

## Question Focus/Question Type/Frame Annotation in Consumer Health Questions:

In the second phase of consumer health question annotation, we annotate individual questions that appear in consumer requests. We consider the following semantic elements: *question focus* (topic), *question type*, and *frames*. The earlier named entity corpus has been expanded to include 1740 requests. Each request will be double-annotated.

An individual request may contain one or more questions. For example:

*A family member has recently been diagnosed with hemangiopericytoma of the skull and a lesion in the lung and possibly the kidney. We are interested in learning more about this condition, including the standard course of treatment.*

The request above can be said to consist of two questions:

- *We are interested in learning more about hemangiopericytoma.*
- *We are interested in learning the standard course of treatment for hemangiopericytoma.*

We annotate the relevant elements of all questions in a request.

### Question Focus:

Question Focus is the named entity that is the central theme of the question. Since we mainly annotated questions about diseases and drugs, we expect the question focus to be of Problem or DrugSupplement type, although other entity types are allowed as Focus as well. In the request above, the Focus element is *hemangiopericytoma*, a Problem.

Generally, a question has a single focus, but multiple focus elements are possible, especially in questions about relations between two diseases or two drugs. An example is below.

*I'd like to learn more about **megalocytic interstitial nephritis** with **malakoplakia**.*

In brat, Focus is annotated as a named entity attribute. Simply select the named entity you'd like to annotate as the Focus, and check the Focus checkbox at the bottom of the dialog. (F) appears next to the named entity.

We assume that all relevant named entities have already been annotated; therefore, you should not have to annotate a named entity. The annotator should avoid adding a new named entity to the extent possible, but it's allowed, in addition to deleting existing incorrect ones and moving those with incorrect boundaries.

Each request should have at least one Focus annotation.

All occurrences of the Focus term, including synonyms and misspellings, in the request should be annotated. (e.g., annotate both *anosmia* and *loss of smell*)

When there are nested/overlapping entities, annotate as Focus the term that question is really asking about. (e.g., in '*Ropinirole HCl 0.25 and ropinirole HCl 0.5 mg. Please tell me if these meds are gluten-free*', the annotator should make the inference that the question is really about *ropinirole*, rather than individual dosage/forms of this drug, and annotate only *ropinirole* as the Focus).

If the Focus is pre-annotated and you don't agree with it, remove the Focus attribute.

### Question Types:

Question type indicates the aspect of a named entity about which the consumer is asking a question. Two question types (INFORMATION and COMPARISON) can apply to any named entity type, while others are limited to diseases and drugs/procedures (see below). Two additional question types serve as catch-all types for drug-related questions (DRUG\_QUESTION) and other questions (OTHER\_QUESTION).

The annotator is expected to annotate the word or the phrase that indicates the question type (trigger). The annotator should strive to annotate the minimal trigger. When unsure about the minimal trigger, one effective strategy is to remove tokens from the potential trigger phrase and ask whether the remaining tokens still indicate the same question. For example, in 'I would like to ask if I can take glutathione supplement', the trigger is 'take' rather than 'can take' or 'ask if I can take'.

If the trigger involves a conjunction (e.g., *reduce or remove, center or organization*) and both conjuncts are equally valid as triggers, it is acceptable to annotate the phrase that contains both.

A trigger should normally contain at least one word.

In the examples below, the triggers are underlined and the named entity relevant to the trigger is in bold.

<b>INFORMATION</b>	<p>Concerned with basic information about a named entity. This includes specific requests for general information and sub-type/super-type information. For drugs, this may include generic/brand name, manufacturer, provider, whether it needs prescription, etc.</p> <p><i>I have <b>osteopoikilosis</b> and would like more <u>information</u> on this condition.</i></p> <p><i>What are the different <u>types</u> of <b>syringomyelia</b>?</i></p>
<b>COMPARISON</b>	<p>Concerned with comparison of <u>several entities</u>, which should be normally of the same or related types (i.e., comparison of treatments, comparison of diseases, etc.)</p> <p><i>Are there any <u>comparison</u> of the outcomes of <b>acupuncture</b> for tennis elbow compared to <b>massage</b> and <b>ultrasound</b>?</i></p>

Based on our previous work on disease questions, we will use the following question type classes for disease-related questions. Note that some of these question types are now applicable to procedures/devices, as well. These are currently COMPLICATION, EFFECT, LIFESTYLE\_DIET, LOCATION, PERSON\_ORGANIZATION, PROGNOSIS, and SUPPORT\_GROUP. PERSON\_ORGANIZATION is also applicable to DiagnosticProcedure.

