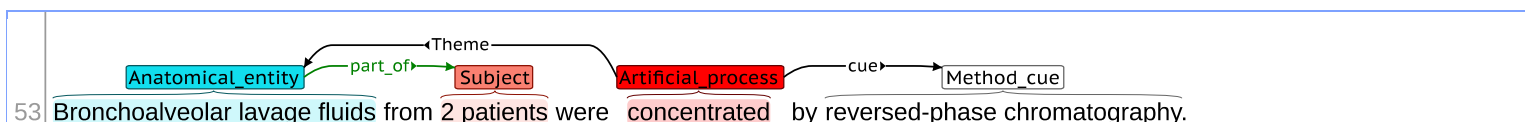# Method cue: Cues by experimental study/clinical diagnosis types

The 'Method cue' may suggest confirmation/speculation degree of events. This cue can be types of experimental studies/clinical examination to analyse biological events.

Specific actions for experimental procedures/clinical examiniations should be annotated as Artificial_process. Such actions may make some effects on entities.

- *(experimental/computational)* analyses/tests/assays

More concrete analysis names for the cue are as follows:

- [reversed-phase *etc.*] choromatography
- enzyme-linked immunosorbent assay (UMLS; C0014441)
- ELISA (UMLS; C0014441)
- chest X-ray (UMLS; C0039985)
- CT scan(s) (UMLS; C0040405)

53 Bronchoalveolar lavage fluids from 2 patients were concentrated by reversed-phase chromatography.

54 Markers of fibrogenesis, including collagen and CTGF were evaluated by measuring mRNA level using RT-PCR.

Events

Phenomena and processes are defined as events in this corpus. The events will be described below.

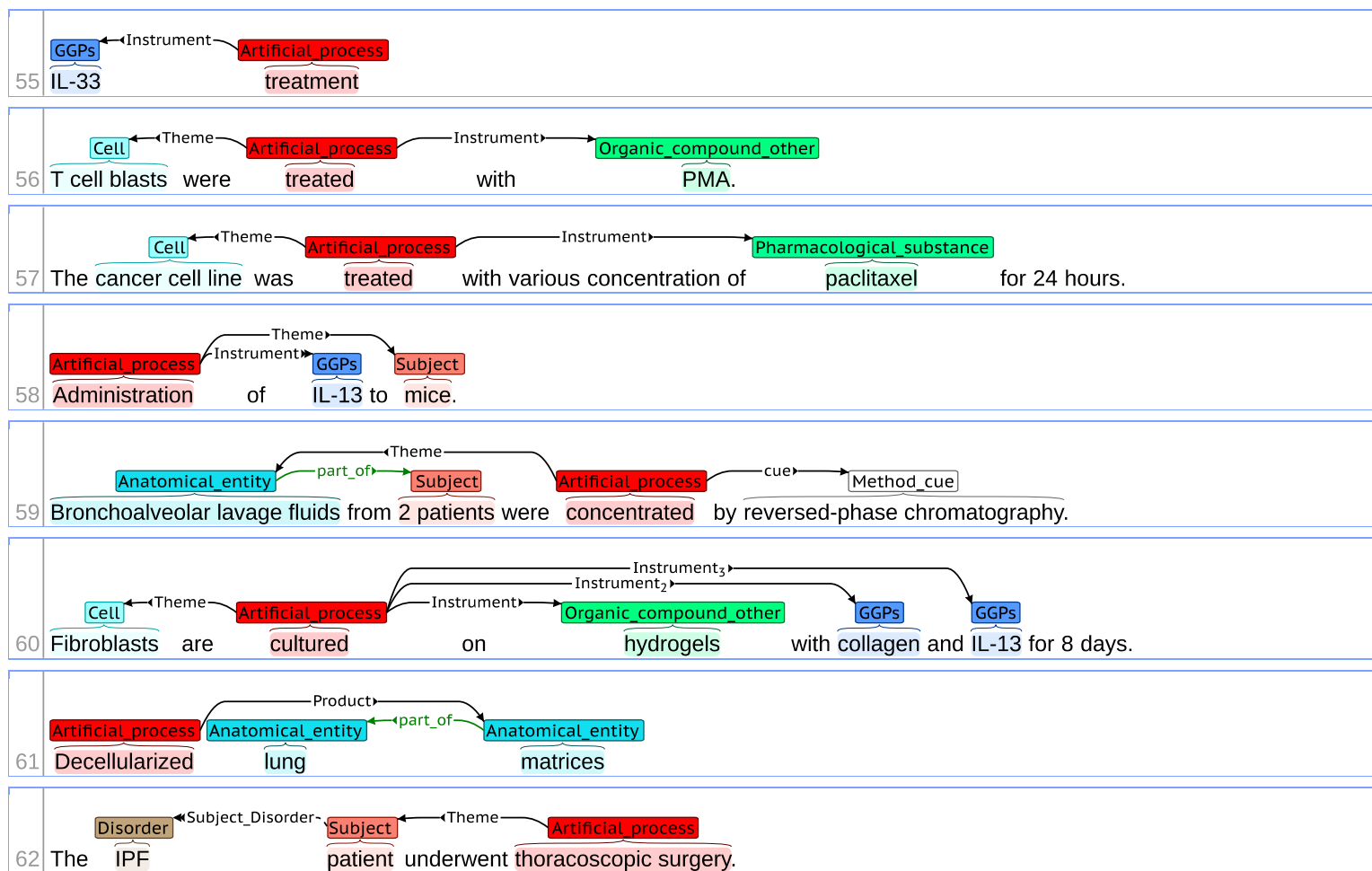# Artificial_process: Artificial process, such as experimental procedures/medical treatments

Category: Artificial process

This event describes 'Artificial processes', which are usually specific actions for experimental procedures/medical treatments. Thus, this event can affect entities and the other events.

Analysis/assay names should be annotated as Method cue, which may suggest confirmation/speculation degree of events.

The following words/phrases can be triggers for this event:

- treat/treatment
    - pretreatment (UMLS; C3539076)
- incubate/incubation (UMLS; C1439852)
    - preincubation
- addition
- culture/cultured (UMLS; C1449619)
    - coculture (UMLS; C0282547)
- transfection (UMLS; C0040669)
- exposure (UMLS; C0332157)
- administer/administration (UMLS; C1621583)
- adoptive transfer (UMLS; C0376518)
- concentrate (UMLS; C2003864)
- decellularize(d)/decellularization (UMLS; C3827823)
- surgery (UMLS; C0543467)

55 | GGPs ←Instrument— Artificial_process
IL-33 treatment

56 | Cell ←Theme— Artificial_process —Instrument→ Organic_compound_other
T cell blasts were treated with PMA.

57 | Cell ←Theme— Artificial_process —Instrument→ Pharmacological_substance
The cancer cell line was treated with various concentration of paclitaxel for 24 hours.

58 | Artificial_process —Theme→ —Instrument→ GGPs Subject
Administration of IL-13 to mice.

59 | Anatomical_entity —part_of→ —Theme— Subject Artificial_process —cue→ Method_cue
Bronchoalveolar lavage fluids from 2 patients were concentrated by reversed-phase chromatography.

60 | Cell ←Theme— Artificial_process —Instrument→ Organic_compound_other —Instrument₂→ —Instrument₃→ GGPs GGPs
Fibroblasts are cultured on hydrogels with collagen and IL-13 for 8 days.

61 | Artificial_process Anatomical_entity —Product→ ←part_of— Anatomical_entity
Decellularized lung matrices

62 | Disorder ←Subject_Disorder— Subject ←Theme— Artificial_process
The IPF patient underwent thoracoscopic surgery.

Arguments:

The *Theme* for this event can be targets of this event, such as Subject, Anatomical_entity, and Cell.

The *Instrument* for this event can be entities, such as GGPs, Organic_compound_other, and

Pharmacological_substance, which are used and given to *Theme*, to carry out an artificial process.

The *Product* for this event can be entities, which can be produced by this event.

Biological_process: Biological process beyond cellular/molecular levels
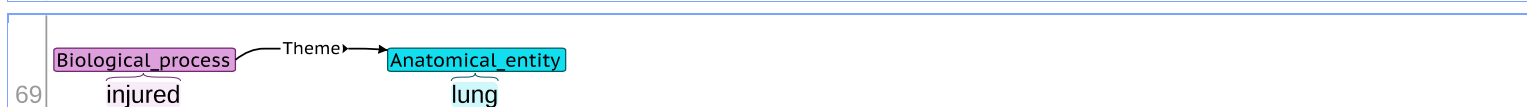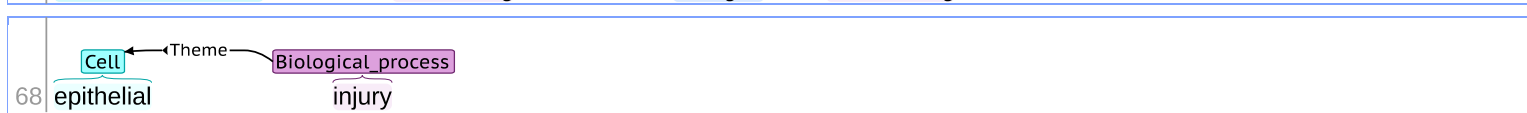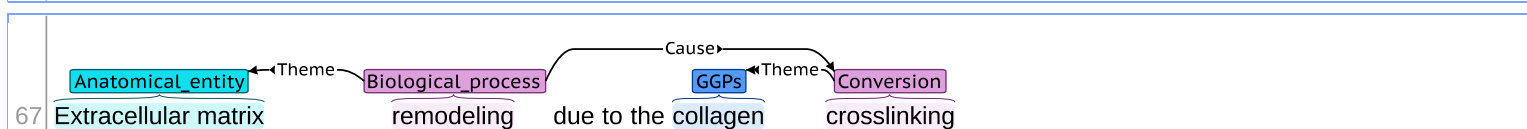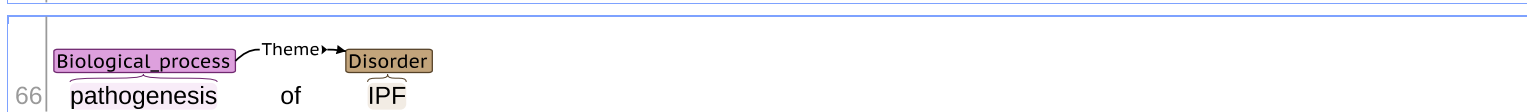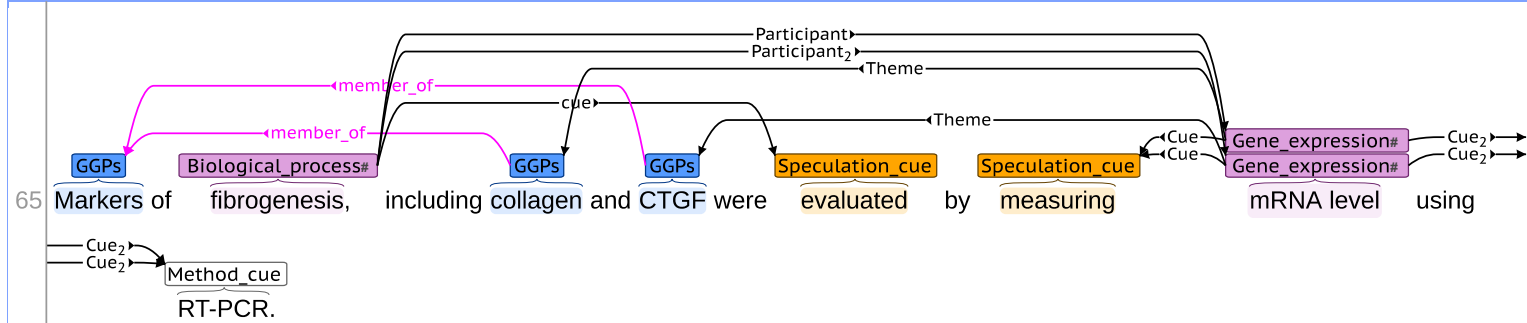
Category: Biological process

This event is based on the GENIA-Meta-knowledge corpus at NaCTeM.

This event will be difficult to be classified into the other biological events, such as Cellular_process and Molecular_function, as it is on higher level than the other events.

