

KEEPHA Annotation Guidelines

Version 03/24

Lisa Raithel*, Shuntaro Yada*,
Hui-Syuan Yeh, Cyril Grouin, Thomas Lavergne,
Aurélie Névéol, Patrick Paroubek, Philippe Thomas,
Tomohiro Nishiyama, Sebastian Möller, Eiji Aramaki,
Yuji Matsumoto, Roland Roller, Pierre Zweigenbaum

March 22, 2024

Contents

1	Overview	3
2	Entities	4
2.1	General Guidelines for Entity Annotation	4
2.2	Drug	7
2.3	Disorder	9
2.4	Function	11
2.5	Anatomy	11
2.6	Test	12
2.7	Opinion	12
2.8	Measure	13
2.9	Time	14
2.10	Route	15
2.11	Doctor	15
2.12	Entities for De-Identification	15
2.13	Other	16
3	Relations	18
3.1	caused	19
3.2	treatment_for	20
3.3	has_dosage	20
3.4	experienced_in	21
3.5	examined_with	21
3.6	has_result	22
3.7	interacted_with	22
3.8	signals_change_of	23

*Equal contribution; Corresponding author: raithel@tu-berlin.de

3.9	has_time	23
3.10	has_route	23
3.11	is_opinion_about	24
3.12	refers_to	24
3.13	misc	25
4	Difficult Cases	25
4.1	Entailment	25
4.2	Advice from other Users	25
4.3	Distinction of Disorder, Function, and Opinion for sentiment expressions	26
4.4	Disorder expressions using general words	26
4.5	English	27
4.6	French	27
4.7	German	27
4.8	Japanese	28

1 Overview

In the following, we describe the annotation scheme we developed for our multi-lingual corpus KEEPHA. The data for this corpus consists of different sources, for example, patient fora like Yahoo Q&A or social media like Twitter (X).

Our annotation scheme includes entities, attributes, and relations. First, we define the annotated entities (Section 2), including possible attributes. Moreover, different examples of their usage are presented for each entity. Subsequently, we define the relations connecting these entities (Section 3).

Examples marked with * denote sentences that have been translated from a different language, and sentences marked with ** denote ‘artificial’ examples that have been completely made up. All other examples are taken from existing corpora. English examples are extracted from the CADEC corpus¹ (?), while German examples are taken from the German patient forum Lifeline². French examples are translated from German using DeepL³ and manual verification. Lastly, the Japanese examples are taken from social media (Twitter) and forum (Yahoo! JAPAN Q&A) posts. Note that sometimes, the translated sentences do not perfectly fit the original translations of the target languages. However, we keep the original annotation of the target language.

You can toggle the translations of the non-English examples by clicking [here](#)⁴:

[show/hide translations](#)

section	entity	attributes
2.2	drug _{drug}	increase, decrease, stopped, started, unique_dose
2.2.1	change_trigger _{change_trigger}	
2.3	disorder _{disorder}	negated
2.4	function _{function}	negated
2.5	anatomy _{anatomy}	
2.6	test _{test}	
2.7	opinion _{opinion}	positive, negative, neutral
2.8	measure _{measure}	
2.9	time _{time}	frequency, duration, date, point
2.10	route _{route}	
2.11	doctor _{doctor}	
2.12	user _{user}	
2.12	url _{url}	
2.12	personal_information _{perso_info}	
2.13	other _{other}	

Table 1: Overview of the entities and their attributes.

¹<https://pubmed.ncbi.nlm.nih.gov/25817970/>

²<https://fragen.lifeline.de/forum/>

³<https://www.deepl.com/translator>

⁴This feature is unfortunately not supported by all PDF viewers.

2 Entities

In Table 1, we show all available entities and their respective attributes. Attributes are used to add more precise semantics to some of the entity annotations.

2.1 General Guidelines for Entity Annotation

We now define general rules for the annotation of entities.

Scope

Regarding scope, we annotate entities in the form of noun or verb phrases together with their modifying parts, e.g., adjectives, adverbs, etc., though complex post modifiers (e.g., relative clauses: “a pain that is ...”, and prepositional-phrase modifiers: “a neck pain from yesterday”) are excluded. We always prefer the smallest core noun phrase. If this does not work, a long span of the whole verb phrase (or even the whole clause/sentence) is allowed.

- (1) (en) I have **absolutely no appetite_{disorder}** and **constantly feel sick_{disorder}** to my **stomach_{anatomy}** .
(fr) (...) ça fait **drolement mal_{disorder}** et c'est **très gênant_{disorder}** .
[It **hurts awfully_{disorder}** , and it is a **great discomfort_{disorder}**]
(de) (...) ich hatte in der Situation auch **absolut keinen Hunger_{disorder}** .
[I was also **absolutely not hungry_{disorder}** in this situation.]
(ja) ** 激しい嘔吐_{disorder} に見舞われました
[I suffered **severe vomiting_{disorder}** .]

Metaphors and descriptive language

Further, we annotate descriptive expressions based on the *patient's perspective* even if they do not occur in a medical dictionary. We always try to follow the patient's perspective: If the patient thinks something (e.g., a symptom, an irregularity) is a disorder, then we annotate it as **disorder**.

- (2) (en) (...) I felt like I would imagine someone that had **MS_{disorder}** would feel like, **swaying_{disorder}** , **pulling myself up stairs_{disorder}** , (...)
(fr) (...) J'ai **failli tomber dans les pommes_{disorder}** , (...)
[I **almost passed out_{disorder}**]
(de) Die **Beine_{anatomy}** **fühlen sich nicht stark_{disorder}** man hat **Gefühle des Ungleichgewichts_{disorder}** .
[The **legs_{anatomy}** **do not feel strong_{disorder}** , there are **feelings of imbalance_{disorder}**]
(ja) ずっと お腹_{anatomy} が **ゴロゴロいってる_{disorder}**
[My **stomach_{anatomy}** **keeps saying grrrr_{disorder}**]

Often, patients use metaphors to describe their suffering or, on the other side, a good experience with medication therapy. Since metaphors are descriptive language as well, we annotate them as we would annotate explicit symptom descriptions. Later, we can re-investigate those “special disorders”.

- (3) (en) I felt always like I had a **veil over my head_{disorder}** and **lead in my legs_{disorder}** .

- (fr) (...) à chaque effort, mon **cœur_{anatomy}** bat la chamade_{disorder} et j'ai ne n'ai pas beaucoup de souffle_{disorder} .
 [With every effort, my **heart_{anatomy}** is **racing_{disorder}** and I am **out of breath_{disorder}**]
- (de) (...) nur 1 Woche später^{rel. point in time}_{time} war ich wie im 7. Himmel^{positive}_{opinion} .
 [Just **one week later^{rel. point in time}_{time}** I was **on cloud nine^{positive}_{opinion}** .]
- (ja) * マイスリー_{drug} を飲むと 脳_{anatomy} が 強制終了_{disorder} する
 [Zolpidem_{drug} makes my **brain_{anatomy}** **shut down_{disorder}**]

Determiners and Possessives

We do not include determiners or possessive pronouns in an entity.

- (4) (en) * I only really felt the **burning sensation_{disorder}** in the **arms_{anatomy}** and **chest_{anatomy}** when I ...
- (fr) (...) J'ai développé une **très forte sinusite chronique_{disorder}** ,(...)
 [I have developed **severe chronic sinusitis_{disorder}**]
- (de) Seit zwei Tagen^{duration}_{time} ist meine **Übelkeit^{negated}_{disorder}** wieder weg, doch da (...)
 [Since two days^{duration}_{time} my **nausea^{negated}_{disorder}** is gone again, but since (...)]
- (ja) ** この 薬_{drug} は 効かなかった^{negative}_{opinion}
 [This **drug_{drug}** **didn't work^{negative}_{opinion}** .]

Enumerations

Consecutive entities or enumerations that are, for example, separated by commas or conjunctions are to be split apart and annotated separately.

- (5) (en) I felt **pain_{disorder}** in my **arms_{anatomy}** , **legs_{anatomy}** , **ears_{anatomy}** and **toes_{anatomy}** .
- (fr) **4_{measure}** opérations_{other} , **himurel_{drug}** , **pentasa_{drug}** , **methotrexate_{drug}** , (...)
 [**4_{measure}** surgeries_{other} , **himurel_{drug}** , **pentasa_{drug}** , **methotrexate_{drug}** , (...)]
- (de) Vor einem Jahr^{rel. point in time}_{time} bekam ich kurz hintereinander^{frequency}_{time} wei **Herzinfarkte_{disorder}** , einen **Schlaganfall_{disorder}** und eine **Embolie_{disorder}** .
 [A year ago^{rel. point in time}_{time} had two **heart attacks_{disorder}** , a **stroke_{disorder}** and an **embolism_{disorder}** in quick succession^{frequency}_{time}]
- (ja) まつ毛_{anatomy} や 眉毛_{anatomy} も 抜けてしまった_{disorder} ので
 [Because my **eyelashes_{anatomy}** and **eyebrows_{anatomy}** have **fallen out_{disorder}** too.]

Nested and Overlapping Entities

Nested entities are not permitted; we always annotate the largest phrase. For example, in the case of “headache,” we do not split the word into “head” and “ache”. The general rule is to apply a syntax-based decision, i.e., basic noun phrases (without prepositions, particles, etc.) are annotated as one entity. This is also language-dependant:

German: Do not split compound words, e.g., “Kopfschmerzen_{disorder}” [*headache*] is annotated as one entity.

Japanese: Do not split compound words based on standard tokenization, but split noun phrases (including compound nouns) at “prepositions” (particles て, に, を, は, の).

French: Basic noun phrases (“douleur thoracique_{disorder}”) [*chest pain*] are annotated as one **disorder**, otherwise we separate the annotations, e.g., “douleur_{disorder} au thorax_{anatomy}” [*chest pain*].

For specific entity types, we provide prioritization rules to help in the decision of which mention to annotate. The prioritization is mentioned for the entity types in their respective sections in these guidelines.