<b>CAUSE</b>	<p>Identifies questions asking about the cause of a disease. This includes both direct and indirect causes, such as factors that might increase the susceptibility to a disease.</p> <p><i>Are there any exogenous (outside) factors that <u>cause</u> <b>Batten disease</b>?</i></p>
--------------	---

	<p>Concerned with longer term effects of a disease, usually in the form of risk factors or statements of a disease causing some other longer-term disease.</p>
COMPLICATION	<p>Are carriers of <b>cystic fibrosis</b> at a higher <u>risk</u> for other health conditions?</p> <p>Is it possible the patient can have <u>complications</u> such as loss of balance after she has <b>medical ear debridement</b>?</p>
DIAGNOSIS	<p>Concerned with the method of diagnosing a particular disease. These questions can simply ask about the diagnostic process for a given disease, specific types of diagnostic tests and even questions asking for a specific diagnosis.</p> <p>Could you please let me know the name of the blood test used to <u>diagnose</u> a <b>vitiligo</b> carrier?</p>
EFFECT	<p>Concerned with the effects of a disease that are not explicitly COMPLICATIONs or SYMPTOMs. This can include questions about a causal link between a disease and another condition or a general link, questions about specific effects of a disease, and unspecified effects that may alter normal life processes. These questions are likely to have multiple focus elements.</p> <p>Can <b>tuberous sclerosis</b> <u>affect</u> <b>eye closure</b>?</p>
FREQUENCY	<p>Concerned with the prevalence of a disease, perhaps in a certain population, whether it is rare, etc.</p> <p>Could you tell me which database I could look at to find the most recent data on <b>neuroblastoma</b> epidemiology; in particular the <u>prevalence</u> in Europe?</p>
INHERITANCE	<p>Concerned with whether a disease is genetic or with the inheritance patterns of a disease.</p> <p>Is there any evidence that hemorrhagic shock and <b>encephalopathy syndrome</b> is <u>passed</u> from parent to child?</p>
LIFESTYLE_DIET	<p>Concerned with the kinds of lifestyle/diet changes needed for a patient of a particular disease or whether a particular activity/food/substance is appropriate for the patient.</p> <p>Are <b>aloe</b> products <u>okay</u> for my aunt who has <b>kidney disease</b>?</p>
LOCATION	<p>Concerned with the body location of a disease, including where the symptoms manifest and if specific anatomical locations are possible sites for a disease.</p> <p>Does <b>IP</b> affect any <u>areas inside the body</u>, such as internal organs?</p>

<b>PERSON_ ORGANIZATION</b>	<p>Concerned with finding individuals or organizations that specialize in a particular disease. This typically includes individual specialists and researchers.</p> <p><i>How can I <u>find a physician</u> who is knowledgeable about <b>carbamoyl phosphate synthase 1 deficiency</b>?</i></p>
<b>PREVENTION</b>	<p>Concerned with prevention of a disease.</p> <p><i>Is there any way to <u>prevent</u> <b>LHON</b> with certain vitamins?</i></p>
<b>PROGNOSIS</b>	<p>Concerned with life expectancy and long-term quality of life. Specific types of questions include general prognosis, life expectancy, survival probability, lifestyle changes, long-term prospects of symptoms, and the long-term probability of symptoms.</p> <p><i>As <b>Burkitt</b> is so rare, there is a lack of research determining whether the treatment is curative or palliative and I want to know what the chances are that it will <u>return</u>.</i></p>
<b>SUSCEPTIBILITY</b>	<p>Concerned with how a disease spreads, its transmission. Similar to INHERITANCE but excludes genetic factors.</p> <p><i>How is the <b>herpes virus</b> <u>transmitted</u>?</i></p>
<b>SUPPORT_GROUP</b>	<p>Concerned with finding support groups for a particular disease.</p> <p><i>Are there any <u>support groups</u> for adults with <b>blepharophimosis type 1</b>?</i></p>
<b>SYMPTOM</b>	<p>Concerned with the signs and symptoms of a disease, including asking what the symptoms of a disease are, whether specific issues are symptoms of the disease, and the typical magnitude of the symptoms.</p> <p><i>Is fatigue a common <u>symptom</u> of <b>polyglucosan body disease</b>?</i></p>
<b>TREATMENT</b>	<p>Concerned with treatment, cure and management of a disease.</p> <p><i>Does this mean that if I were to get <b>cancer</b> in the future I would not be able to be <u>treated</u> with chemotherapy?</i></p> <p><i>Is there a <u>cure</u> for <b>HTLV-1</b> in general?</i></p>

The following drug/procedure question types are defined based on the preliminary annotation of 20 questions.