The following words/phrases can be triggers of this event:

- organogenesis (UMLS; C0242290)
    - angiogenesis [process] (UMLS; C0302600)
- fibrogenesis (UMLS; C0596570)
    - fibrotic response
    - profibrotic response
- pathogenesis (UMLS; C0699748)
    - immunopathogenesis
    - etiopathogenesis
    - development [of some disease]
- pathophysiological mechanism
- acute exacerbation (UMLS; C0743630)
- angiostasis (UMLS; C3179230)
- resistance (UMLS; C1514892)
- homeostasis (UMLS; C0019868)
- sensitivity (UMLS; C2349185)
- [tissue] remodeling (UMLS; C1820201)
- cilium biogenesis (UMLS; C1155941)
- cilium assembly (UMLS; C1155941)
- inflammatory response (UMLS; C1155266)
- immune response [process] (UMLS; C0301872)
- injury
    - tissue damage (UMLS; C0010957)
    - scarring (UMLS; C0008767 (for Cicatrization))
- adverse event (UMLS; C0877248)

63 | Anatomical_entity ←Theme— Biological_process
matrix   remodeling   phase

64 | Anatomical_entity ←Theme— Biological_process
lung   fibrogenesis

65 Markers of fibrogenesis, including collagen and CTGF were evaluated by measuring mRNA level using RT-PCR.

(annotations: GGPs, Biological_process#, member_of, cue, GGPs, GGPs, Speculation_cue, Speculation_cue, Gene_expression#, Gene_expression#, Participant, Participant₂, Theme, Theme, Cue, Cue, Cue₂, Cue₂, Method_cue)

66 pathogenesis of IPF

(annotations: Biological_process — Theme → Disorder)

67 Extracellular matrix remodeling due to the collagen crosslinking

(annotations: Anatomical_entity ← Theme — Biological_process — Cause → GGPs ← Theme — Conversion)

68 epithelial injury

(annotations: Cell ← Theme — Biological_process)

69 injured lung

(annotations: Biological_process — Theme → Anatomical_entity)

Arguments:

The *atLoc*, *fromLoc* and *toLoc* arguments for this event must be Subject, Anatomical_entity, Cell, and Cell_component, which are the locations where this event occurs. The *atLoc* argument indicates the location at which this event occurs. The *fromLoc* indicates the location from which this event starts, whereas the *toLoc* indicates the location to which this event proceeds.

The other arguments, such as *Cause*, *Theme*, *Participant*, and *Product*, for this event can be any entities or events.

The *Cause* for this event is entities/events, which cause this event, whereas the *Theme* for this event is entities/events, which are targets of this events.

The *Participant* is entities/events, which are involved in this event.

The *Product* is entities/events, which is produced by this event.

The *disorder* argument is Disorder.

The *Cue* argument is cues, such as Negation cue, Speculation cue or Method cue.
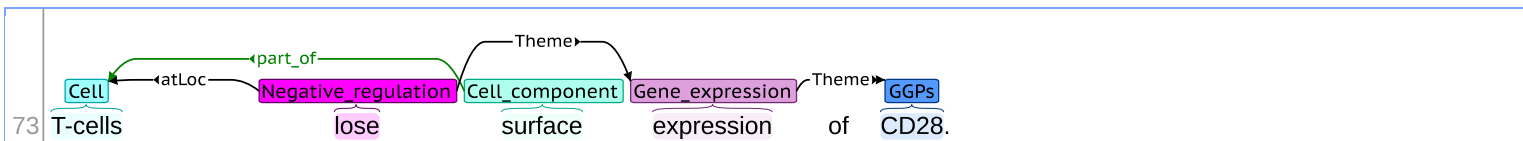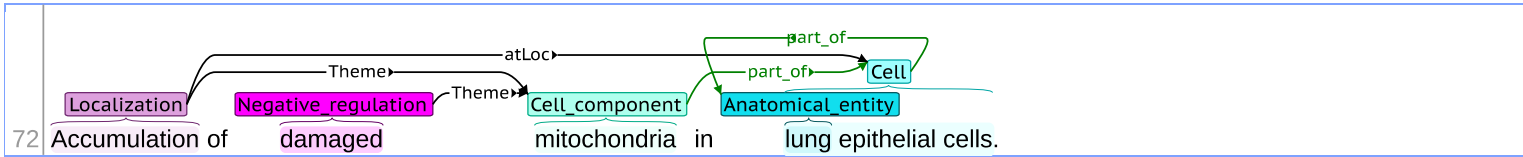
Regulation: Regulation

Category: Biological process

This event describes "regulation process" event (UMLS ID:C1327622).

If it is not clear whether the trigger words are "positive" or "negative", this "Regulation" event will be selected.

The following words/phrases can be triggers of this event:

- *affect*; *affected*; *affects*

- *alter*; *altered*; *alters*
- *change*; *changed*; *changes*
- *effect*
- *influence*; *influenced*; *influences*
- *modify*; *modified*; *modifies*; *modification*

The following words/phrases may be triggers, depending on the situations:

- *[play a/an ~] role [in ~ing]* (Depending on "*~ing*")
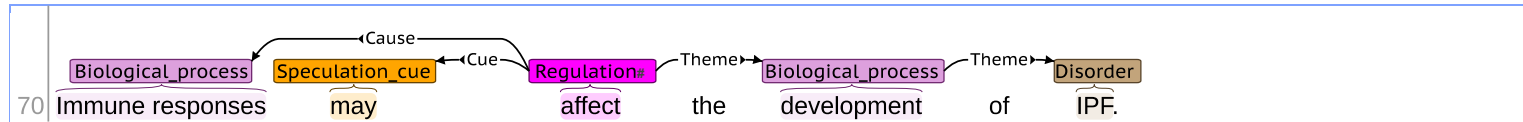
Arguments:

The *Theme* (optional; zero or one) indicates events, such as Gene_expression, or entities, such as GGPs, and Organic_compound_other, that are regulated. Function or quality of *Theme* can be affected.

The *Cause* (optional; zero or one) indicates events or entities, such as GGPs, Organic_compound_other, that are the stated cause of the `Regulation`.

The *atLoc* (optional) indicates the location where the `Regulation` event occurs: Cell_component, Cell or Anatomical_entity.

The *disorder* (optional) indicates the Disorder for which the `Regulation` event occurs.

The *Cue* argument (optional) is cues, such as Negation cue, Speculation cue or Method cue.

Negative_regulation: Negative regulation

Category: Biological process

This event is based on the GENIA-Meta-knowledge corpus at NaCTeM.

This event describes "inactivation/inhibition/down-regulation" event (UMLS ID:C3463820).

The following words/phrases, which are verbs and nominalized verbs, can be triggers of this event:

- *abolish*, *abolishes*, *abolished*, *abolishment*
- *abrogate*, *abrogates*, *abrogated*, *abrogation*
- *attenuate*, *attenuates*, *attenuated*, *attenuation*
- *block*, *blocks*, *blocked*, *blocking*
- *decline*, *declines*, *declined*
- *decrease*, *decreases*, *decreased*
- *deprive*, *deprives*, *deprived*, *deprivation*
- *dimish*, *diminishes*, *dimished*
- *down-regulate*, *down-regulates*, *down-regulated*, *down-regulation*
- *impair*, *impairs*, *impaired*, *impairment*
- *inactivate*, *inactivates*, *inactivated*, *inactivation*

- *inhibit*, *inhibits*, *inhibited*, *inhibition*
- *negatively regulate*, *negatively regulates*, *negatively regulated*, *negative regulation*
- *prevent*, *prevents*, *prevented*, *prevention*
- *reduce*, *reduces*, *reduced*, *reduction*
- *remove*, *removes*, *removed*, *removal*
- *repress*, *represses*, *repressed*, *repression*
- *suppress*, *suppresses*, *suppressed*, *suppression*
- *damage*
- *lose*, *loses*, *lost*, *loss*

71 Phosphorylation of AKT2 was downregulated.

72 Accumulation of damaged mitochondria in lung epithelial cells.

73 T-cells lose surface expression of CD28.

Moreover, the following words, which indicate roles, can also be triggers of this event.

- *inhibitor*, *inhibitors*

74 PLC inhibitor, U-73122

75 Rho inhibitor C3 exotoxin

76 mTOR inhibitor attenuated EMT in AECs.
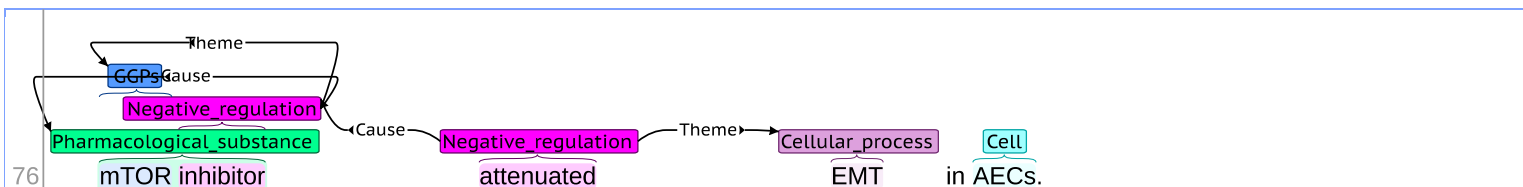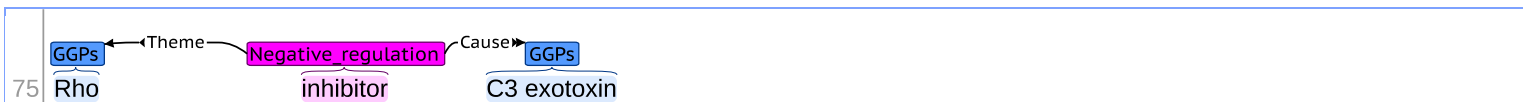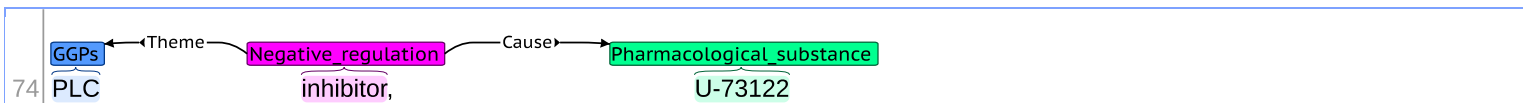
The following phrase can also be triggers of this event.

- *inhibitory effect(s)*

Moreover, Disorder can be annotated along with this event.
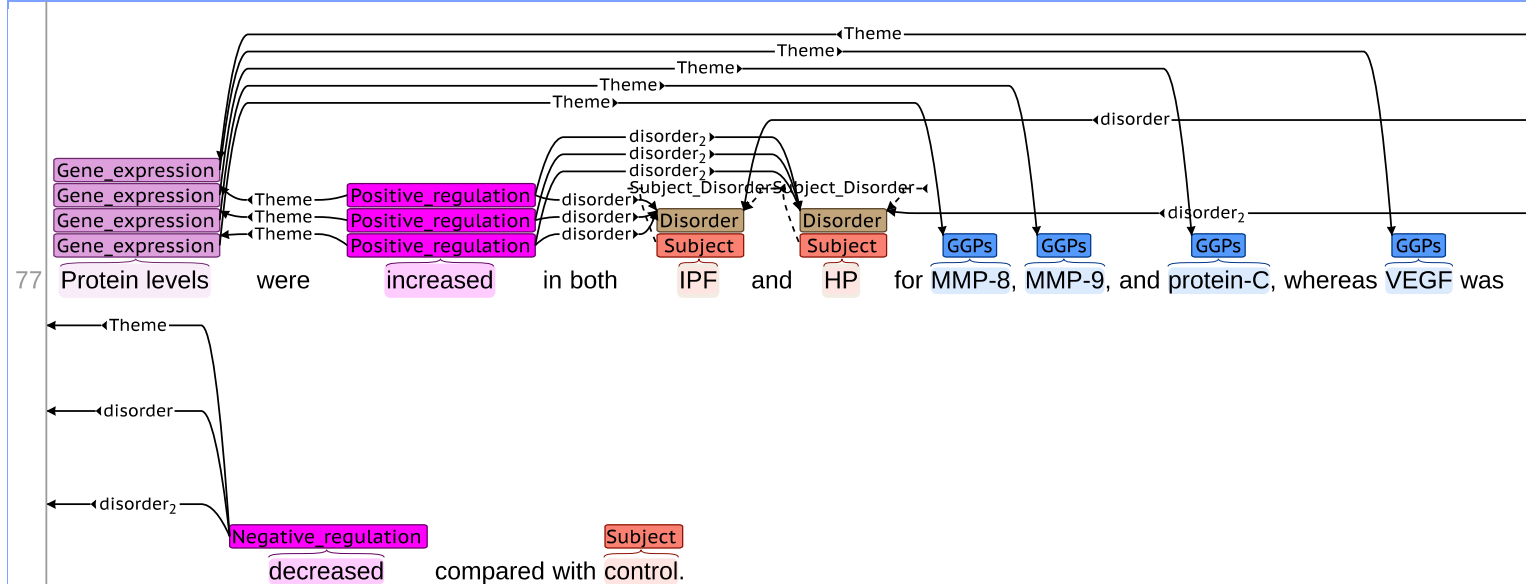
Arguments:

The *Theme* (optional; zero or more) indicates events or entities, such as GGPs and Organic_compound_other, that are negatively regulated. Function or quality of *Theme* can be affected negatively.