Discontinuous Entities

Discontinuous entities are allowed if there is no easier solution, i.e., when annotation as one entity is not possible or does not make sense. Please see the following examples for a comparison of different cases:

1. “it hurts in the left lung and right lung”

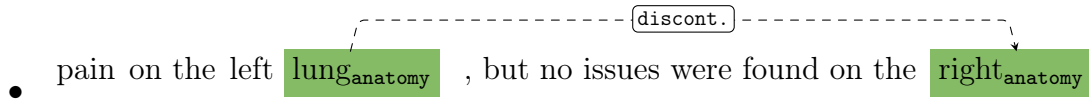
→ Here, two entities are annotated separately:

- “it hurts in the left lung_{anatomy} and right lung_{anatomy}”

2. “pain on the left lung, but no issues were found on the right”

→ The annotation uses two separate, partially overlapping entities, one continuous and one discontinuous, as shown below:

- “pain on the left lung_{anatomy}, but no issues were found on the right” (continuous entity)

- 
pain on the left lung_{anatomy}, but no issues were found on the right_{anatomy}
(discontinuous entity)

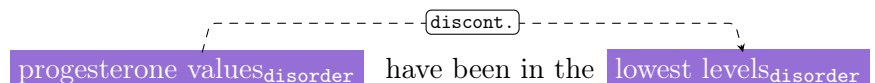
3. “nerves / muscle cramps”

→ Similarly as above, the phrase “nerves / muscle cramps” is annotated as two separate, *partially overlapping* entities:

- 
nerves_{disorder} / muscle cramps_{disorder} (discontinuous entity)

- “nerves / muscle cramps_{disorder}” (continuous entity)

In case a discontinuous entity is necessary, we limit the number of entity fragments to two to reduce the burden for the annotators.

- (6) (en) * (...) although my  progesterone values_{disorder} have been in the lowest levels_{disorder} for years.

- (fr) (...) et tous mes symptômes_{disorder} se sont à nouveau intensifiés_{disorder}, (...)
 [(...) and all my symptoms_{disorder} have intensified_{disorder} again, ...]
- (de) (...) obgleich meine Progesteronwerte_{disorder} seit Jahren im Keller_{disorder} waren.
 [(...) although my progesterone values_{disorder} have been in the lowest levels_{disorder} for years.]
- (ja) ...

Punctuation

Punctuation markers are not included in the annotation (except for, e.g., hyphens, abbreviations, etc.).

Spelling Mistakes / Colloquial Language

Since spelling mistakes can occur often, particularly when writing medication names, we treat them as if they were correctly written as long as we can easily understand what the user meant.

Abbreviations

Further, abbreviated expressions are to be annotated as well. Abbreviations are often used for drug names, but also for disorders or doctors' profession names.

- (7) (en) (...) I felt like I would imagine someone that had MS_{disorder} would feel like (...)
 (fr) Pour l'instant le gastro_{doctor} ne me parle pas de rémission_{other} !
 [so far, my GI_{doctor} hasn't said anything about remission_{other} !]
 (de) Ich glaube, ich bleibe auch erst einmal bei dem Gyn_{doctor} der letzten Jahre, (...).
 [I think I'll stay with my gyn_{doctor} of the last years for the time being.]
 (ja) * ドセ_{drug} の 浮腫み_{disorder} も気になる
 [I'm concerned about swelling_{disorder} by doce(taxel)_{drug}]

2.2 Drug

With drug_{drug}, we annotate any mention of a medication name, brand, or agent. We also include dietary supplements. As an exception, we include drug-based treatments, too, such as “chemotherapy” and “PUVA therapy” (PUVA = Psoralen and Ultraviolet A). Therapies that do not involve any drugs whatsoever are labeled as other.

- (8) (en) Lipidor_{drug} did the job_{positive} on my cholesterol_{function} both LDL and HDL.
 (fr) l'entyvio_{drug} c'est pas un miracle mais bon ça aide_{positive}. bon courage à tous.
 [entyvio_{drug} is no miracle, but well, it does help_{positive}. good luck to all.]
 (de) Ich nehme Betablocker_{drug}.
 [I take beta blockers_{drug}.]
 (ja) 夜_{time} アí デパス_{drug} で 眠れている_{function} なんですけど、(...)
 [I can sleep_{function} at night_{time} with Etizolam_{drug}]

Prioritization of drug In case *route* or *anatomy* and *drug* are one word (e.g., a compound) or connected by a hyphen, which might be the case for German, we prioritize *drug* over *route* and label the entire spans as *drug*.

- (9) (en) * Have the greatest respect for pills_{drug} in general, (...)
 (fr) (...) que tu prenais un peu de crème_{drug} le matin.
 (de) “Urogestgel_{drug}” or “Estradiol-Creme_{drug}”
 / Urogest_{drug} gel / Estradiol_{drug} cream/
 (ja) 吐き気を抑えるのに 飲み薬_{drug} と 胃薬_{drug} を使いました
 / I used an internal medicine_{drug} and a stomach medicine_{drug} to control the nausea./

2.2.1 Triggers for Medication Status Changes

Here, we annotate phrases describing or triggering a change in medication intake with the entity tag *change_trigger*_{change_trigger}. Typical words might be, for example, *increase*, *reduce*, *begin*, *one dose of*, etc.

To specify what kind of change occurred, we add the attributes *start* (the trigger specifies that medication was just started), *stop* (a medication was stopped), *increase* (the dosage was increased), *decrease* (the dosage was decreased), and *unique_dose* (the medication was only given/taken once) *to the drug they refer to* (see below in Section 2.2.2).

- (10) (en) I now have increased_{change_trigger} my intake of Vitamin C_{increased drug} to 16,000 mg / day_{measure}, and we’ll see how this works out.
 (fr) J’ai oublié_{change_trigger} de prendre mon citalopram_{decreased drug} pendant 3 jours_{duration time}
 / I forgot_{change_trigger} to take my citalopram_{decreased drug} for 3 days_{duration time}
 (de) Ich hatte vor einer Woche_{rel. point in time time} die Betablocker_{decreased drug} reduziert_{change_trigger} aber nur von einer 23mg_{measure} auf die Hälfte_{measure}.
 / I had reduced_{change_trigger} beta blockers_{decreased drug} a week ago_{rel. point in time time} but only from 23mg_{measure} to half of it_{measure} /
 (ja) 10日後くらいに_{rel. point in time time} ドセタキセルが 始まります_{change_trigger}
 / Docetaxel_{started drug} starts_{change_trigger} in about 10 days_{rel. point in time time} /

2.2.2 Attribute: Drug Changes

The attributes referring to the entity *drug* provide additional details about the context in which the medication is taken. We consider the attributes *started*, *stopped*, *increased*, *decreased* and *unique_dose*. Some examples are provided below:

- (11) (en) I now have increased_{change_trigger} my intake of Vitamin C_{increased drug} to 16,000 mg / day_{measure}, and we’ll see how this works out.
 (fr) Je suis sous sertraline_{drug} (25_{measure} puis_{change_trigger} 50_{measure} en dosage)
 / I’ve been taking sertraline_{drug} (dosage 25_{measure} then_{change_trigger} 50_{measure})/
 (de) Ich hatte vor einer Woche_{rel. point in time time} die Betablocker_{decreased drug} reduziert_{change_trigger} aber nur von einer 23mg_{measure} auf die Hälfte_{change_trigger}.

- (ja) * 副作用_{disorder} が強くて、 レキサプロ^{stopped}_{drug} 辞めました_{change_trigger} !
 [/I ^{stopped}_{change_trigger} Lexapro^{stopped}_{drug} because a _{side effect}_{disorder} was strong/]

If the patient describes having forgotten the medicine, we add the attribute **decrease** except the context suggests a complete stop (then we add **stop**). Conversely, if she describes taking, for example, a double dose, this will be marked as an **increase**.

Abbreviations of drug names

We also annotate abbreviations or colloquial names of drugs.

- (12) (en) *In my desperation, I also tried an AD_{drug} .
 (fr) (...) je passe une radio_{test} des 2 mains_{anatomy} mercredi matin^{rel. point in time}_{time} avant ma perf_{drug} , (...)
 [/I'm having a _{CT}_{test} for both hands_{anatomy} on Wednesday morning^{rel. point in time}_{time} before my IV_{drug} /]
 (de) Nun habe ich also mit Candesartan^{stopped}_{drug} und Op^{stopped}_{drug} (...) aufgehört_{change_trigger} .
 [/So now I have ^{stopped}_{change_trigger} using Candesartan^{stopped}_{drug} and Op^{stopped}_{drug} ./]
 (ja) なんか SSRI_{drug} 増やしてほしい。
 [/I need more _{SSRI}_{drug} /]

Mentions referring to a drug

Further, mentions referring to a drug but not clearly stating the drug's name are also annotated (see also Example 14 (en)).

- (13) (en) The pill_{drug} does its job^{positive}_{opinion} , (...)
 (fr) (...), j'ai pris deux_{measure} pilules_{drug} par jour_{measure} , (...)
 [I took two_{measure} pills_{drug} a day_{measure}]
 (de) Ich nehme die Tabletten_{drug} seit 2 Tagen^{duration}_{time}
 [/I have been taking the pills_{drug} for two days^{duration}_{time} /]
 (ja) 漢方薬_{drug} を購入して 飲み始めた_{change_trigger} 所、 (...)
 [/Once I _{started}_{change_trigger} taking purchased Chinese medicine_{drug} , (...)]

2.3 Disorder

A **disorder**_{disorder} annotation denotes any disease, sign, or symptom related to the patient's health, including mental issues. Sometimes a **disorder** may be expressed as a parameter in combination with a value: e.g., **high** **LDL**_{disorder} (parameter=LDL, value=high)⁵. When the value is outside the normal range, this describes a **disorder**. Sometimes, disorders are only referred to very

⁵If the value does not directly modify the target noun, the corresponding appropriate entities should be labeled separately. Example: "LDL_{function} was high_{measure}"

broadly, e.g., it might happen that the patient simply says “I do not feel well”; These expressions are also treated as disorders.