<b>ACTION</b>	<p>Concerned with how a drug acts in the body, its pharmacokinetics, how fast it shows its effects, etc.</p> <p><i>How quick does <b>Lidocaine</b> <u>start working</u> to relieve the pain?</i></p>
---------------	--

<b>ALTERNATIVE</b>	<p>Concerned with alternatives for a drug/device/procedure.</p> <p><i>Is there an efficacious <u>alternative</u> to <b>allopurinol</b>?</i></p>
<b>CONTRAINDICATION</b>	<p>Concerned with the conditions in which a drug or a therapeutic procedure should not be used because it may be harmful to the person.</p> <p><i>I would like to ask if I can <u>take</u> <b>glutathione supplement</b> if I have lupus.</i></p>
<b>COST</b>	<p>Concerned with the pricing of a procedure or a drug.</p> <p><i>I need to know the <u>cost</u> of <b>extracorporeal shock treatment</b>.</i></p>
<b>DOSAGE</b>	<p>Concerned with the appropriate dosage of a drug.</p> <p><i>What is the maximum daily <u>dosage</u> of <b>Metformin</b>?</i></p>
<b>INDICATION</b>	<p>Concerned with the conditions that the drug or therapeutic procedure is used for.</p> <p><i>Does <b>castor oil</b> <u>cure</u> hair loss?</i></p>
<b>INGREDIENT</b>	<p>Concerned with ingredients of a drug/supplement.</p> <p><i>Please tell me if <b>ropinirole</b> <u>is</u> gluten-free.</i></p>
<b>INTERACTION</b>	<p>Concerned with interactions between drugs.</p> <p><i>I am taking <b>Allopurinol</b> for the treatment of gout. My feet swell. How <u>dangerous</u> is it <u>to take</u> a <b>diuretic</b>?</i></p>
<b>LONG_TERM_EFFECT</b>	<p>Concerned with long term consequences of a drug/supplement.</p> <p><i>What are the <u>long term effects</u> of <b>psychoneurotropic drugs</b>?</i></p>
<b>OVERDOSE</b>	<p>Concerned with the consequences of an overdose.</p> <p><i>Will swallowing <b>clotrimazole 10mg lozenge</b> <u>hurt</u> me?</i></p>
<b>SIDE_EFFECT</b>	<p>Concerned with short term effects or adverse reactions of a drug/supplement.</p> <p><i>Are the <b>nicotine patches</b> <u>affecting</u> my asthma?</i></p>
<b>STORAGE_DISPOSAL</b>	<p>Concerned with instructions about how to store/dispose a drug/device. Includes questions about whether the drug is stable/potent after being stored in unusual conditions, whether a drug can be used beyond expiration date, etc.</p>

	<i>Can I use <b>Ringers Lactate</b> after it freezes and is <u>thawed</u>?</i>
<b>TAPERING</b>	<p>Concerned with instructions about how to stop using a drug/procedure/device and its effects.</p> <p><i>What are the <u>withdrawal symptoms</u> for <b>pregabalin</b>?</i></p>
<b>USAGE</b>	<p>Concerned with patient instructions for using a drug, a medical device, application of a procedure. Includes questions like how, how often (frequency), how long (duration), at what age.</p> <p><i>I want to know how long before I go to bed I am supposed to <u>take</u> <b>Ambien</b>.</i></p>

We have two additional types, which can be used if the question is not a disease question or if the question type does not neatly fit in one of the classes above. If either of the following types is used, the annotator should add a note indicating the proposed question type (use all uppercase and underscore for consistency). In brat, Notes text box is at the bottom of the dialog that comes up when an entity is selected.

<b>DRUG_QUESTION</b>	<p>Concerned with other drug information.</p> <p><i>Do you have to <u>be on</u> <b>suboxone</b> and <b>orbuprenorphine</b> to <u>use</u> <b>Lofexidine</b>? (Notes: COINDICATION)</i></p>
<b>OTHER_QUESTION</b>	<p>Concerned with a type of information not covered with the classes above.</p> <p><i>I was wondering if you could tell me about the <u>risks</u> that come with <b>genetically engineered food</b> (Notes: EFFECT)</i></p>

### Question Frames:

In question answering, frames are structured representations extracted from the question that serve the task of answer retrieval. For example, the question:

*They have diagnosed my niece with **Salla disease**. Can you please let me know what to expect?*

can be represented with the following frame, which indicates that the consumer is asking about the prognosis of *Salla disease*.