The *Cause* (optional; zero or one) indicates events or entities, such as GGPs and Organic_compound_other, that are the stated cause of the `Negative regulation`.

The *atLoc* (optional) indicates the location where the `Negative regulation` event occurs: Cell_component, Cell or Anatomical_entity.

The *disorder* (optional) indicates the Disorder for which the `Negative regulation` event occurs.

The *Cue* argument (optional) is cues, such as Negation cue, Speculation cue or Method cue.

Positive_regulation: Positive regulation

Category: Biological process

This event is based on the GENIA-Meta-knowledge corpus at NaCTeM.

This event describes "activation/up-regulation" event (UMLS ID:C1879547).

The following words/phrases can be triggers of this event:

- *activate*; *activates*; *activated*; *activation*
- *cause*; *causes*; *caused*
- *control*; *controls*; *controled*
- *depend*; *depends*; *depended*; *dependent*
- *elevate*; *elevates*; *elevated*; *elevation*
- *enhance*; *enhances*; *enhanced*
- *increase*; *increases*; *increased*
- *induce*; *induces*; *induced*; *induction*
- *mediate*; *mediates*; *mediated*
- *modulate*; *modulates*; *modulated*; *modulation*

- *necessary*
- *positively regulate*; *positively regulates*; *positively regulated*; *positive regulation*
- *promote*; *promotes*; *promoted*; *promotion*
- *require*; *requires*; *required*
- *regulate*; *regulates*; *regulated*; *regulation*
- *stimulate*; *stimulates*; *stimulated*; *stimulation*
- *transactivate*; *transactivates*; *transactivated*; *transactivation*
- *trigger*; *triggers*; *triggered*
- *up-regulate*; *up-regulates*; *up-regulated*; *up-regulation*
- *[in] response [to]*

Regarding the trigger word, '*mediate*', it is used differently from other trigger words, such as '*activate*' and '*stimulate*', as follows:

A mediates B and C

Here, this sentence may indicate the two following relationships:

- *A => B*
- *A => C*

('=>' indicates 'positive regulation')
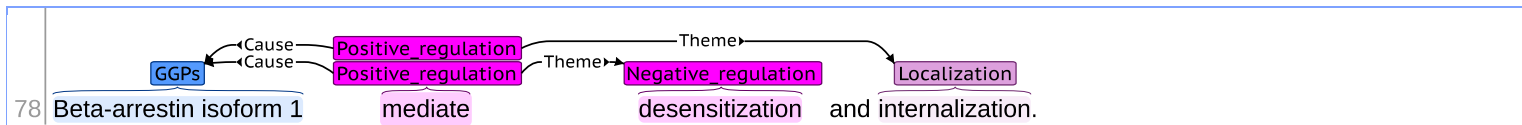
The following sample sentence indicates:

- CXCL12 => phosphorylation of SMAD3
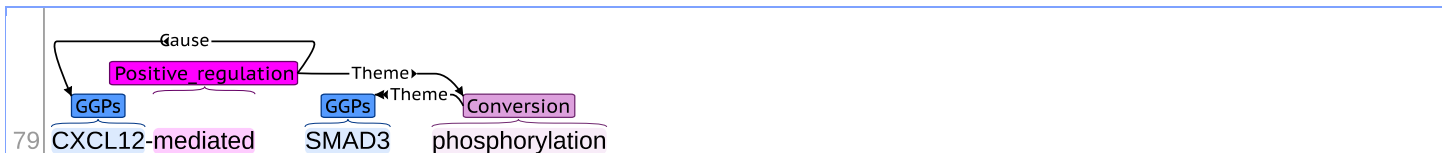
The following sample sentence indicates:

- CXCL12 => expression of CTGF
- *SMAD3 => expression of CTGF*
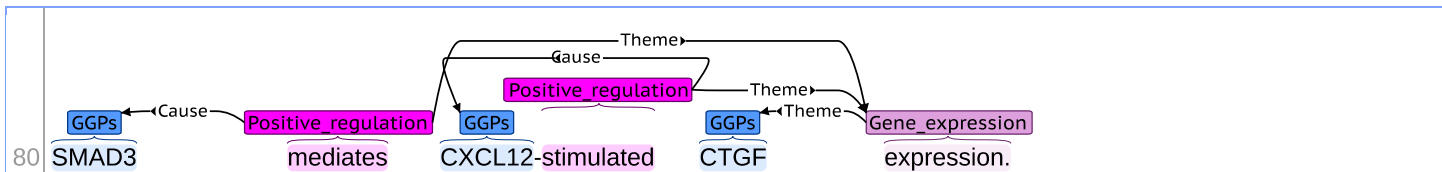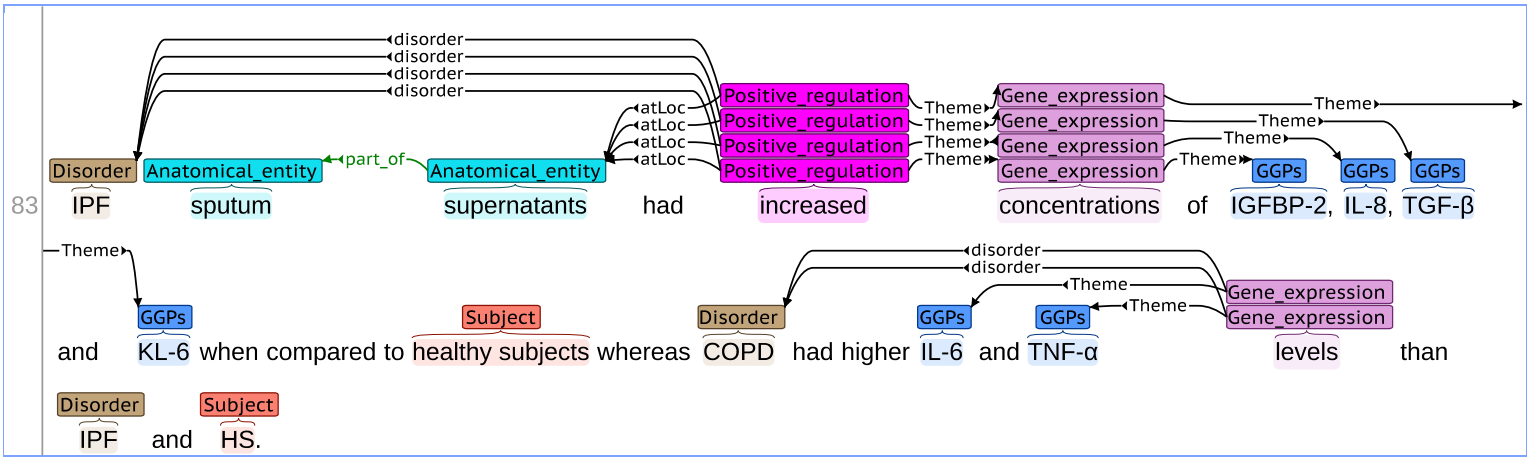

Considering the sample sentences, 2 & 3, the conclusion is:

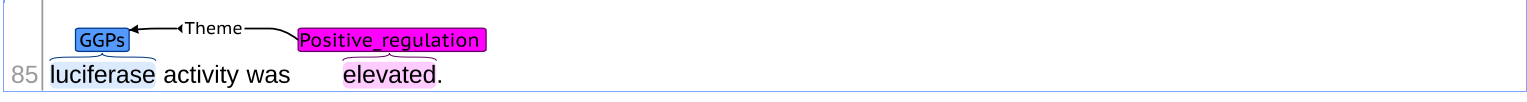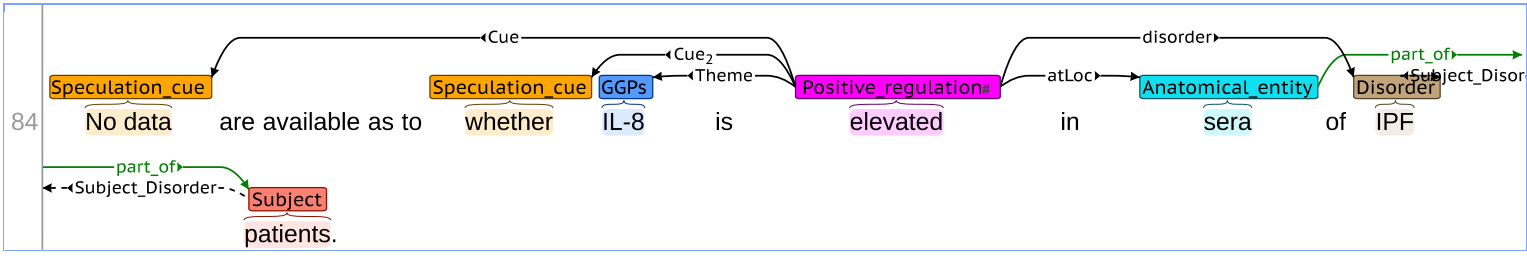CXCL12 => phosphorylation of SMAD3 => expression of CTGF

81 LOXL1/LOXL2 gene expression and protein levels were increased.

82 Protein levels were increased in both IPF and HP for MMP-8, MMP-9, and protein-C, whereas VEGF was decreased compared with control.
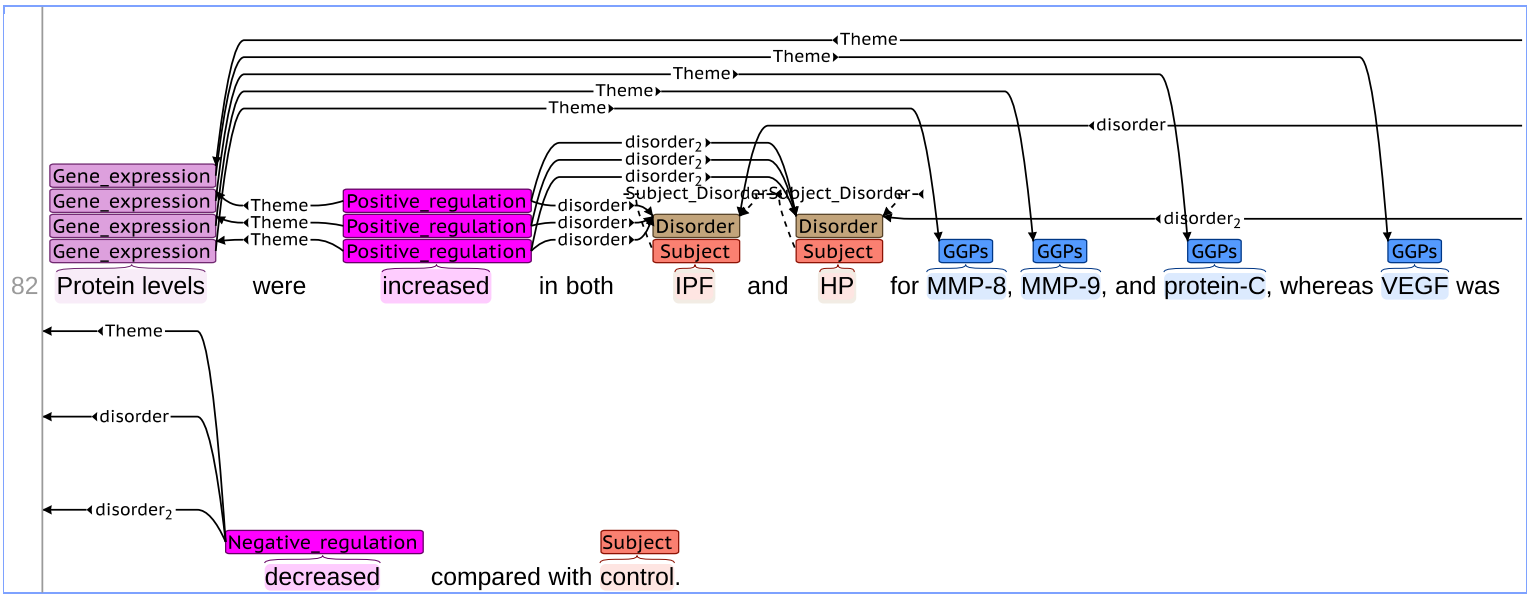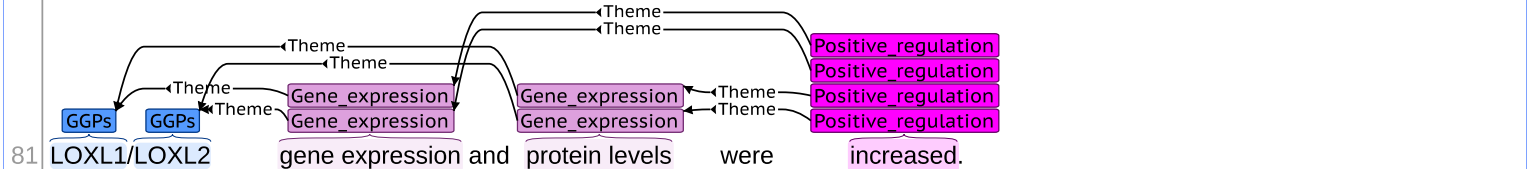
The phrases, *higher/lower*, should not be annotated as Regulation events.