- (14) (en) I tried_{change_trigger} the advertised Arthritis medicines_{stopped drug} with severe side-effects_{disorder} and only tried this one because the doctor_{doctor} had samples.
- (fr) J’ai une maladie de crohn_{disorder} depuis 36 ans_{duration time} (...) *[I’ve had crohn’s disease_{disorder} for 36 years_{duration time} (...)]*
- (de) Nach langen, qualvollen 11 Monaten_{duration time} wurde eine Gürtelrose ohne Ausschlag_{disorder} diagnostiziert. *[After a long, agonizing 11 months_{duration time} I was diagnosed with shingles without a rash_{disorder} .]*
- (ja) 昼間_{rel. point in time time} アロレクサプロ_{drug} が 副作用_{disorder} ひどくて_{negative opinion} 未だに 気持ち悪い_{disorder} *[Daytime_{rel. point in time time} Lexapro still makes me sick_{disorder} because the side effects_{disorder} are so bad_{negative opinion}]*

Disorder vs. Function

We annotate adverse biological processes as **disorder** and neutral/positive processes as **function**, i.e., in general, a negated function is a **disorder** (but not vice versa). To distinguish **disorder** versus **function**, we apply the following rules:

1. If the entity in question is a noun phrase and describes a malfunction, we label it as a **disorder** rather than a negated function: “I have no appetite_{disorder} .”
2. If, however, a malfunction is expressed beyond a continuous noun phrase, we label it negated function: “My appetite_{negated function} has disappeared.”

2.3.1 Negation of Disorders

Currently, we only apply the negation attribute to the entity annotation **disorder** and **function**. For **disorder**, note that we do not assign the “negated” attribute to the disorders that do not completely disappear (e.g., “a long-running headache_{disorder} was almost eased_{positive opinion}”) because they still exist at least a little. Most such cases would accompany **opinion** entities, which are regarded as richer descriptions of the disorders’ status.

- (15) (en) There is no abnormality_{negated disorder} .
- (fr) Par contre j’étais pas plus fatigué_{negated disorder} que la normal. *[however, I wasn’t more tired_{negated disorder} than usual.]*
- (de) Folgende Beschwerden sind schon weg: Nackenschmerzen_{negated disorder} , (...) *[The following ailments are already gone: neck pain_{negated disorder} ,]*
- (ja) * イライラ_{negated disorder} などではなかった *[I didn’t feel irritated_{negated disorder} or something]*

2.3.2 Partial Negation (of Disorders)

We *do not* annotate partial negations since these require a more detailed annotation, which is, for now, out of scope.

- (16) (en) * My **hair_{anatomy}** has not completely fallen out.
 (fr) * Les **attaques de panique_{disorder}** ont presque disparu.
*[The **panic attacks_{disorder}** are almost gone.]*
 (de) Die **Brust_{anatomy}** tut schon viel weniger weh.
*[The **chest_{anatomy}** hurts much less already.]*
 (ja) **髪_{anatomy}** が完全に抜けたわけではありません
*[My **hair_{anatomy}** has not completely fallen out.]*

2.4 Function

With **function_{function}**, we mark all body functions and processes. Body functions are often represented in biomarkers (e.g., “HDL” and “white blood cells (WBC)”). This includes mental functions, too.

- (17) (en) **Lipitor_{drug}** did the job^{positive}_{opinion} on my **cholesterol_{function}** both LDL and HDL.
 (fr) La seule chose qu'on m'a donné pour essayer de **dormir_{function}** c'est du **Xanax_{drug}**
*[**Xanax_{drug}** is the only thing I was given to help me **sleep_{function}**]*
 (de) **Opripramol_{drug}**^{stopped} **nahm_{change_trigger}** ich nur **zwei Abende lang_{duration}** zum **Schlafen_{function}**
 (...).
*[I only **took_{change_trigger}** **Opripramol_{drug}**^{stopped} **for two evenings_{duration}** **for sleeping_{function}** (...)]*
 (ja) ** **白血球_{function}** は問題なかった。
*[no problem with **leukocyte_{function}**]*

2.4.1 Negation of Functions

For **function**, we add the negation attribute when the function goes wrong. It is not limited to absence or stopping (e.g., “**appetite_{negated function}** disappears”), but also includes abnormality, i.e., when the original body processes changes in any way (e.g., “**WBC_{negated function}** is decreasing”). Note that *continuous (noun/verbal) phrases* of malfunctions should be annotated not as **negated function**, but as **disorder**: e.g., “I have **no appetite_{disorder}**” and “**decreased WBC_{disorder}** was observed” because malfunctions can be regarded as **disorder**. See also Section 2.3.

2.5 Anatomy

With **anatomy_{anatomy}**, we annotate all organs or anatomical parts. We usually do not annotate smaller parts such as partial tissues and blood cells as **anatomy**, but as **function**. However, if a sentence describes a disorder found in a cell, the cell could be an **anatomy** entity.

- (18) (en) Had numerous odd **aches_{disorder}**, especially in the **leg area_{anatomy}**.
 (fr) (...) mais elle a fait une **réaction_{disorder}** au **pancréas_{anatomy}**, (...).
*[but her **pancreas_{anatomy}** was **affected_{disorder}**,]*

- (de) Ich besitze nur noch eine Niere_{anatomy} .
 [I only have one kidney_{anatomy} left.]
 (ja) ** お腹_{anatomy} が 刺すように痛い_{disorder}
 [I feel a stabbing pain_{disorder} in my stomach_{anatomy}]

Anatomy entities are *not annotated* when within a larger entity such as within a **disorder** or **test**. Therefore, we prioritize the annotation of **disorder** and **test**.

- (19) (en) Pain_{disorder} under ribs_{anatomy} , restless legs_{disorder} .
 (fr) J'ai beaucoup de douleurs_{disorder} au quotidien_{frequency} maux de tête_{disorder} ,
 douleurs musculaires_{disorder}
 [I am in pain_{disorder} everyday_{frequency} headaches_{disorder} , sore muscles_{disorder}]
 (de) Magen/Darmspiegelung_{test} , weil ich immer Magenschmerzen_{disorder} hatte.
 [Gastrointestinal endoscopy_{test} because I always had stomach pain_{disorder}]
 (ja) ** 肺がん_{disorder} と宣告されてしまった
 [I was diagnosed with lung cancer_{disorder} .]

2.6 Test

With **test**, we mark all medical tests, interviews, examinations or any other procedure that produces a result to be used in medical diagnoses.

- (20) (en) Blood test_{test} was normal_{measure} .
 (fr) je passe une radio_{test} des 2 mains_{anatomy} mercredi matin_{rel. point in time} avant ma perf_{drug} ,
 (...)
 [I'm having a CT_{test} for both hands_{anatomy} on Wednesday morning_{rel. point in time} before my IV_{drug}]
 (de) Ich habe ganz zu Beginn_{rel. point in time} der WJ_{function} auch mal Speicheltests_{test} machen lassen, mit zum Teil abstrusen Werten.
 [At the very beginning_{rel. point in time} of the MP_{function} , I also had saliva tests_{test} done, some of which showed abstruse values.]
 (ja) ** 血液検査_{test} の結果が怖いです
 [I'm afraid of the blood test_{test} results.]

2.7 Opinion

This denotes the *personal* evaluation or opinion of a medication (**drug**), health state (**disorder**) or biological process (**function**). This entity could be a detailed description of a mental function or disorder (e.g., “My feeling_{function} is stable_{positive opinion}”). Usually, these assessments are rather colloquial, and it is difficult to find an appropriate span. Therefore, we once again follow the principle of taking the shortest span possible.

The opinion entity always comes with a sentiment attribute: either **positive**, **negative**, or **neutral**. A **positive** assessment is often associated with an improvement of a disease or with a good experience of the patient with a certain medication. We do not assign *negated* (the *negation* attribute) to the corresponding **disorder**. We annotate *all* opinions to make it easier for the annotators and

keep the annotation consistent with the other entities. Patient-related opinions are then explicitly expressed by relations.

- (21) (en) * This **drug_{drug}** is a **nightmare^{negative}_{opinion}** and **should be discontinued^{negative}_{opinion}** .
- (fr) l' **entyvio_{drug}** c' est pas miracle mais bon **ça aide^{positive}_{opinion}** .
 [/ **entyvio_{drug}** is no miracle, but well, it does help^{positive}_{opinion} .]
- (de) Ich nehme doch jetzt **12 Wochen^{duration}_{time}** **kein_{change_trigger}** **Lyrica^{stopped}_{drug}** **mehr_{change_trigger}** und **dafür_{change_trigger}** **50mg_{measure}** **Opi^{started}_{drug}** und das **lief auch richtig gut^{positive}_{opinion}** .
 [/ I did not take_{change_trigger} **Lyrica_{drug}** for **12 weeks^{duration}_{time}** and **instead_{change_trigger}** took **50mg_{measure}** of **Opi_{drug}** and that **went really well^{positive}_{opinion}** , too.]
- (ja) * **アモキサン_{drug}** **効いてる^{positive}_{opinion}** **おかげか_{function}** **気持ち_{function}** は **少し楽になった^{positive}_{opinion}** .
 [/ My **feeling_{function}** becomes **a little better^{positive}_{opinion}** now that the **amoxan_{drug}** is working^{positive}_{opinion} .]

Emojis / Emoticons We include emojis or emoticons if they explicitly encode an opinion.

2.7.1 Attribute: Sentiment

When patients describe their current state of health or when they assess the consequences of a medication they took, they often use emotional words. Therefore, we add, if applicable, the sentiment markers **positive**, **negative**, or **neutral** to the phrases annotated with **opinion**.

- (22) (en) I really made a difference, **improvement of quality of life^{positive}_{opinion}** .
- (fr) C' est nul^{negative}_{opinion} le **Xanax_{drug}** **ça marche même pas^{negative}_{opinion}** .
 [/ **Xanax_{drug}** sucks^{negative}_{opinion} , it doesn't even work^{negative}_{opinion} //]
- (de) Dann, nachdem ich mit **Triseqenz Hormontabl.^{started}_{drug}** **angefangen_{change_trigger}** habe, nur **1 Woche später^{duration}_{time}** war ich **wie im 7. Himmel^{positive}_{opinion}** .
 [/ Then, after I **started_{change_trigger}** **Triseqent hormon pills^{started}_{drug}** , just **one week later^{rel. point in time}_{time}** I was **on cloud nine^{positive}_{opinion}** .]
- (ja) **レクサプロ_{drug}** は **強迫_{disorder}** には **効かなかったです^{negative}_{opinion}** .
 [/ **Lexapro_{drug}** didn't help^{negative}_{opinion} with **obsessive thoughts_{disorder}** .]