83 IPF sputum supernatants had increased concentrations of IGFBP-2, IL-8, TGF-β and KL-6 when compared to healthy subjects whereas COPD had higher IL-6 and TNF-α levels than IPF and HS.

Occasionally, other events, which could be the *Theme* for this Positive_regulation event, are abbreviated in the sentences. In such cases, molecular entities can be the *Theme* for this event.

84 No data are available as to whether IL-8 is elevated in sera of IPF patients.

85 luciferase activity was elevated.

Other events can be arguments, such as *Theme* and *Cause*.
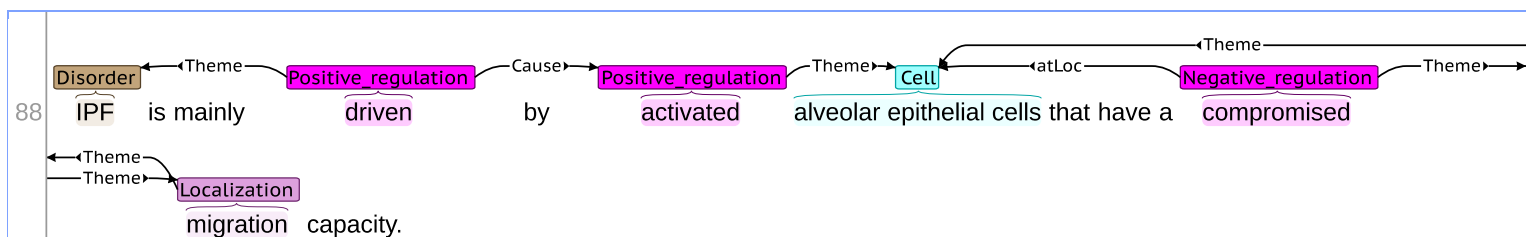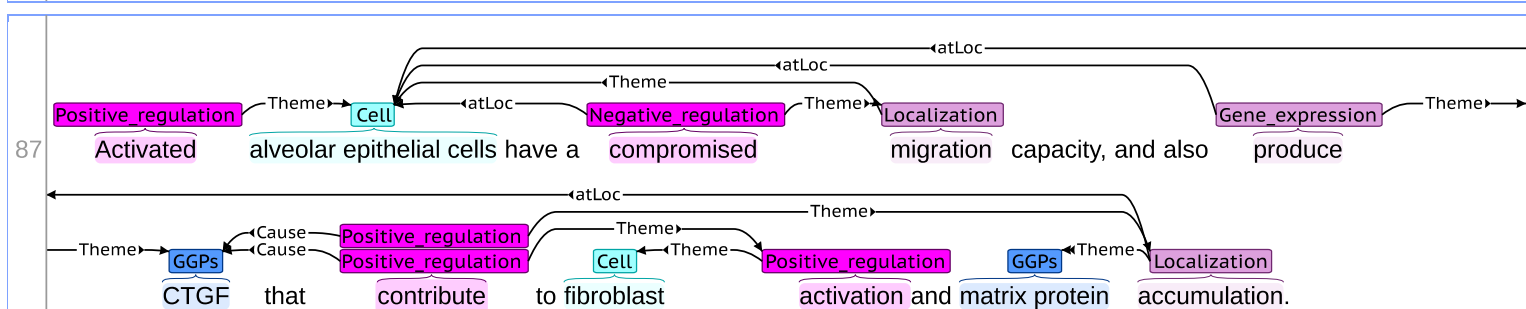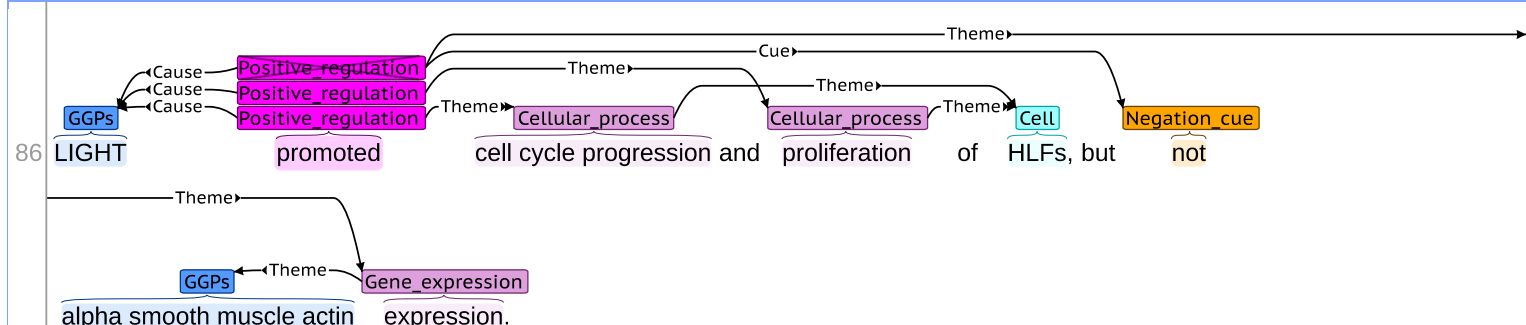
86 LIGHT promoted cell cycle progression and proliferation of HLFs, but not

alpha smooth muscle actin expression.

87 Activated alveolar epithelial cells have a compromised migration capacity, and also produce

CTGF that contribute to fibroblast activation and matrix protein accumulation.

88 IPF is mainly driven by activated alveolar epithelial cells that have a compromised

migration capacity.

Arguments:

The *Theme* (optional; zero or more) indicates events or entities, such as GGPs and Organic_compound_other, that are positively regulated. Function or quality of *Theme* can be affected positively.

The *Cause* (optional; zero or one) indicates events or entities, such as GGPs and Organic_compound_other, that are the stated cause of the `Positive regulation`.

The *atLoc* (optional) indicates the location where the `Positive regulation` event occurs: Cell_component, Cell or Anatomical_entity.

The *disorder* (optional) indicates the Disorder for which the `Positive regulation` event occurs.

The *Cue* argument (optional) is cues, such as Negation cue, Speculation cue or Method cue.

Correlation: Correlation between several entities/events; co-occurence of several events

Category: Biological process

This event describes "Correlation" or "Co-occurence" (UMLS ID:C0332281).

When several events/entities are correlated, this event will be selected. Or, when several events occur simultaneously, those events can be connected with this event.

If the cause and effect (or the causality) are not clear, the relation can be annotated with this `Correlation` event.

The following words/phrases can be triggers of this event:

- *involve*; *involved*; *involves*; *involvement*
- *accompany*; *accompanied*; *accompanies*
- *characterize*; *characterized*; *characterizes*; *characterization*
- *coincide*; *coincided*; *coincides*; *coincident*
- *correlate*; *correlated*; *correlates*; *correlation*
- *concomitant*; *concomitantly*
- *parallel*; *paralleled*; *parallels*
- *relate*; *related*; *relates*; *relation*
- *synergy*

The following words/phrases may be triggers, depending on the situations:

- *associate*; *associated*; *associates*; *association* (These can be also triggers for Binding event)
- *combine*; *combined*; *combines*; *combination* (These can be also triggers for Binding or Conversion)
- *couple*; *coupled*; *couples* (These can be also triggers for Binding or Conversion)
- *link*; *linked*; *links* (These can be also triggers for Binding or Conversion)
- *[play a/an ~] role [in ~ing]* (Depending on "*~ing*")
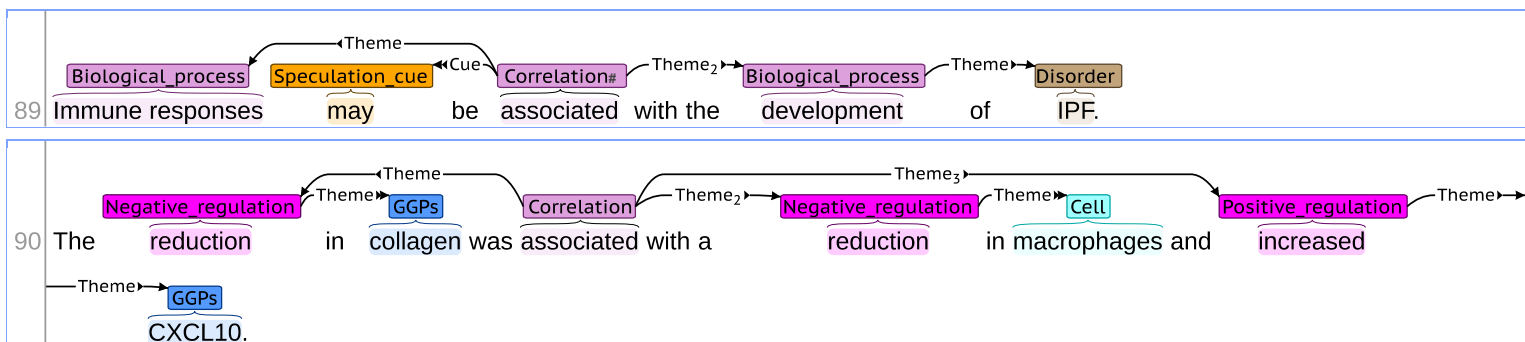
Arguments:

Theme (two or more) indicates events, such as Gene_expression, or entities, such as GGPs and Organic_compound_other, that are related. Function or quality of *Theme* are related.

The *atLoc* (optional) indicates the location where the `Correlation` event occurs: Cell_component, Cell or Anatomical_entity.

The *disorder* (optional) indicates the Disorder for which the `Correlation` event occurs.

The *Cue* argument (optional) is cues, such as Negation cue, Speculation cue or Method cue.
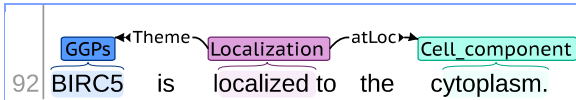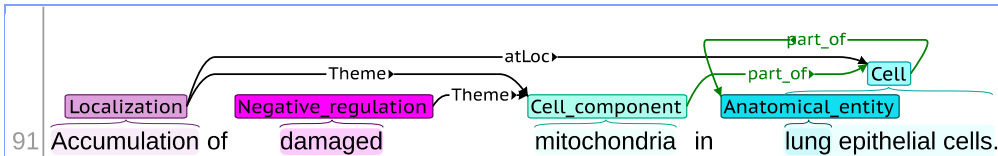
Localization: Localization/movement of entity

Category: Biological process

This event is based on the GENIA-Meta-knowledge corpus at NaCTeM.

This event indicates localization or movement of entities, such as Cell and molecular entities including GGPs.

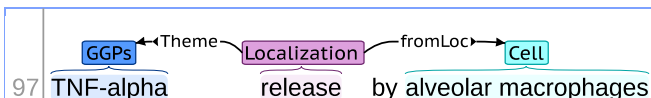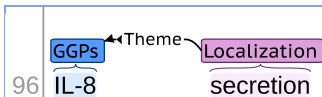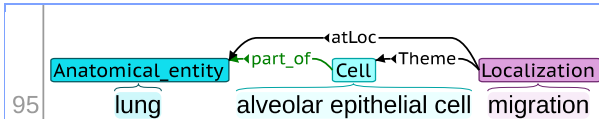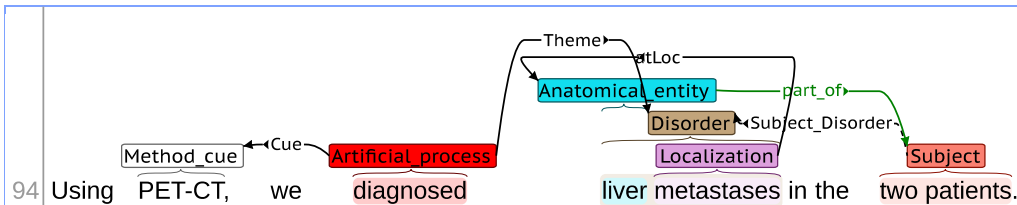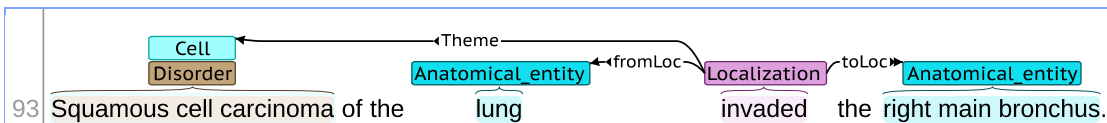The following words/phrases can be triggers of localization event:

- accumulation
- localization, localized (UMLS ID:C0475264 -> C1744691)
- adhesion

91 Accumulation of damaged mitochondria in lung epithelial cells.


92 BIRC5 is localized to the cytoplasm.

The following words/phrases can be triggers of movement event:

- infection (This can be Disorder)
- metastasis
- invade, invasion
- migration
  - cell migration (UMLS ID:C1622501)
- chemotactic activity (UMLS ID:C0008018)
- chemotaxis (UMLS ID:C0008018)
- cellular infiltrate/infiltration
- (cellular) secretion (UMLS ID:C0036536)
  - release (UMLS ID:C0036536)

93 Squamous cell carcinoma of the lung invaded the right main bronchus.

94 Using PET-CT, we diagnosed liver metastases in the two patients.

95 lung alveolar epithelial cell migration

96 IL-8 secretion

97 TNF-alpha release by alveolar macrophages

The following cases, in which the words such as "foci" or "focus" can be annotated as event trigger for this event, might be special: Here, these words may be tentatively assigned the same ID as "localization" (UMLS ID:C0475264).

- fibroblast foci
- fibroblastic foci
- myofibroblast-rich focus core
- fibroblastic focus

98	Cell ←Theme Localization fibroblastic foci
99	Cell ←Theme Localization myofibroblast-rich focus core
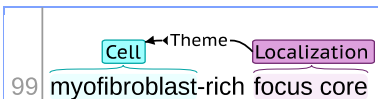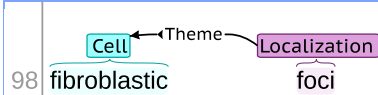
Arguments:

The *atLoc* argument indicates the location at which this event occurs. The *fromLoc* indicates the location from which this event starts, whereas the *toLoc* indicates the location to which this event proceeds. The *atLoc*, *fromLoc* and *toLoc* for this event must be

- Subject
- Anatomical_entity
- Cell
- Cell_component

The other argument, *Theme*, for this event can be any entities, which are targets of this event.

The *disorder* argument must be Disorder.

The *Cue* argument is cues, such as Negation cue, Speculation cue or Method cue.

Cellular_process: Processes on cellular levels

Category: Cellular process

This event is based on the GENIA-Meta-knowledge corpus at NaCTeM.

Among the biological processes, those processes on the cellular levels are categorized into this event.

The following words/phrases can be triggers of this event:

- apoptosis (UMLS; C0162638)
- [cell] differentiation [process] (UMLS; C0007589)
- carcinogenesis (UMLS; C0596263)
- cellular crosstalk (UMLS; C0007582)
- intercellular communication process (UMLS; C0007582)
- cell-cell interaction (UMLS; C0007582)
- cell injury (UMLS ID:C0599732)
- endoplasmic reticulum (ER) stress (UMLS; C3178870)
  - unfolded protein response (UMLS; C1155342)

The following words/phrases related to cell cycle (UMLS; C0007586) can be triggers of this event as well (see Cell cycle):

- cell cycle progression (UMLS; C1516334)
- cell cycle control (UMLS; C1155872)
- G1 [cell cycle] arrest (UMLS; C3178834)

100	**Cellular_process** —Theme→ **Cell** apoptosis of mesenchymal cells
101	**Cellular_process** —Theme→ **Cell** apoptosis in fibroblasts
102	**Cellular_process** —Participant→ **Cell** —Participant₂→ **Cell** Cellular crosstalk between epithelial cells and fibroblasts.
103	**Cell** ←Theme— **Cellular_process** Epithelial cell injury

Arguments:

The *Cause* for this event is entities/events, which cause this event, whereas the *Theme* for this event is entities/events, which are targets of this events. Usually, the *Theme* for this event is Cell.

The *Participant* is entities/events, which are involved in this event.

The *Product* is entities/events, which is produced by this event.

The *atLoc* argument (optional) indicates the location at which this event occurs. The *fromLoc* (optional) indicates the location from which this event starts, whereas the *toLoc* (optional) indicates the location to which this event proceeds.

The *disorder* (optional) indicates the Disorder for which the `Cellular process` event occurs.

The *Cue* argument (optional) is cues, such as Negation cue, Speculation cue or Method cue.

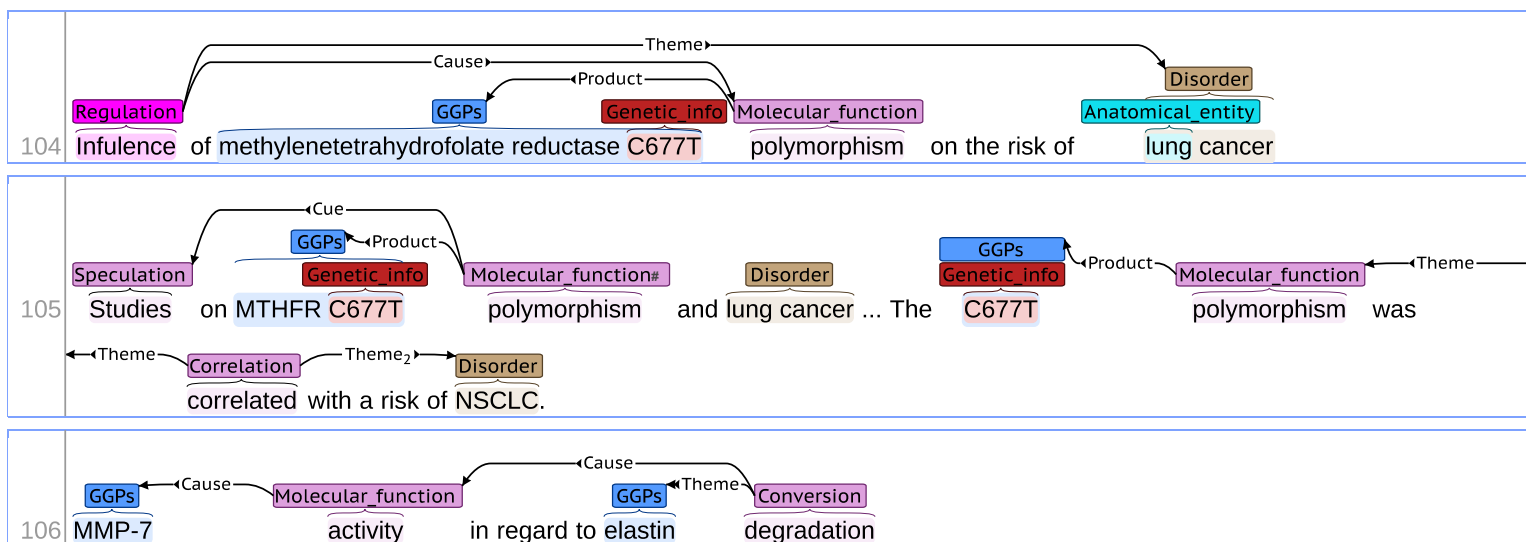# Molecular_function: Processes/functions on molecular level

Category: Molecular function

This event is based on the GENIA-Meta-knowledge corpus at NaCTeM.

This event describes the biological events on the molecular levels.

The following words/phrases can be triggers of this event:

- mutation (if naturally occured) (UMLS ID:C0026882)
  - deletion [mutation] (UMLS ID:C1511760)
- polymorphism (genetic polymorphism) (UMLS ID:C0032529)
- molecular mechanism (UMLS; C0678659)
- biochemical mechanism (UMLS; C0678659)
- [enzyme] activity (UMLS ID:C0243102)
  - catalytic activity (UMLS ID:C0243102)
- biosynthesis (anabolism, or biosynthetic process) (UMLS; C0220781)
- degradation (catabolism, or breakdown) (UMLS; C0699900)
- chemoattractant activity (UMLS; C1149381) (If chemotaxis, Cellular_process)
- immunoreactivity (UMLS; C0597879) (If it meant immunoassay, Method cue)
- (indirect) interaction (UMLS; C1148560 (originally for Molecular function))(in case where it is difficult to

determine whether it is a direct interaction between molecules Binding)

104	Infulence of methylenetetrahydrofolate reductase C677T polymorphism on the risk of lung cancer
105	Studies on MTHFR C677T polymorphism and lung cancer ... The C677T polymorphism was correlated with a risk of NSCLC.
106	MMP-7 activity in regard to elastin degradation

Arguments:

The *Cause* for this event is entities/events, which cause this event, whereas the *Theme* for this event is entities/events, which are targets of this events. Usually, the *Theme* for this event is molecular entities.

- GGPs
- Pharmacological_substance
- Organic_compound_other
- Inorganic_compound

The *Participant* is molecular entities, which are involved in this event.

The *Product* is molecular entities, which are produced by this event.

The *atLoc* argument (optional) indicates the location at which this event occurs. The *fromLoc* (optional) indicates the location from which this event starts, whereas the *toLoc* (optional) indicates the location to which this event proceeds.

The *atLoc*, *fromLoc* and *toLoc* for this event can be

- Cell,
- Cell_component

The *disorder* (optional) indicates the Disorder for which the `Molecular function` event occurs.

The *Cue* argument (optional) is cues, such as Negation cue, Speculation cue or Method cue.

Pathway: Molecular pathway networks (metabolism, signaling)

Category: Molecular function

This event describes molecular pathway networks, such as metabolism and signaling pathways.