2.8 Measure

With **measure_{measure}** , we mark clinically relevant measurements, such as drug dosages and test results. The expression is typically a numerical value that accompanies a measurement unit.

- (23) (en) Started **2 years ago^{rel. point in time}_{time}** with **10 mg_{measure}** then **6 mos later^{duration}_{time}** **doc_{doctor}** upped to **20_{measure}** .
- (fr) **pentasa_{drug}** , **cortancyl_{drug}** (**10mg_{measure}**) , (...)
 [/ **pentasa_{drug}** , **cortancyl_{drug}** (**10mg_{measure}**) , ...]

- (de) Das **Utrogest_{drug}** sind **Weichkapseln_{route}** mit **100 mg_{measure}** **naturid. Progesteron_{drug}** und sie sind verschreibungspflichtig.
 [/ **Utrogest_{drug}** are **soft capsules_{route}** with **100 mg_{measure}** of **nature-id. progesterone_{drug}** and they are prescription-only.]
- (ja) **生検_{test}**の結果 **Ki67_{function}** が **46%_{measure}** だったため、通院にて治療中です
 [/The **biopsy_{test}** results showed a **Ki67_{function}** of **46%_{measure}** , and I am in the hospital for treatment.]

In some cases, it might happen that it is difficult to distinguish between a phrase being a **opinion** or a **measure** when they occur together with a **test** (see, for example, Section 4.5, Ambiguity). In these cases, we prioritize **measure**.

Temporal Measurements

Note that temporal measurements (e.g. “5 times per month”) should be annotated as **time** entities (see Section 2.9) *unless* they indicate the amount/dosage of a drug: If we can relate the expression with the **dosage** relation to a **drug**, then it is probably a **measure**.

1. ** I took **three days' worth_{measure}** of **medicines_{drug}** at once **today_{date}** .
2. ** My mom should take this **pill_{drug}** **three times a day_{measure}** .

2.9 Time

As temporal markers (**time_{time}**), we define all mentions of *frequencies*, *durations*, *dates*, or *relative points in time*. For a more narrow description, we use those characteristics (frequency, duration, ...) as attributes. Also, we include prepositions (e.g., “in”, “from”, “before”, and “since”) in the entity since these carry relevant information specifying the semantics of the expression. If there is no suitable attribute, we leave it blank (i.e., we do not add an attribute).

Time expressions include e.g., “night”, “afternoon”, “for one week” (mostly **duration**), but also expressions like “last Monday”, “in two weeks” (**relative points in time**), “every morning”, “after lunch” (**frequency**) or “11.07.2022” (**date**).

- (24) (en) Instead of taking the **pill_{decreased}** **2X daily_{measure}** , as prescribed I take it (...) usually **about 3-4 times per week_{measure}** , but often skipping **a week or two at a time_{frequency}**
- (fr) Bonjour, Je suis sous **entyvio_{drug}** **depuis juillet 2019_{duration}** et ce traitement **fonctionne fort bien_{positive}** **opinion** au niveau de la **colite_{disorder}** .
 [/Hi, I've been taking **entyvio_{drug}** **since july 2019_{duration}** and the drug **has been working very well_{positive}** **opinion** for my **colitis_{disorder}** .]
- (de) Hallo liebe <user>, vor 2 Monaten_{rel. point in time} habe ich meine **Frauenärztin_{doctor}** gefragt, da ich zwischendurch einen **Hormontest_{test}** wollte.
 [/Hello dear <user>, **2 months ago_{rel. point in time}** asked my **gynecologist_{doctor}** because I wanted a **hormone test_{test}** in the meantime.]
- (ja) **3年前の2月ごろに_{date}** 乳がんが発覚
 [/ **Breast cancer_{disorder}** was discovered **around February 3 years ago_{date}**]

2.9.1 Attribute: Temporality

The following attributes of time expressions are considered in our annotations: **frequency**, **duration**, **date** and **relative point in time**. The attributes should help to specify the annotated time expressions in more detail. See examples in Section 2.9. If none of the attributes fits, we do not add one.

2.10 Route

Route_{route} annotates the means of medication intake, e.g., via pills or via injection. We, thus, annotate verbs indicating medication intake, such as “drink” or “inject”, but exclude too general verbs (e.g. “have” and “take”). If a mention like “pill” refers to a drug (and not to the means of intake), it should be annotated as **drug**.

- (25) (en) It has not eliminated the need for **oral_{route}** **pain meds_{drug}** in all situations but **has helped_{positive opinion}**
- (fr) (...), depuis 1 an et demi^{duration}_{time} **perfusion_{route}** d **entyvio_{drug}** tous les mois^{frequency}_{time} (...)
 [/for the past 18 months^{duration}_{time} I get **entyvio_{drug}** **IV_{route}** every month^{frequency}_{time}]
- (de) Ich habe das leider nur mit **Tabletten_{route}** gemacht (mir wollte keiner meiner **Ärzte_{doctor}** eine **Infusion_{route}** verschreiben).
 [Unfortunately, I only did this with **pills_{route}** (none of my **doctors_{doctor}** would prescribe me an **infusion_{route}**).]
- (ja) *現在 **内服_{route}** しているのは **レトロゾール_{drug}** です
 [I am currently taking **letrozole_{drug}** **orally_{route}** .]

2.11 Doctor

We annotate medical job descriptions, such as *physician*, *dentist*, *nurse*, (*psycho*) *therapist*, etc., with **doctor_{doctor}**. This entity label is for job descriptions, *not* for the given names of doctors.

- (26) (en) I took **lipitor^{increased}_{drug}** **10mg_{measure}** for only about two months^{duration}_{time} then **cardiologist_{doctor}**
increased_{change_trigger} it to **20mg_{measure}** .
- (fr) (peut etre de **l'arthrose_{disorder}** dixit mon **docteur traitant_{doctor}**) (...)
 [(could be **arthritis_{disorder}** according to my **GP_{doctor}**) (...)]
- (de) (...), hatte ich **gestern^{rel. point in time}_{time}** einen Termin bei meinem **Internisten_{doctor}** .
 [(...), I had an appointment with my **internist_{doctor}** **yesterday^{rel. point in time}_{time}**]
- (ja) **病院の **先生_{doctor}** がこの薬を **飲み_{route}** って言うんです。でも信用できません
 [The **doctor_{doctor}** at the hospital says I should **drink_{route}** this medicine. But I don't trust him.]

2.12 Entities for De-Identification

To make sure personal information like **user names_{user}**, **URLs_{url}**, **e-mail_{perso_info}** addresses, etc., do not end up in the final corpus, even if we pre-processed the data, we also mark these information with the respective labels to later apply a final post-processing. Replacing those entities with their corresponding marker, e.g., <user>, should not change the content of the document.

User

Since social media users use creative names that are not necessarily easy to find using pre-processing tools, we mark “left-over” user names to de-identify them afterward if necessary. The examples below were randomly generated from scratch and not taken from real existing data.

- (27) (en) ** Dear **Leopard Footballer_{user}** , thank you for your message.
(fr) ** Bonjour **Mûre Vive_{user}** pour ma part ils n'ont pas été efficace^{negative}_{opinion} (...)
[Hi **Mûre Vive_{user}** , for me they were not effective^{negative}_{opinion}]
(de) ** Liebe **Broccoli Klarinette_{user}** danke für deinen Bericht.
[Dear **Broccoli Klarinette_{user}** , thanks for your report.]
(ja) ** ゆーりん_{user} さん、ありがとうございます！
[Thank you, **Yurin_{user}** !]

URL

In case there are any URLs or e-mail addresses missed when de-identifying the data, we mark them to be removed later.

- (28) (en) ** Go to **www.xxxxx.xx.xx_{url}** for more info.
(fr) ** Si vous souhaitez discuter plus avant: **www.xxx xxxx xxx xxx_{url}** .
[If you wish to discuss further: **www.xxx xxxx xxx xxx_{url}** .]
(de) ** Wenn du möchtest, geh doch mal auf die Seite **www.xxx xxxx xxx xxx_{url}** .
[If you like, take a look at this website: **www.xxx xxxx xxx xxx_{url}** .]
(ja) ** ここに詳しく載ってます **www. x x x. x x_{url}**

Personal information

With this marker, we annotate all other personal information that needs to be removed before data publication (jobs, city, doctor/hospital' names ..). This kind of annotation is only necessary if the previous (automatic) de-identification failed. The examples below were randomly generated.

- (29) (en) ** My name is **Emanuel Streich_{perso_info}** and I am **16 years old_{perso_info}** .
(fr) ** J'ai appris que j'avais **crohn_{disorder}** il y a cinq ans^{rel. point in time}_{time} J' avais **20 ans_{perso_info}** .
[I was diagnosed with **crohn_{disorder}** **five years ago^{rel. point in time}_{time}** I was **20 years old_{perso_info}** .]
(de) ** Ich wohne in **Musterhausen_{perso_info}** , das kennst du sicher.
[I live in **Musterhausen_{perso_info}** , you probably know it.]
(ja) ** 伊藤晴美_{perso_info} 、 大学2年生_{perso_info} です。
[I am **Harumi Ito_{perso_info}** , a **second-year college student_{perso_info}** .]

2.13 Other

With **other_{other}** , we annotate all remaining entities that refer to a kind of treatment or medical event (pregnancy, wound, etc.), but for which we do not have a category. For example, since we do not have an entity for treatments that are not drugs (e.g., *cognitive behavioral therapy*), we would annotate this term as **cognitive behavioral therapy_{other}** . Additional examples are clinical tools (e.g.

“syringe”) and medical devices (e.g. “wig”, “dental implant”, and “pacemaker”), as well as operations, e.g., “aesthetic surgery”.

- (30) (en) Just physical therapy_{other} and pain medication_{drug}.
 (fr) Mon médecin généraliste_{doctor} m’a recommandé la psychothérapie_{other} et le Femiloges_{drug}.

[My GP_{doctor} recommended psychotherapy_{other} and Femiloges_{drug}.]

- (de) Ohne BH fallen die Implantate_{other} runter.

[Without bra the implants_{other} fall off.]

- (ja) * 抗がん剤_{drug}の副作用_{disorder}で髪_{anatomy}が抜け_{disorder}て、ウィッグ_{other}を着用していました。

[Wore a wig_{other} because of hair loss_{disorder} due to side effects_{disorder} of anti-cancer drugs_{drug}.]

3 Relations

In the following, we will give details on how we annotate the relations between the entities. All relations were defined to be language-independent and to be used also with little linguistic knowledge. As an exception, `otherother` can take and accept all relations for practical reasons. We further take into account the following:

- The interpreted intention of the author if the annotator(s) judges it to be clearly identifiable.
- Common world knowledge (e.g., the action-reaction principle) and common sense (e.g., a broken leg can not be caused by brushing one’s teeth even if the text author insisted that it was).
- A possible contextual cue from metadata (e.g., the metadata might contain a drug name, which is not necessarily mentioned in the text itself).

We do *not* annotate relations if one of the following points apply:

- The entities described in Section 2 are not concerned, i.e., again, we follow the perspective of the patient.
- The document at hand is hypothetical or speculative (from the patient’s perspective).
- The document is formulated as a question.

For better visualization, we highlight only the relations and entities in question. Also, we cut off parts of the original documents to keep every example on one line. In Table 2, we provide an overview of all relations and their arguments.

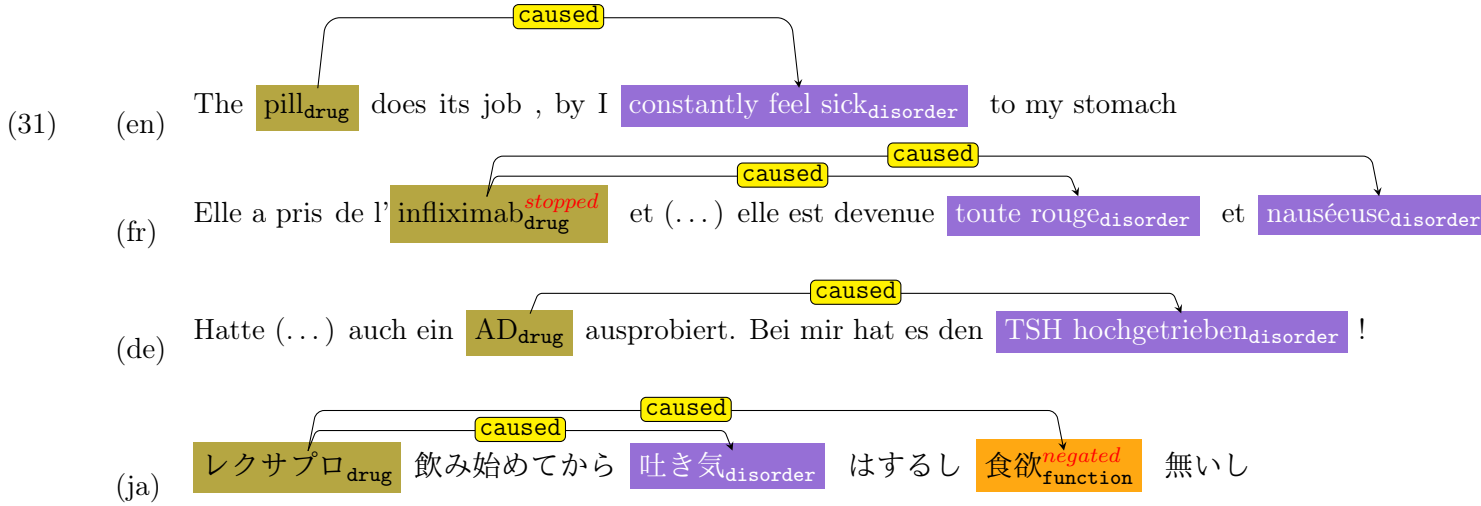
section	relation	argument 1	argument 2
3.1	caused	{drug, disorder}	{disorder, function}
3.2	treatment_for	{drug}	{disorder, function}
3.3	has_dosage	{drug}	{measure}
3.4	experienced_in	{disorder}	{anatomy}
3.5	examined_with	{disorder, anatomy, function}	{test}
3.6	has_result	{test}	{measure, disorder, function}
3.12	refers_to	{disorder}	{disorder, function ^{negated} }
3.12	refers_to	{drug}	{drug}
3.12	refers_to	{anatomy}	{anatomy}
3.12	refers_to	{function}	{function}
3.7	interacted_with	{drug}	{drug}
3.8	signals_change_of	{change_trigger}	{drug}
3.9	has_time	{drug, disorder}	{time}
3.10	has_route	{drug}	{route}
3.11	is_opinion_about	{opinion}	{drug, disorder, function}
3.13	misc	{ANY}	{ANY}
	<i>not tracked</i>	doctor, user, URL, personal info, other	

Table 2: Overview of available relations and the entities they associate.

3.1 caused

We only annotate a **caused** relation when the entities **drug**, **disorder**, or **function** are concerned. For this, we take into account the following:

- Explicit formulation of a **<cause>**–**<consequence>** relation, i.e., supported by linguistic markers like:
 - Explicit discourse inter-clause/sentence articulators: *because, so, then, ...* (including conjunction coordination hinting at causal entailment like “and”, “then”).
 - Conditional constructions: “when **<cause>**, **<consequence>**”, “**<cause>** and **<consequence>**”, ...
 - Restrictive formulations, e.g., “only when **<cause>**, **<consequence>**”.
- Lexical semantics of nouns or verbs, e.g.,:
 - **<cause>** provokes **<consequence>**
 - **<consequence>** is the consequence of **<cause>**
 - **<cause>** was followed by **<consequence>**
 - **<cause>** entailed **<consequence>**
 - **<consequence>** is correlated with **<cause>**
 - **<cause>** is correlated with **<consequence>** and **<cause>** is preceding **<consequence>**
 - **<cause>** is probably linked with **<consequence>** and **<cause>** is preceding **<consequence>**
- **<cause>** **<consequence>** relations are reported by the author of the text (first person) or attributed to another person by the author of the text (second/third person, like parents or siblings). This does not apply when one user provides advice to other users.
- Being part of a juxtaposition of linguistic element/clauses/sentences (but only if supported by the context, either co-text and/or metadata elements):
 - **<cause>** **<consequence>**
 - **<cause>**, **<consequence>**
 - **<cause>**. **<consequence>**.
- Being part of a successive temporal relation:
 - **<consequence>** after **<cause>**
 - before **<consequence>** **<cause>**
 - every time **<cause>** **<consequence>**
- Juxtaposition of **<cause>**–**<consequence>** clause and change of factuality/belief/veracity/opinion:
 - **<cause>**–**<consequence>** belief clause **<veracity expression about previous clause>**
 - “I though I would not have **<consequence>** because of **<cause>**. I was wrong!”
 - “I had never never thought that **<cause>** would yield **<consequence>**
 - “I would not have expected that **<cause>** would yield **<consequence>**



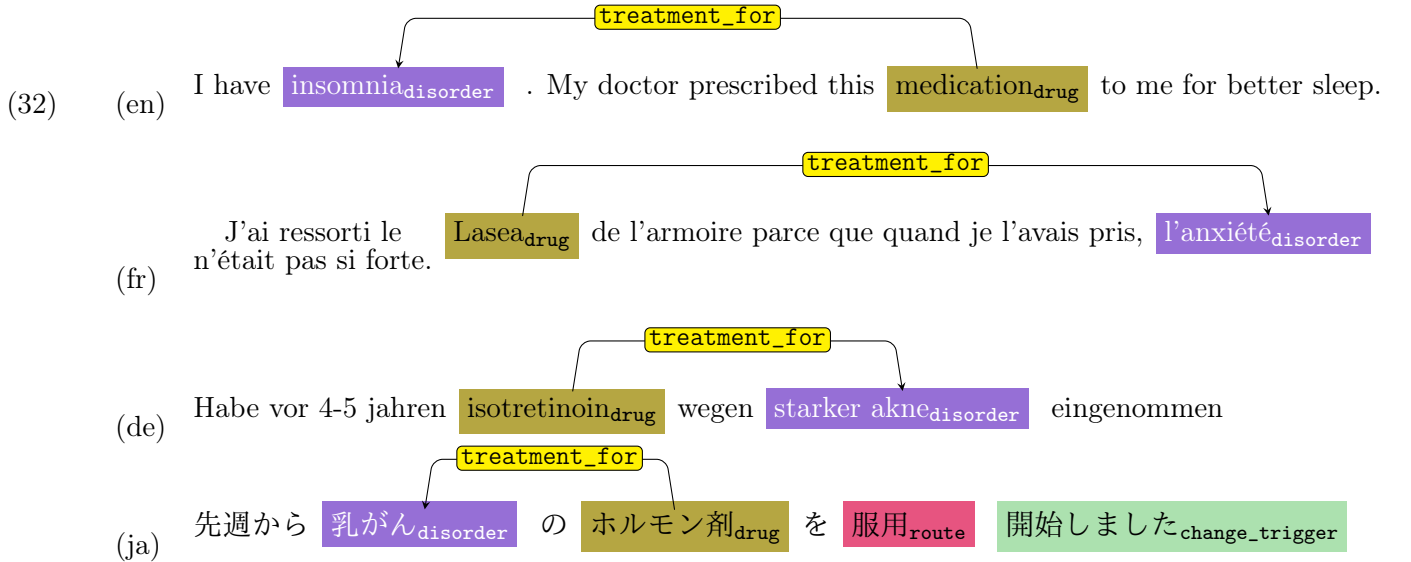
3.1.1 Special Case: Effects of Stopping Medication Intake

We include “drug withdrawal” (i.e., symptoms often caused by discontinuation of an addictive drug) and draw a **caused** relation between the **drug** and the withdrawal symptoms (**disorder**).

Stopping a medication might, however, also bring back the original symptoms or disorder. To ease the annotation burden and make the annotation consistent, we always draw a **caused** relation between a *stopped* drug and the resulting **disorder**, if there is any. Note that for the latter, there is most probably also a **treatment_for** relation.