The following words/phrases can be triggers of this event:

- metabolic pathway (UMLS ID:C1291081)

- ○ metabolism
- signaling pathway (UMLS ID:C0037080)
  - ○ signaling
  - ○ signaling cascade
  - ○ signalling
  - ○ [protein] signaling
  - ○ [protein] signaling pathway
  - ○ [cell] signaling pathways
- molecular pathway (UMLS ID:C1706062)
  - ○ pathway[s] (UMLS ID:C1706062)

The following examples are more specific pathways:

- EGFR signaling (UMLS ID:C3271839)
- ErbB signaling pathway (UMLS ID:C2984323)
- MAPK signaling (UMLS ID:C1518102)
- PDGFR signaling (UMLS ID:C1155400)
- PI3K signaling (UMLS ID:C1817666)
- VEGFR2 signaling pathway (UMLS ID:C3549205)
- integrin signaling (UMLS ID:C1512812)
- mTOR signaling (UMLS ID:C1515673)

107 MAPK signaling pathway

108 AKT2-mediated signaling pathway
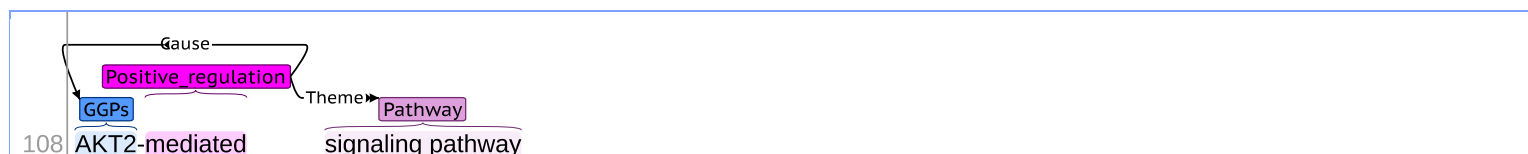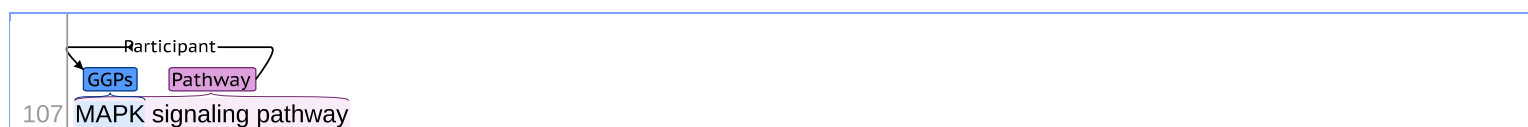
Occasionally, more than one participant molecule can be included in this event.

109 MEKK1/MEK1/ERK1 and AP-1 pathway
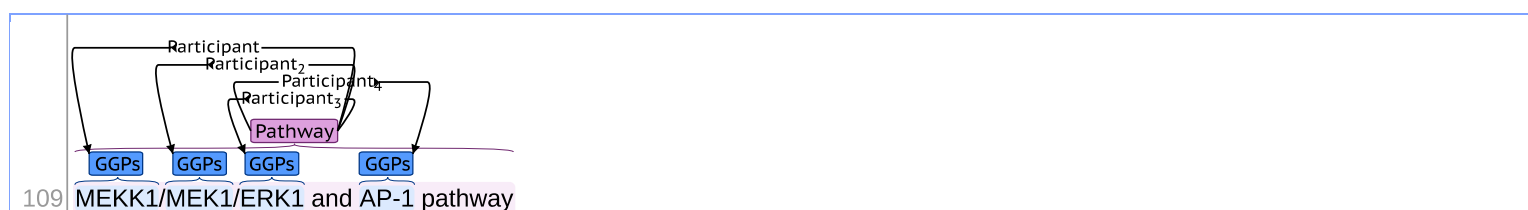
Arguments:

The *Participant* (optional; zero or more) for this event must be molecular entities involved in the `Pathway`, such as GGPs, Organic_compound_other or Pharmacological_substance.

The *atLoc* (optional; zero or one) indicates the location, such as Cell_component, Cell or Anatomical_entity, where the `Pathway` event occurs.

The *disorder* (optional) indicates the Disorder for which the `Pathway` event occurs.

The *Cue* argument is cues, such as Negation cue, Speculation cue or Method cue.

Conversion: Changes in covalent bonds of molecular entities
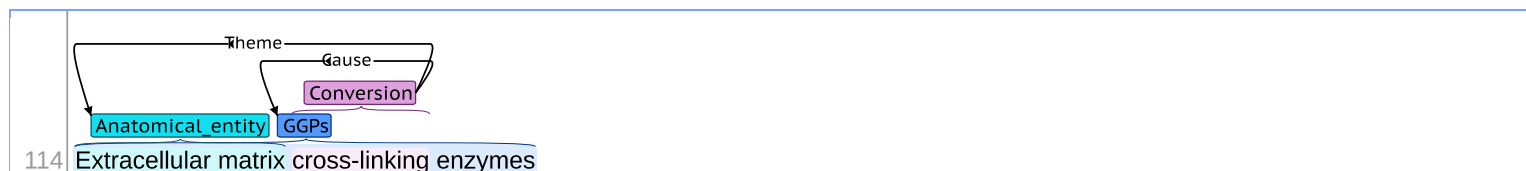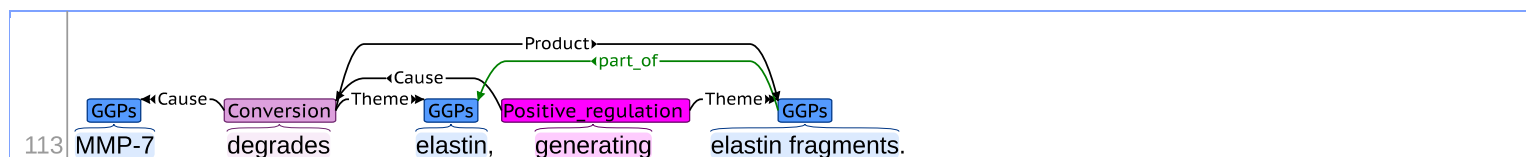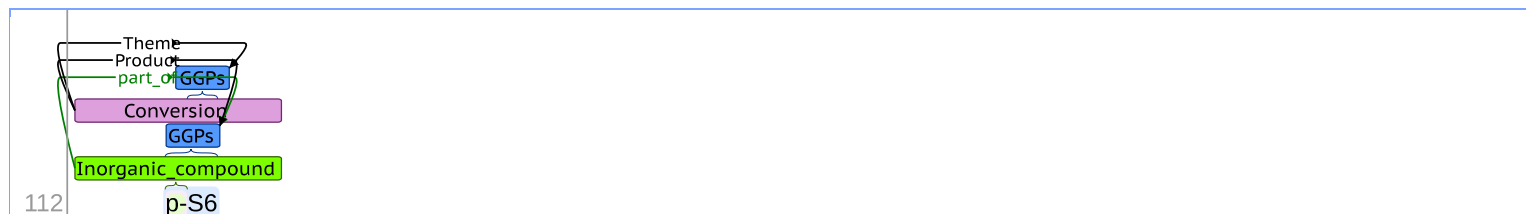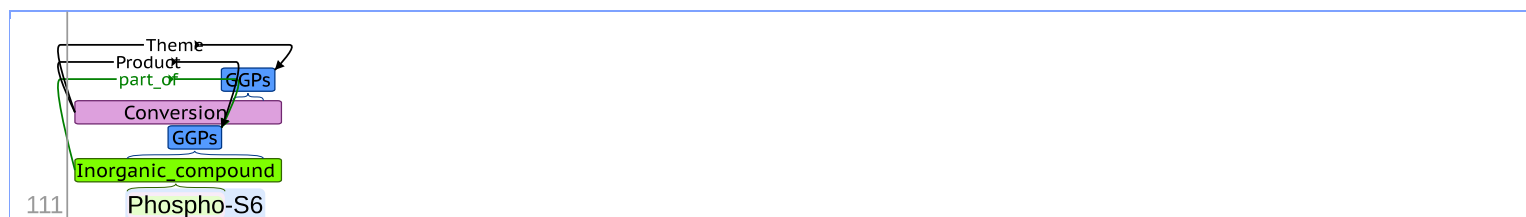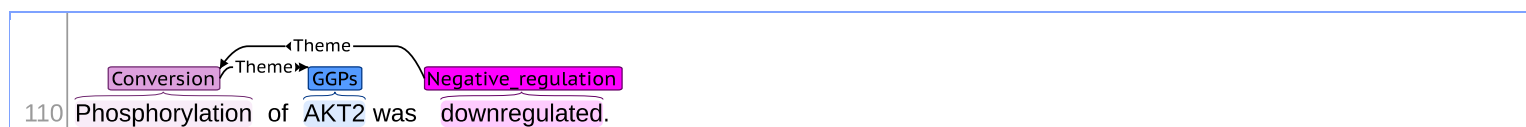
Category: Molecular function

This event describes the changes in covalent bonds of molecules, including modification of protein molecules and cleavage of covalent bonds.

The following words/phrases can be triggers of this event:

- *phosphorylate*, *phosphorylates*, *phosphorylated*, *phosphorylation(s)* (UMLS ID:C0031715)
    - autophosphorylation (UMLS ID:C0813988)
    - self-phosphorylation (UMLS ID:C0813988)
    - protein phosphorylation (UMLS ID:C0031715)
    - tyrosine autophosphorylation (UMLS ID:C0813988)
    - tyrosine phosphorylation (UMLS ID:C0031715)
    - hyperphosphorylation (UMLS ID:C0031715)
- *dephosphorylate*, *dephosphorylates*, *dephosphorylated*, *dephosphorylation* (UMLS ID:C3160734)
- *methylate*, *methylates*, *methylated*, *methylation* (UMLS ID:C0025723)
- *acetylate*, *acetylates*, *acetylated*, *acetylation* (UMLS ID:C0001038)
- *carboxymethylate*, *carboxymethylates*, *carboxymethylated*, *carboxymethylation* (UMLS ID:C0596262)
- *cleave*, *cleaves*, *cleaved*, *cleavage* (UMLS ID:C0596311)
    - *degrade*, *degrades*, *degraded*, *degradation* (UMLS ID:C0596311)
- *cross-link*, *cross-links*, *cross-linked*, *cross-linking* (UMLS ID:C0332220)

The following ones can also be trigger words:

- [enzyme] hydrolysis (UMLS ID:C0020291)
- proteolysis (UMLS ID:C0597304)

110 Phosphorylation of AKT2 was downregulated.

111 Phospho-S6

112 p-S6

113 MMP-7 degrades elastin, generating elastin fragments.

114 Extracellular matrix cross-linking enzymes

Arguments:

The *Theme* for this event must be mostly molecular entities whose covalent bonds are converted: GGPs, Organic_compound_other or possibly Pharmacological_substance. However, the other entities, such as Cell, Cell_component and Anatomical_entity, which are composed of molecules to be converted, can also be *Theme* for this event.

The *Cause* (optional) for this event is usually enzyme proteins (GGPs) or molecular events, which cause the `Conversion` event.

The *Product* (optional) indicates the molecule(s) that could be produced by the `Conversion` event: GGPs or Organic_compound_other.

The *atLoc* (optional) indicates the location where the target molecules, to which the `Conversion` event occurs, exist: Anatomical_entity, Cell, and Cell_component.

The *disorder* (optional) indicates the Disorder for which the `Conversion` event occurs.

The *Cue* argument is cues, such as Negation cue, Speculation cue or Method cue.

Gene_expression: Gene expression

Category: Molecular function

This event describes *gene expression* (UMLS ID:C0017262), which can be either *transcription* (UMLS ID:C0040649) or *translation* (UMLS ID:C1519614), or both the events.

The following words/phrases can be triggers of *transcription* (UMLS ID:C0040649):

- *transcribe*; *transcribes*; *transcribed*; *transcription*
- *[mRNA] expression*
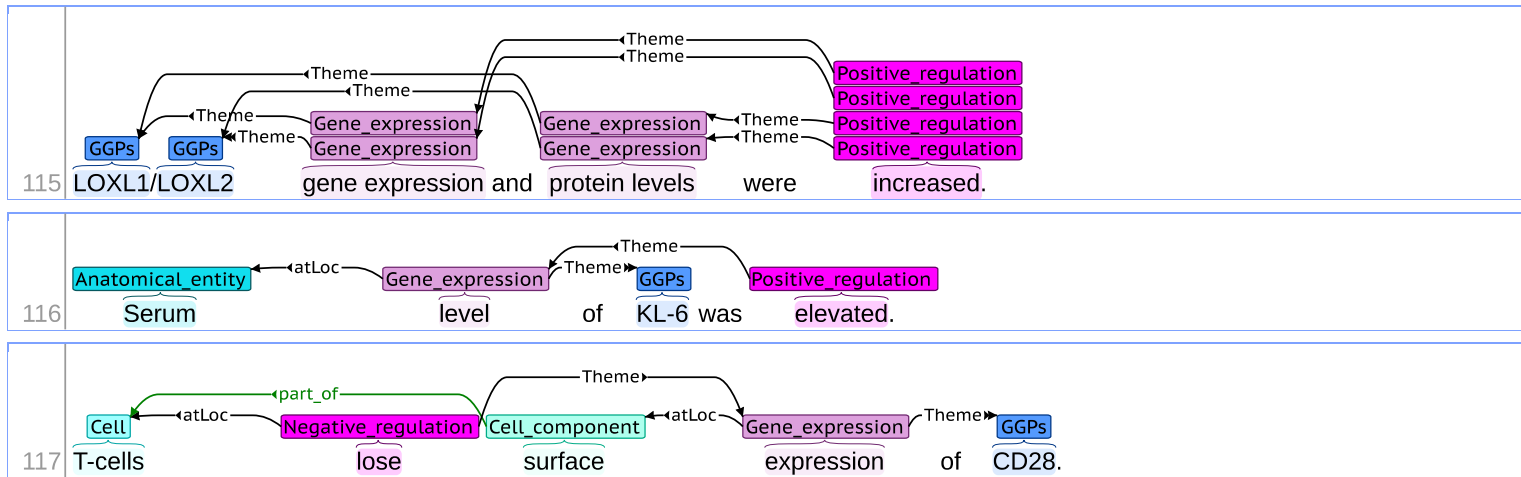- *[mRNA] production*
- *[mRNA] synthesis*

The following words/phrases can be triggers of *translation* (UMLS ID:C1519614):

- *translate*; *translates*; *translated*; *translation*
- *[protein] biosynthesis*
- *[protein] formation*
- *[protein] expression*
- *[protein] synthesis*
- *protein level [increased/decreased]*

The following words/phrases can be triggers of *gene expression* (UMLS ID:C0017262), if it is not clear whether *transcription* or *translation* is stated.

- *co-express*; *co-expresses*; *co-expressed*; *co-expression*
- *coexpress*; *coexpresses*; *coexpressed*; *coexpression*
- *hyperexpress*; *hyperexpresses*; *hyperexpressed*; *hyperexpression*
- *express*; *expresses*; *expressed*; *expression*
- *overexpress*; *overexpresses*; *overexpressed*; *overexpression*
- *produce*; *produces*; *produces*; *production*,

- *synthesize*; *synthesizes*; *synthesized*; *synthesis*
- *up-expression*

115 | LOXL1/LOXL2 gene expression and protein levels were increased.

116 | Serum level of KL-6 was elevated.

117 | T-cells lose surface expression of CD28.

Arguments:

The *Theme* for this event must be genes/gene products: GGPs.

No *Cause* is annotated for this event.

The *atLoc* (optional) indicates the location where the `Gene expression` event occurs : Anatomical_entity, Cell, and Cell_component.

The *disorder* (optional) indicates the Disorder for which the `Gene expression` event occurs.

The *Cue* argument (optional) is cues, such as Negation cue, Speculation cue or Method cue.

Binding: Non-covalent interaction between molecules

Category: Molecular function

This event describes "non-covalent interaction" between molecules, such as GGPs, Organic_compound_other, Pharmacological_substance and Inorganic_compound. Moreover, this event is opposite to Dissociation.

However, covalent bond formation will be categorized in Conversion.

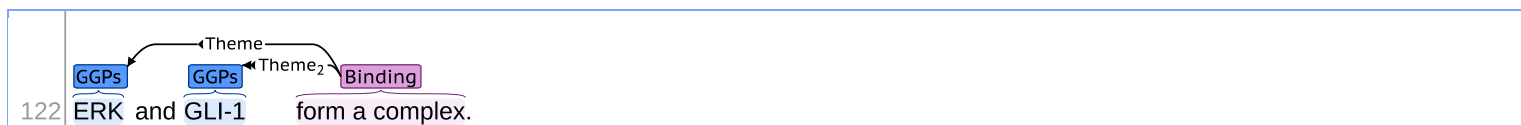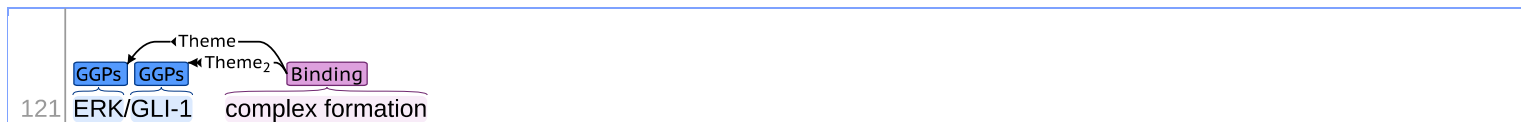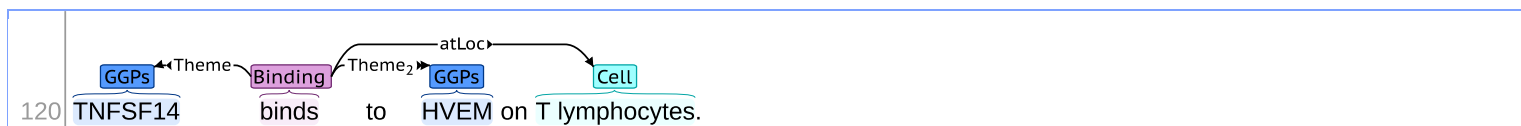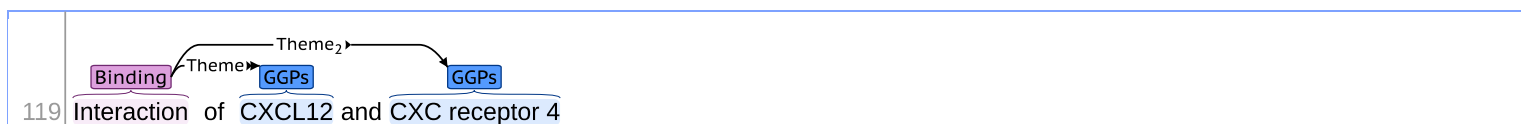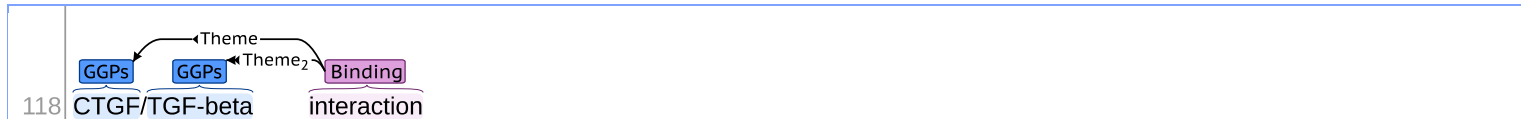Moreover, interaction between molecules and Cell/Cell_component will be categorized in Localization.

Interaction between Cell/Cell_component will be categorized in Cellular_process.

UMLS ID for this event; Molecular Interaction Process (UMLS; C1167622).

The following words/phrases can be triggers of this event:

- *adhere*; *adheres*; *adhered*; *adhesion*
- *affinity*; *affinities*
- *assemble*; *assembles*; *assembled*; *assembly*
- *associate*; *associates*; *associated*; *association*
- *attach*; *attaches*; *attached*; *attachment*

- *bind*; *binds*; *bound*; *binding*
- *conjugate*; *conjugates*; *conjugated*
- *couple*; *couples*; *coupled*; *coupling*
- *dimer*; *dimers*
- *dimerize*; *dimerizes*; *dimerized*; *dimerization*
- *dock*; *docks*; *docked*; *docking*
- *form [a] complex*; *complex formation*
- *interact*, *interacted*, *interacts*, *interaction*
- *oligomerize*; *oligomerizes*; *oligomerized*; *oligomerization*
- *polymerize*; *polymerizes*; *polymerized*; *polymerization*

118 CTGF/TGF-beta interaction

119 Interaction of CXCL12 and CXC receptor 4

120 TNSF14 binds to HVEM on T lymphocytes.

121 ERK/GLI-1 complex formation

122 ERK and GLI-1 form a complex.

Arguments:

The *Theme* (optional; multiple) indicates molecular entities such as GGPs, Organic_compound_other, Pharmacological_substance, and Inorganic_compound.

No *Cause* is annotated for this event.

The *atLoc* (optional; zero or one) indicates the location where `Binding` event occurs: Cell_component, Cell or Anatomical_entity.

The *disorder* (optional) indicates the Disorder for which the `Binding` event occurs.

The *Cue* argument (optional) is cues, such as Negation cue, Speculation cue or Method cue.

Dissociation: Dissociation of multiple molecules

Category: Molecular function

This event describes "Dissociation" of molecules, such as GGPs, Organic_compound_other, Pharmacological_substance and Inorganic_compound, which had interacted with each other, as described in Binding. Thus, this event is opposite to Binding.

UMLS ID for this event; Molecular dissociation (UMLS; C0301643).

Moreover, dissociation between molecules and Cell/Cell_component will be categorized in Localization.

Dissociation between Cell/Cell_component will be categorized in Cellular_process.

The following words/phrases can be triggers of this event:

- *disaggregate*, *disaggregates*, *disaggregated*, *disaggregation*
- *disassemble*, *disassembles*, *disassembled*, *disassembly*
- *disrupt*, *disrupts*, *disrupted*, *disruption*
- *dissociate*, *dissociates*, *dissociated*, *dissociation*
- *free*, *frees*; *freed*
- *release*, *releases*, *released*, *releasing*

Arguments:

The *Theme* (optional; multiple) indicates molecular entities, such as GGPs, Organic_compound_other, Pharmacological_substance, and Inorganic_compound.

No *Cause* is annotated for this event.

The *atLoc* (optional) indicates the location where the `Dissociation` event occurs: Cell_component, Cell or Anatomical_entity.

The *disorder* (optional) indicates the Disorder for which the `Dissociation` event occurs.

The *Cue* argument (optional) is cues, such as Negation cue, Speculation cue or Method cue.
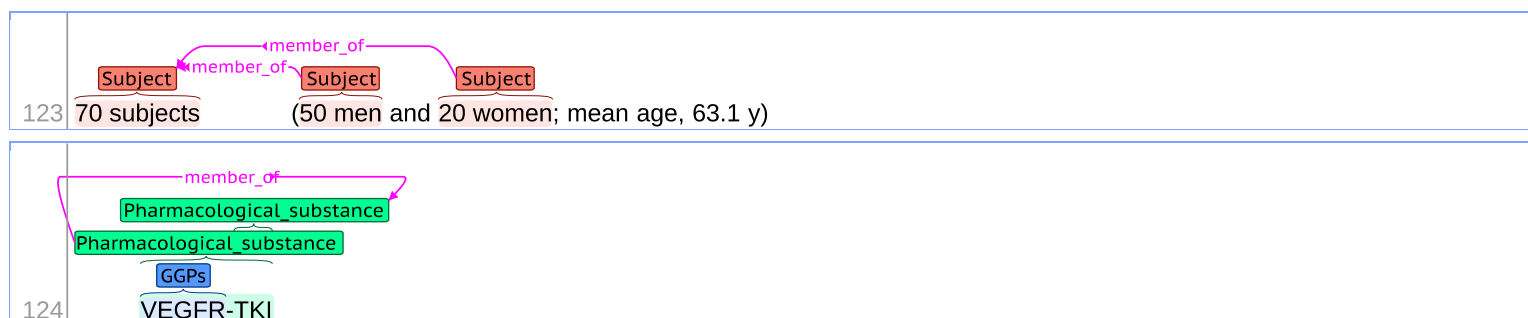
Relations

In addition to the relations between event triggers and their arguments (*Theme*, *Cause*, etc.), which are described above, several types of relations were defined.
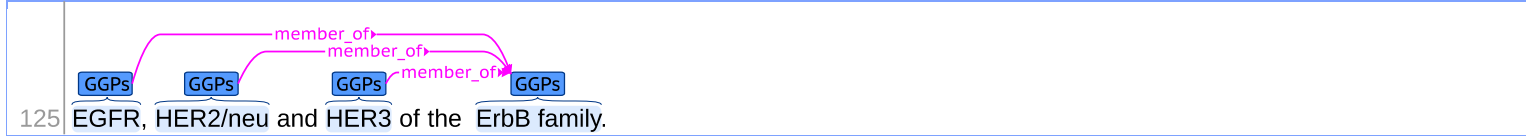
Those relations are defined as follows:

member_of: Relationships between member and its group

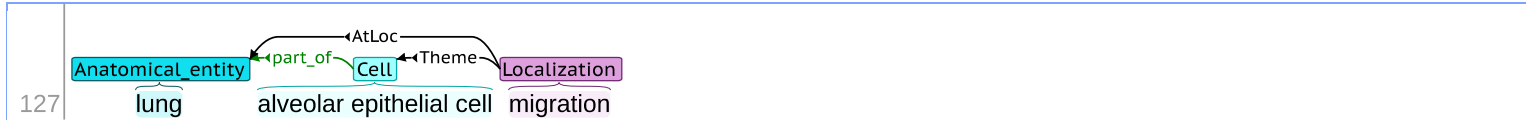This relation describes relationships of a member with a group, to which the member belong.