3.2 treatment_for

This relation connects a **drug** and the targeted **disorder**, describing the medication that was used to treat the patient’s symptoms.



3.3 has_dosage

has_dosage relates a measurement, e.g. the number of pills (**measure**), with a medication name (**drug**).

- (33) (en) I took **lipitor_{drug}**^{increased} **10mg_{measure}** **for only about two months_{time}** (duration) (...)
- (fr) (...) **le premier essai_{change_trigger}** de **l'Utrogest_{drug}**^{started} **200_{measure}** a été un **échec total_{opinion}** (negative).
- (de) **Utrogest_{drug}** sind verschreibungspflichtige Kapseln mit **100 mg_{measure}** **naturid. Progesteron_{drug}**.
- (ja) **エビリファイ_{drug}** **初日_{date}** ですが **3mg_{measure}** からなので **効き目はよくわかりません_{opinion}** (neutral).

3.4 experienced_in

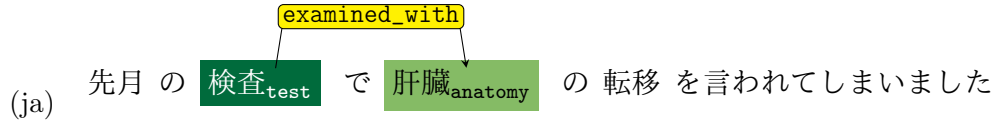
This relation associates a **disorder** with the location it is felt/experienced in, i.e., part of the body (anatomy).

- (34) (en) **Pain_{disorder}** under **ribs_{anatomy}** , **restless legs_{disorder}** .
- (fr) j'ai eu un gros souci de **désydrrose_{disorder}** aux **pieds_{anatomy}**
- (de) ** Ich hatte einen **seltsamen Schmerz_{disorder}** im **Bein_{anatomy}** .
- (ja) **顔_{anatomy}** と **足_{anatomy}** が **凄くむくむ_{disorder}**

3.5 examined_with

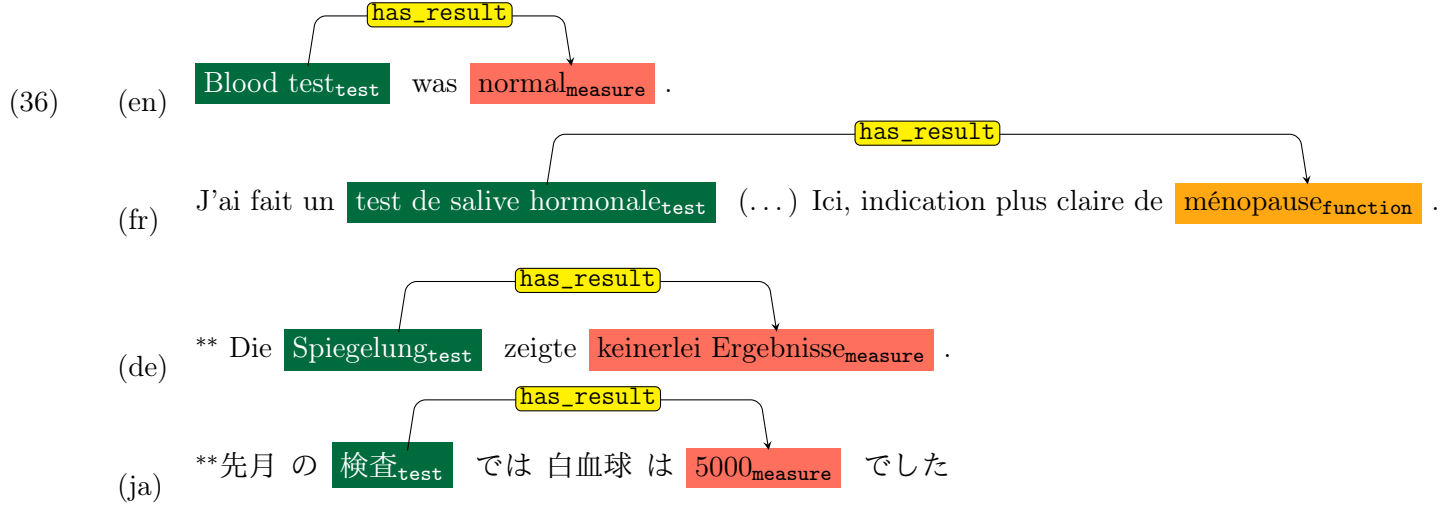
This relation associates a **disorder** or a **function** with a **test**: the **test** is used to examine the **disorder**. **anatomy** entities can be examined with a **test**, too.

- (35) (en) * I'm having **both hands_{anatomy}** **x-rayed_{test}** **on Wednesday morning_{time}** (rel. point in time) , (...)
- (fr) je passe une **radio_{test}** des 2 **mains_{anatomy}** **mercredi matin_{time}** (rel. point in time) , (...)
- (de) ** Meine **Schmerzen_{disorder}** wurden dann mit einer **Magenspiegelung_{test}** untersucht.



3.6 has_result

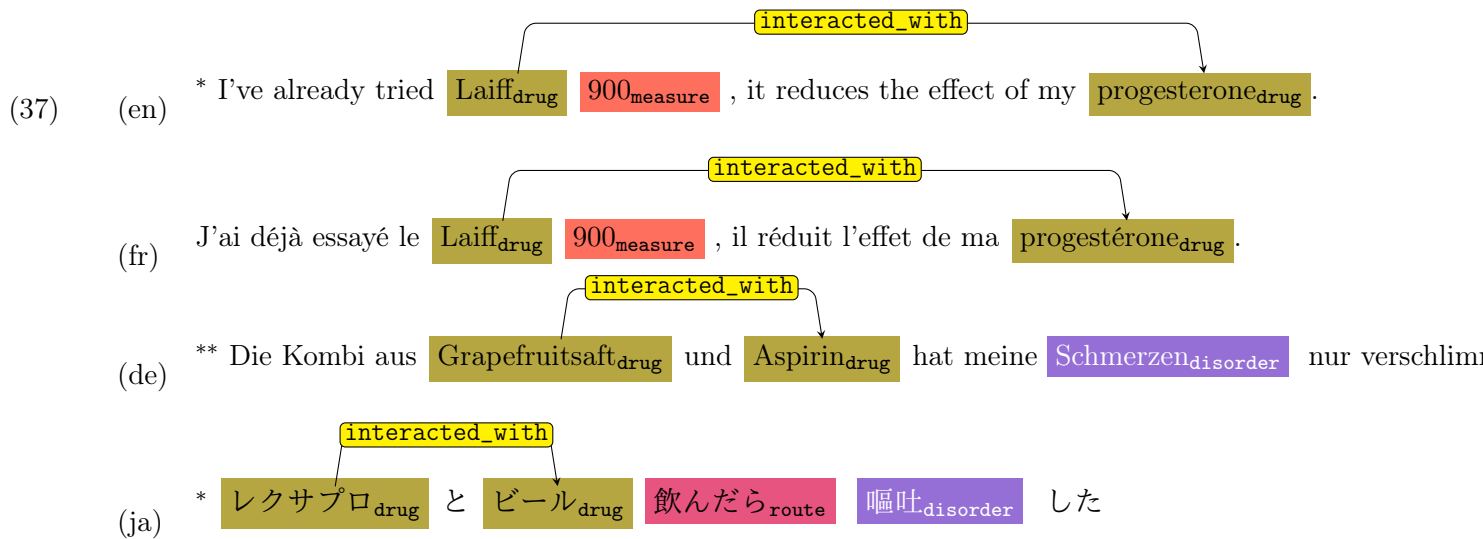
Following the line of thought from above, the `has_result` relation then relates the `test` with its result, usually a `measure`, in rare cases also a `disorder` or a `function`. This relation should not be used for `caused` scenarios.



Note that the mention of a `function` can implicitly refer to a test, resulting in a measure. Therefore, we also allow `has_result` relations between functions and measures.

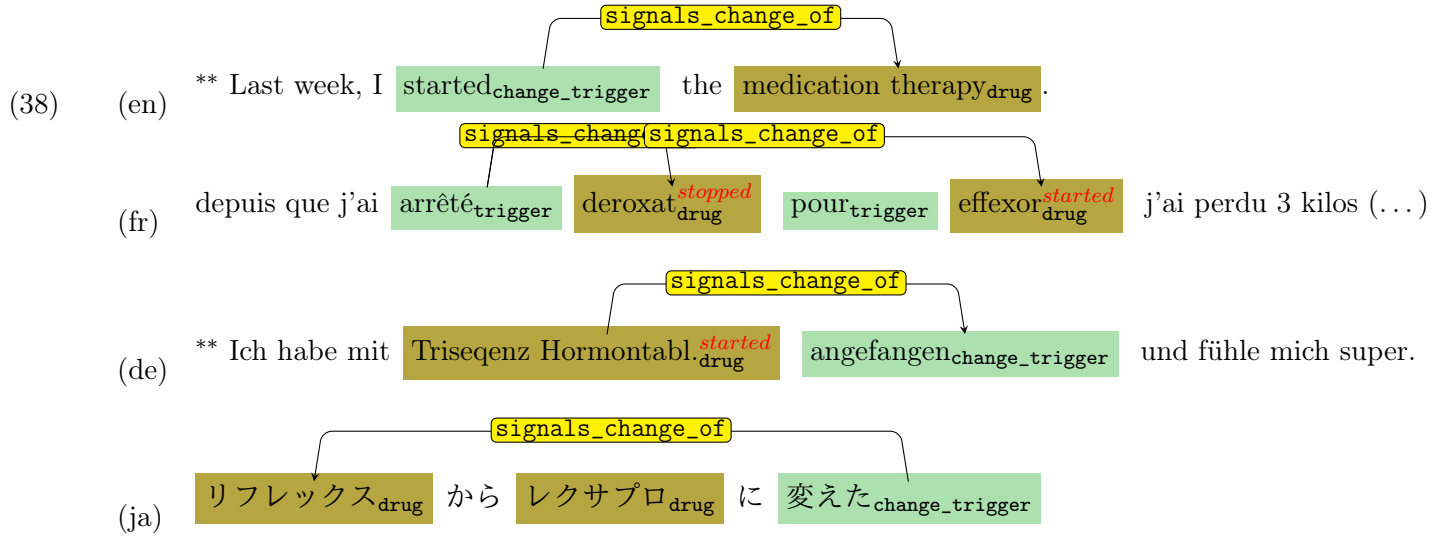
3.7 interacted_with

In some cases, a `drug-drug` interactions result in a `disorder`. The relation `interacted_with` is defined for exactly this purpose.