This relaction can be used to describe the relation of a protein with its protein family.

125 EGFR, HER2/neu and HER3 of the ErbB family.

member_of *member_of* *member_of*

GGPs GGPs GGPs GGPs

part_of: Relationships between whole and part

This relation indicates the relationships between whole entity and its part.

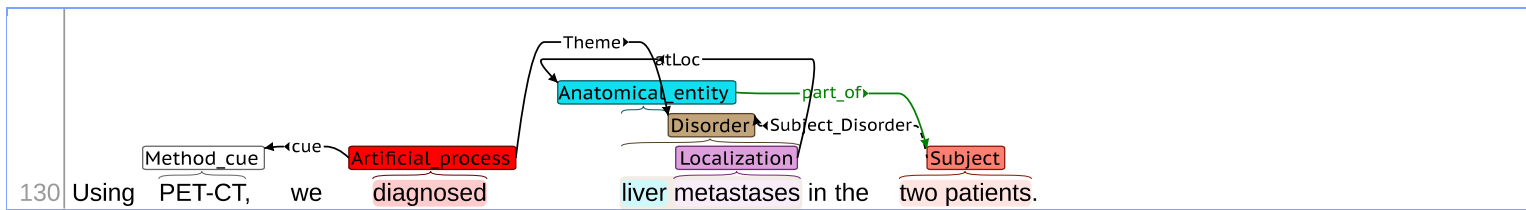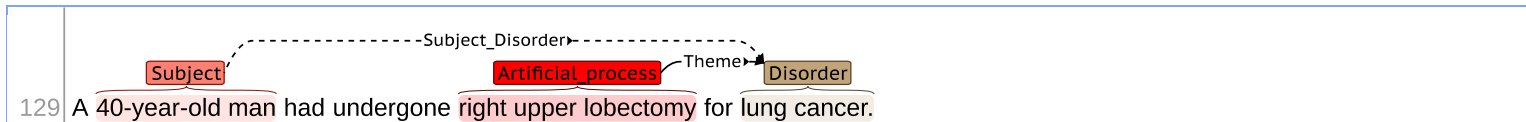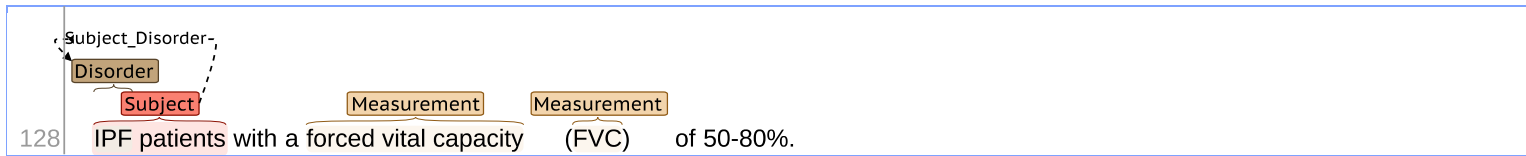Typical examples are relationships between Cell and its Anatomical_entity, or Cell_component and Cell.

126 Mitochondria in lung epithelial cells.

Cell_component — part_of → Anatomical_entity ← part_of — Cell

127 lung alveolar epithelial cell migration

Anatomical_entity ← part_of — Cell — Theme — Localization
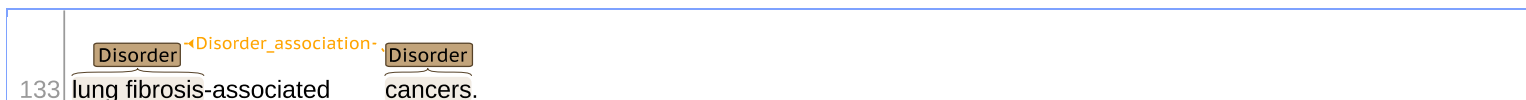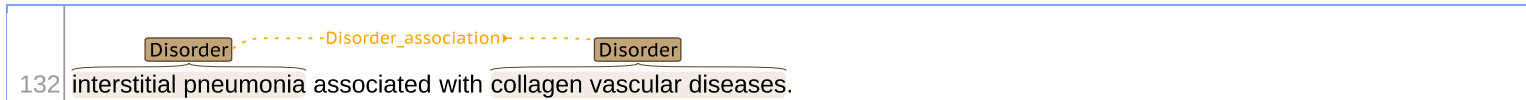AtLoc

Subject_Disorder: relationships between subject and disorder (PHAEDRA)

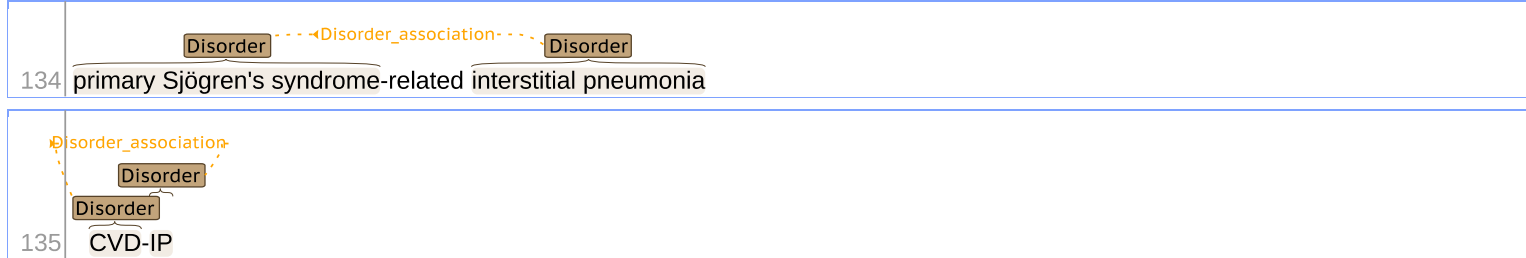This relation indicates relationships of Subject with its Disorder.

This relation is based on the PHAEDRA corpus at NaCTeM.

128 IPF patients with a forced vital capacity (FVC) of 50-80%.

Subject_Disorder — Disorder — Subject — Measurement — Measurement

129 A 40-year-old man had undergone right upper lobectomy for lung cancer.

Subject — Subject_Disorder → Artificial_process — Theme → Disorder

130 Using PET-CT, we diagnosed liver metastases in the two patients.

Method_cue ← cue — Artificial_process — Theme → Anatomical_entity — part_of → Subject
AtLoc — Disorder ← Subject_Disorder — Localization

Disorder_association: relationships between two disorders

When two different `Disorders` (Disorder) occur simultaneously, these `Disorders` can be connected with this relation.

131 IPF-associated pulmonary hypertension

Disorder ← Disorder_association — Disorder

132 interstitial pneumonia associated with collagen vascular diseases.

Disorder — Disorder_association → Disorder

133 lung fibrosis-associated cancers.

Disorder ← Disorder_association — Disorder

134 primary Sjögren's syndrome-related interstitial pneumonia


135 CVD-IP

Attributes

Attributes indicate the states/characteristics of events or entities.

Those attributes are defined as follows:

Negated: Negation for Events

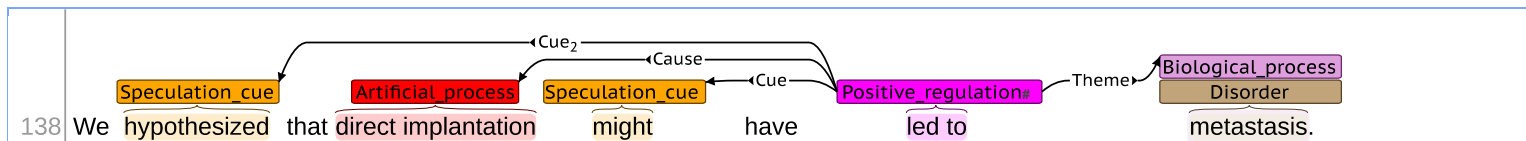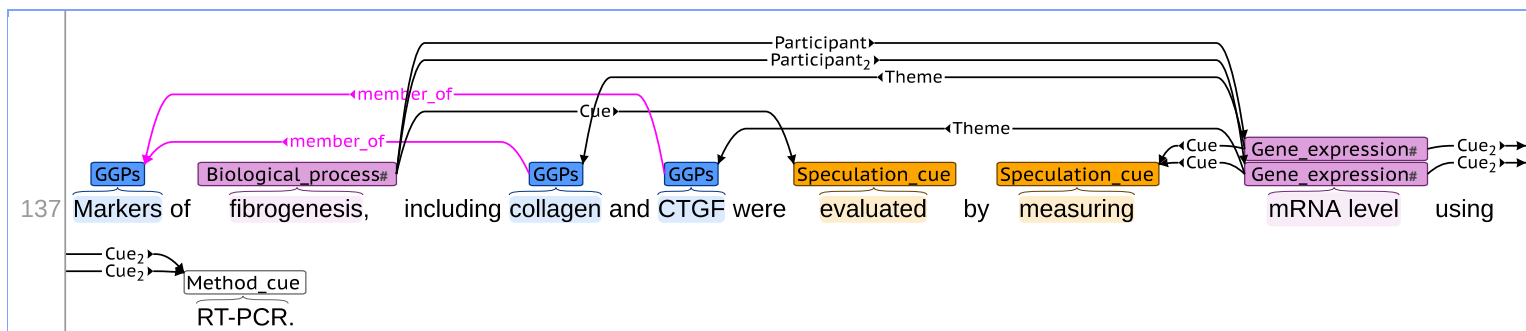The `Negated` attribute is for event triggers that are negated.

In the following case, the event trigger, which is connected with the Negation cue, is negated, and indicated with a cross:

136 The patient did not undergo lung resect.
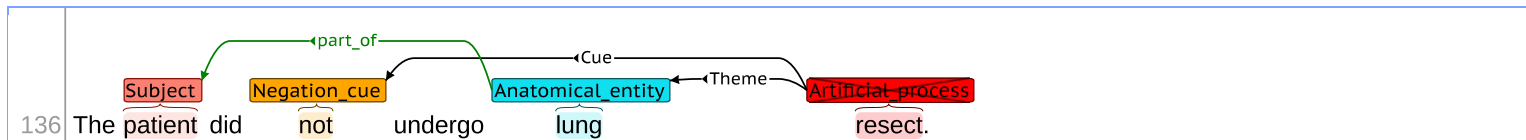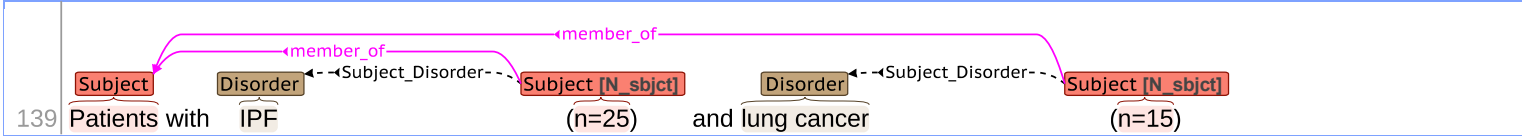
Speculated: Speculation attributes for events

The `Speculated` attribute is for event triggers for which the `Speculated` states may be suggested.

Event triggers, which are connected with Speculation cue or Method cue, should be ticked off with `Speculated` attributes.

137 Markers of fibrogenesis, including collagen and CTGF were evaluated by measuring mRNA level using RT-PCR.

138 We hypothesized that direct implantation might have led to metastasis.

Count_sbjct: Attribute for Subject with N count

The `Count_sbjct` attribute is for the Subject entity that is expressed "N= ***".

139 | Patients with IPF (n=25) and lung cancer (n=15)

member_of · member_of · Subject_Disorder · Subject_Disorder
Subject · Disorder · Subject [N_sbjct] · Disorder · Subject [N_sbjct]

The following case may have two possible annotations:

140 | Subject [N_sbjct]
controls (N=24)

141 | Subject Subject [N_sbjct]
controls   (N=24)