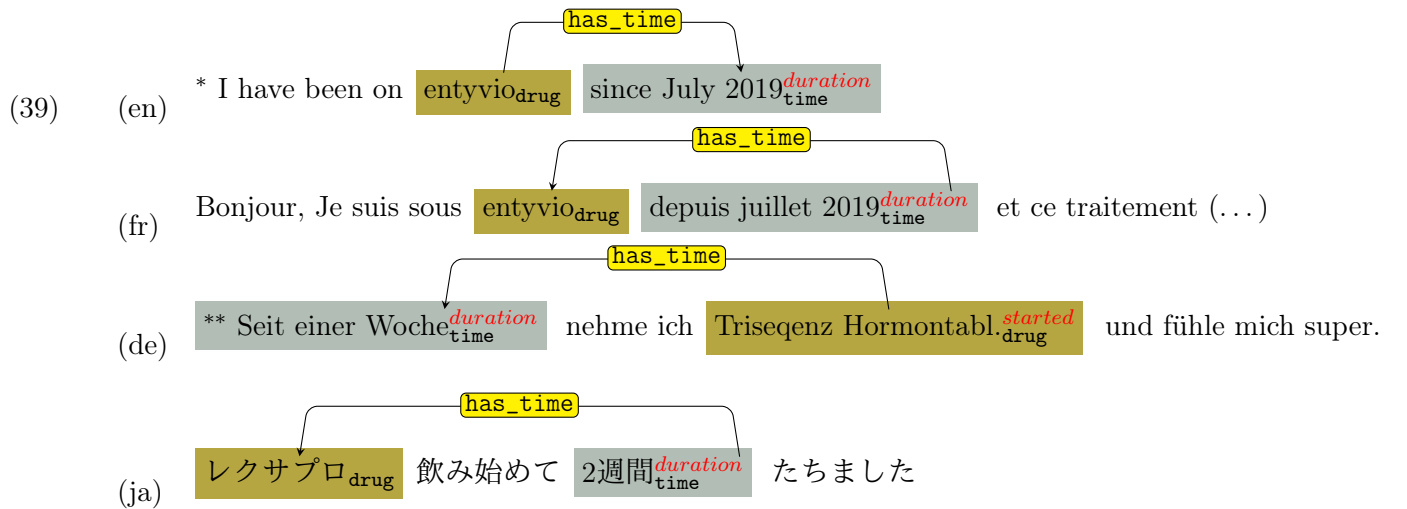
3.8 signals_change_of

For connecting the change triggers with the drug that is affected, we use the `signals_change_of` relation.



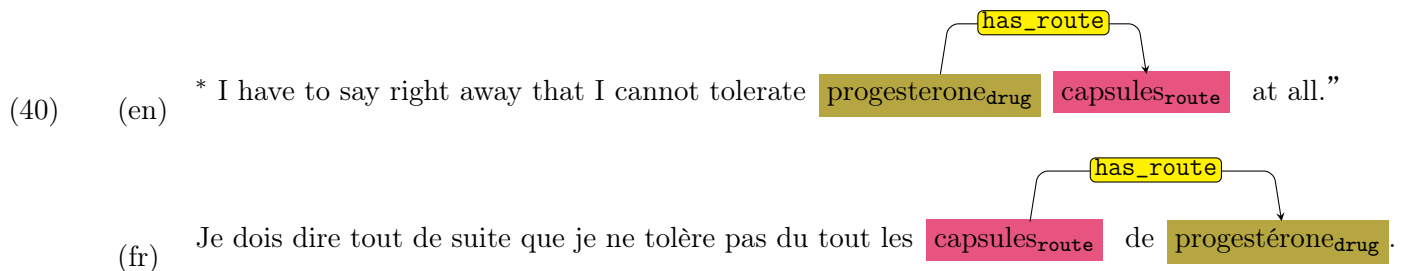
3.9 has_time

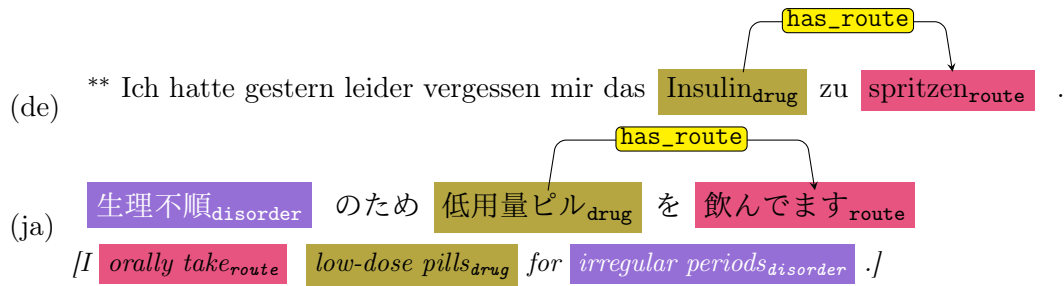
The relation time connects a time expression with all possible entity types except other time expressions.



3.10 has_route

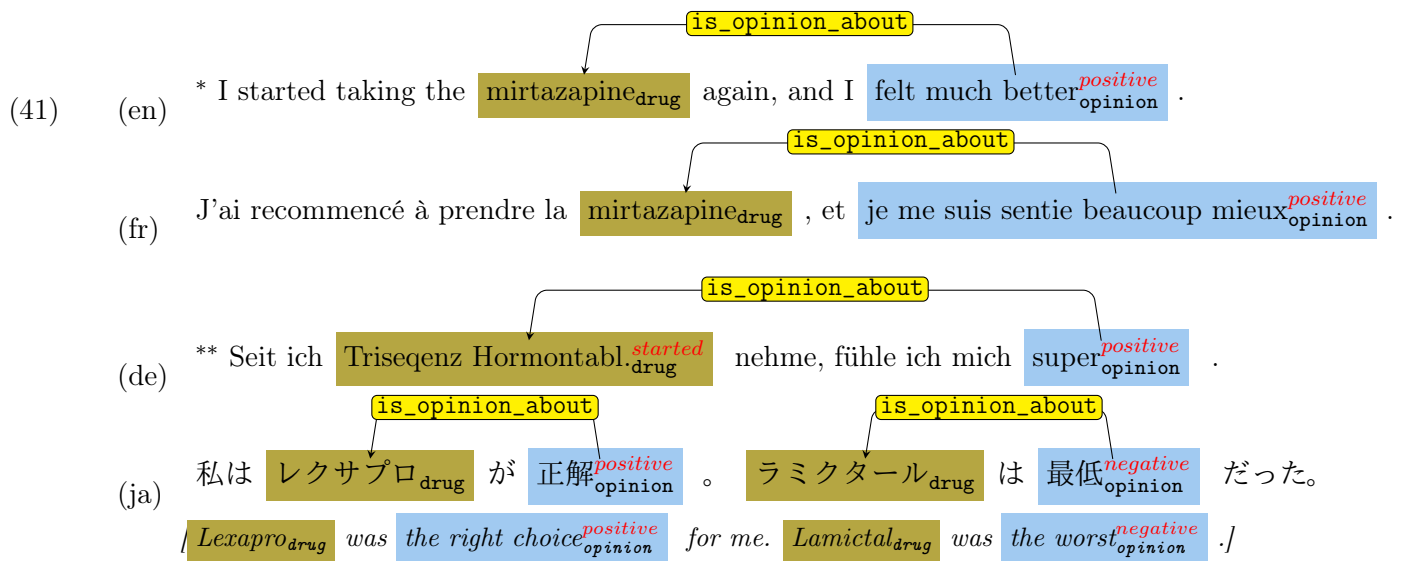
The `has_route` relation associates a drug with how it is consumed, i.e., the means of intake.





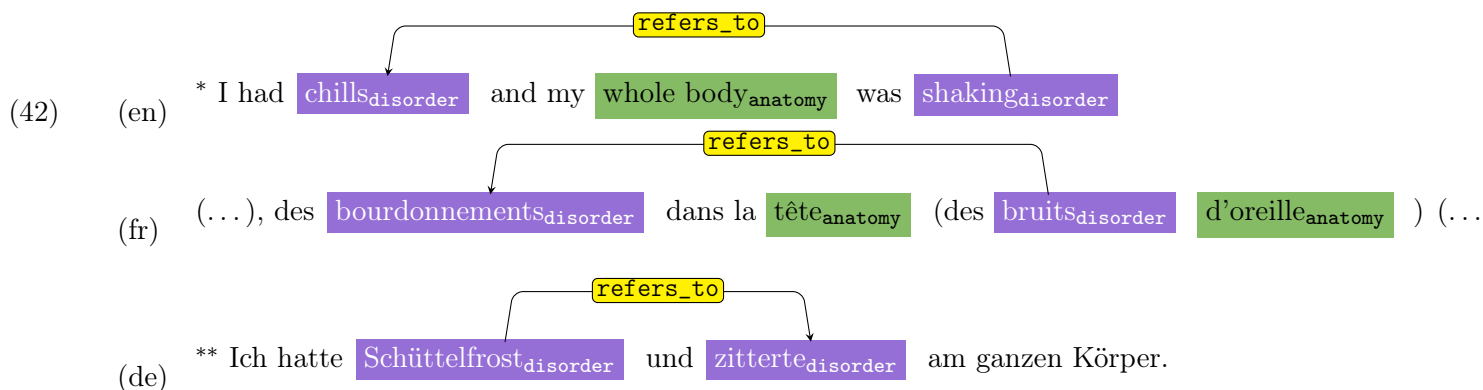
3.11 is_opinion_about

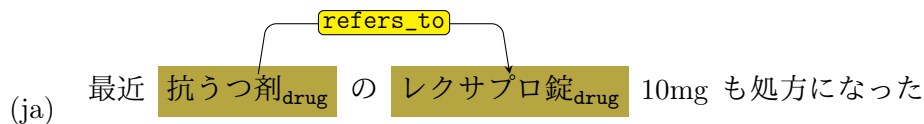
The **is_opinion_about** relation connects any entity with an **opinion** entity. Mostly, these will be **drug** or **function** entities, along with **disorder** and **other** ones. Note that sometimes, it is not possible to find two explicit entities to be connected.



3.12 refers_to

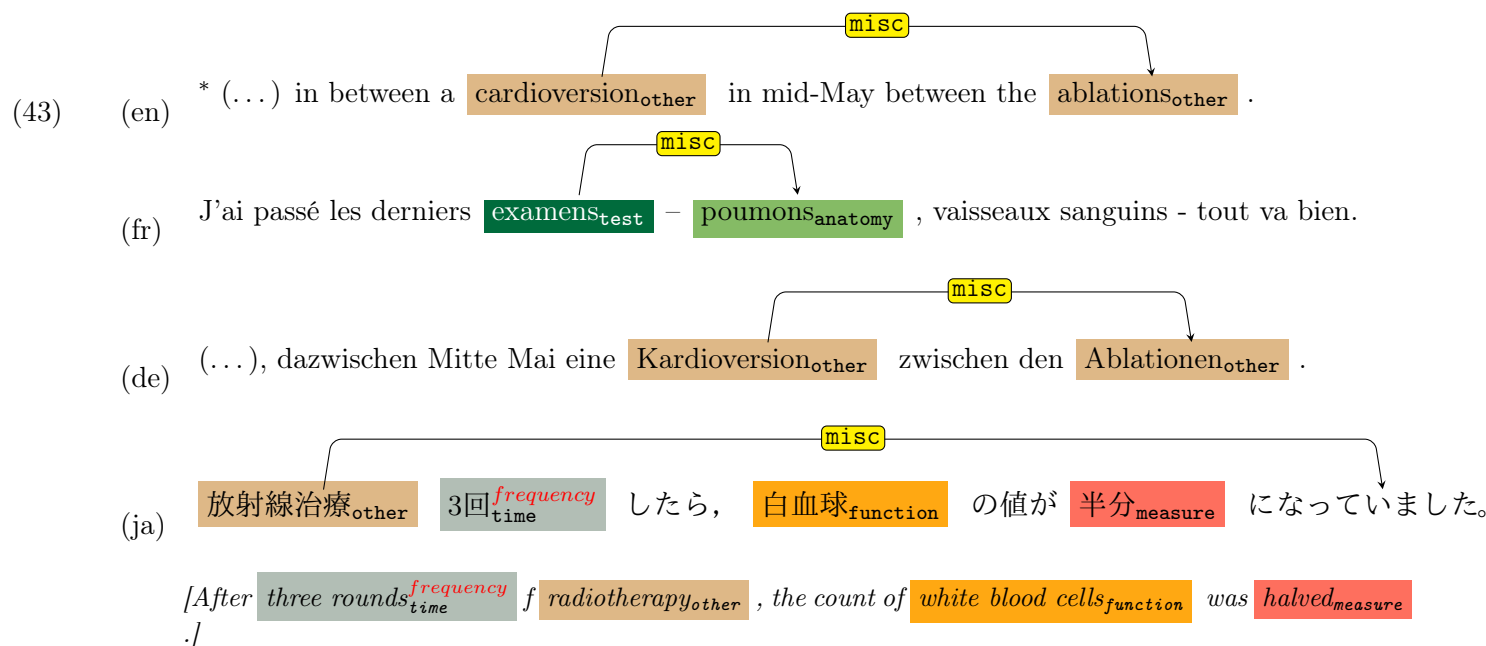
The **refers_to** relation represents a unidirectional link for the entities **drug**, **disorder**, **anatomy**, and **function**, and also for the link between **disorder** and **function^{negated}**. It is used to associate similar concepts, but *not necessarily* to connect the same surface forms (e.g., “Lipitor” and “Lipitor”). Via the **refers_to** relation, we can automatically propagate the relation to all connected targets. The relation allows for chaining corresponding entities.





3.13 misc

This relation is a joker similar to the **other** entity label. If we clearly see a worthwhile medical relation between two entities, but there is no relation defined (yet), we can use the **misc** link.



4 Difficult Cases

In the following sections, we show examples that we found difficult to annotate. They are either ambiguous or it is in general not clear what is meant by the patient.

4.1 Entailment

Implicit statements entailing or implying disorders or functions are ignored for now and are a subject of future work.

- (44) (en) ** My eyebrows remain halfway. (*meaning the other half is gone*)
 (fr)
 (de)
 (ja) まつ毛は半分残りました

4.2 Advice from other Users

In some cases, one user repeats the symptoms of another user in their own text and gives some advice. Since we are annotating the data document-wise, we do not take the original patient's view into account in this case. Therefore, we only annotate the entities but not the relations.

- (45) (en) ** Your fatigue_{disorder} is certainly due to the heavy bleeding_{disorder} .
 (fr)

(de) ** Gegen die **Hitzewellen_{disorder}** würde ich **Johanniskraut_{drug}** nehmen.
 (ja)

4.3 Distinction of Disorder, Function, and Opinion for sentiment expressions

Sentiment words such as “afraid” and “scare” may represent signs of mental disorders in some cases (disorder), whereas they may exhibit just functional psychological reactions in yet other cases (function). Also, they form opinions towards drugs (opinion). We need to distinguish them according to the local context.

- **レキサプロ_{drug}** **副作用_{disorder}** が **怖い_{function}** 。

[*Lexapro_{drug}* . *Afraid_{function}* of *side effects_{disorder}* .]
- 今日から **duration_{time}** 漢方_{time} のみにされるらしい、 **こわ_{function}**

[They are going to make me take only *Chinese medicine_{drug}* from today_{time} *scary_{function}* !]
- **不眠_{disorder}** の **症状_{disorder}** も **酷い_{negative opinion}** 。

[I also have *terrible_{negative opinion}* *symptoms_{disorder}* of *insomnia_{disorder}* . *Painful_{disorder}* .]
- **髪の毛_{anatomy}** が **なくなる_{disorder}** ということは女性にとって、 **最も辛い_{negative opinion}** ことです。

[The *loss_{disorder}* of *hair_{anatomy}* is the *most painful_{negative opinion}* thing for a woman.]

4.4 Disorder expressions using general words

- **体力_{disorder}** が **ガクンと落ちて** **しまった**

[My physical strength has gone through the roof_{disorder} .]

“体力 (my physical strength)” itself does not fall into an entity, but the whole phrase means a disorder.
- 何をするにも **体_{anatomy}** が **重い_{disorder}** 、 **だるい_{disorder}** 、 **きつい_{disorder}**

[My *body_{anatomy}* feels *heavy_{disorder}* , *sluggish_{disorder}* , and *tired_{disorder}* in any way]

“重い (heavy)” itself does not fall into an entity, but it represents a disorder in combination with “体 (body)”.
- **地獄_{negative opinion}** のように **体調が悪い_{disorder}**

[My *body condition goes bad_{disorder}* as *hell_{negative opinion}* /]

The phrase “体調が悪い (body condition goes bad)” forms a disorder only when it appears as a whole. Although it looks like a clause, we did not annotate it with Function (“**体調_{negated function}**が悪い (body condition_{negated function} goes bad)”). This is because this particular phrase “体調が悪い (body condition goes bad)” is often used like an adjective “体調悪い (bad-feeling)”, which should be labeled “**体調悪い_{disorder}** (bad-feeling)”.

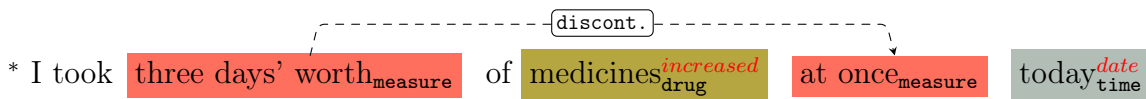
4.5 English

Semantics Some descriptions can be ambiguous.



Here, it is not entirely clear whether the right *lung* or right side of the entire body is meant.

Discontinuity As mentioned before, some entity spans need to be constructed from several sub-spans to correctly capture the meaning. This can lead to difficult constructions and makes the processing of the data more complex.



Ambiguity Natural language is often ambiguous, and more so for colloquial language taken out of a conversation online. In many cases, users are not explicitly stating, e.g., a specific test they did.

EKG_{test} perfect_{measure}, tri's_{test} were perfect_{measure}, blood sugar_{test} perfect_{measure}, etc.

In the above example, it is not entirely clear whether *blood sugar* is a **function** or a **test**. Following up on this, is *perfect* an **opinion** or a **measure**? We decide to treat implicit mentions as if they were explicit, therefore, in this example, we annotate *EKG*, *tri*, as well as *blood sugar* as **test**, and *perfect* as **measure**, as if the person would have given exact test results and reported that they did a “blood sugar test”, e.g., a A1C test.

Specific manifestations of the same function

Lipitor_{drug} did the job^{positive}_{opinion} on my cholesterol_{function} both LDL and HDL.

→ Should *LDL* and *HDL* be annotated as **function**?

4.6 French

Unclear metaphors (...), des boutons_{disorder} de la tête au pied_{anatomy}, (...)

→ What exactly does the person mean?

Function or not? Suite à ma deuxième grossesse_{other} j'ai souhaité passé sur un traitement à faire à domicile, (...)

[After my second pregnancy_{other} I wanted to go ahead with home care, (...)]

Depending on the context, *grossesse* (pregnancy) might be a function. UMLS is not clear about that.

4.7 German

Descriptive Language “Danach bin ich erstmal immer im Tal.”

4.8 Japanese

4.8.1 Body weight changes with weight values

-

Here, “2キロ (two kilograms)” is inserted between “体重 (my weight)” and “減った (dropped)” in the original Japanese sentence, which forced the annotation to be *discontinuous*.

-

“8キロ (8 kg)” also splits the “体重が増加 (my weight has increased)” in this case, resulting in a discontinuous annotation