PROGNOSIS	
<b>Trigger</b>	<i>expect</i>
<b>Theme</b>	<i>Salla disease</i>

A question frame has a question type (e.g., PROGNOSIS) and minimally consists of two elements: a trigger that indicates the question type and a theme, which corresponds to a named entity (often, the Focus

element). In brat, a question frame is constructed by creating a link from the trigger to the named entity (or entities) that act as the Theme. The type of the question frame is inherited from its trigger.

Currently, brat is set up to accept Problem named entities as Theme for questions involving diseases (TREATMENT, PROGNOSIS, etc.). EFFECT questions are allowed to have multiple Theme elements that are Problems. Some question types involving disease now also accept ProcedureDevice entities as Theme (e.g., COMPLICATION, PROGNOSIS, PERSON\_ORGANIZATION).

DRUG\_QUESTION questions allow DrugSupplement, Substance, GeneProtein as Theme elements. INFORMATION, COMPARISON and OTHER\_QUESTION questions allow any named entity as the Theme element. COMPARISON requires multiple Theme elements.

ACTION, CONTRAINDICATION, DOSAGE, INGREDIENT, INTERACTION, LONG\_TERM\_EFFECT, OVERDOSE, and SIDE\_EFFECT questions accept DrugSupplement, Substance as Theme elements, while ALTERNATIVE, COST, INDICATION, STORAGE\_DISPOSAL, TAPERING, and USAGE allow ProcedureDevice as Theme elements, as well. PROC\_COMPLICATION is now obsolete and the annotator can instead use COMPLICATION with a ProcedureDevice entity as Theme.

A question frame can be annotated with the Research attribute, if the person asking the question is interested in information that goes beyond consumer health information. For example, the frame corresponding to the question 'What is the latest research on Parkinson disease treatment?', the annotator should annotate a TREATMENT frame with Parkinson disease as the Theme and also check the Research attribute on the trigger indicating the frame.

In this annotation study, we allow adding Keyword and ExcludeKeyword elements to the frame, as well. A Keyword element is a named entity can be helpful in finding more specific answers to a question. An ExcludeKeyword is an element that should be used to exclude some answers. A question frame can contain multiple Keyword and ExcludeKeyword elements. In the following question (Keyword elements are bold and underlined):

*Is it ok to drink quinine in seltzer water to ease leg cramps?*

The following frame can be annotated, provided that quinine and seltzer water have been annotated as named entities.

TREATMENT	
Trigger	<i>ease</i>
Theme	<i>leg cramps</i>
Keyword	<i>quinine</i>
Keyword	<i>seltzer water</i>

When selecting Keywords for a specific frame, ensure that they are crucial in answering the question represented by the frame. A possible strategy to determine whether a Keyword is crucial in answering the question is to ask whether including the term in a Google search would give us a better answer.

In the following question, amputation can be annotated as an ExcludeKeyword in a TREATMENT frame.

*My father is 70 years old, his **arteries in the both limbs have blockage**. Doctor's suggest to go for amputation but there is high risk in it. Please suggest treatment without **amputation**.*

When there are multiple mentions of the term that acts as the Theme, the following steps can be followed to select the one to annotate as the Theme:

- a) If one of the mentions is syntactically linked to the trigger, select that one. The syntactic link does not need to be a subject or direct object relation. In general, the trigger and the mention being in the same sentence should be sufficient.
- b) Otherwise, select the closest mention that precedes the trigger.

Similarly, if there are multiple possible triggers, select one that is syntactically linked to a relevant mention. If none, select the closest trigger that precedes the mention. If none, select the closest trigger that follows the mention.

In the question *'My son has 1q21.1 deletion. There is not much information on it. Can you help or if not, where can we go to understand about 1q21.21 more. Please, we'd like to know more.'*, there are multiple trigger that can indicate an INFORMATION frame ('know', 'understand', 'help', 'information'). We pick 'understand' as the trigger, since it's syntactically linked to the mention 1q21.1.

There can be multiple questions indicated by the same trigger. In such cases, the trigger needs to be annotated multiple times, one for each frame. For example, in the question *'Aspirin and clotrimazole. Are these medications gluten-free'*, the following should be annotated.

INGREDIENT	
<b>Trigger</b>	are
<b>Theme</b>	Aspirin
<b>Keyword</b>	gluten

INGREDIENT	
<b>Trigger</b>	are
<b>Theme</b>	clotrimazole
<b>Keyword</b>	gluten

Annotating the following frame only is wrong.

INGREDIENT	
<b>Trigger</b>	are
<b>Theme</b>	Aspirin
<b>Theme</b>	clotrimazole
<b>Keyword</b>	gluten

Multiple instances of the same frame can be annotated in the same request. For example, one request has *'I am anemic... Does this have anything to do with me not getting enough B12? .... So does this kind of anemic have anything to do with B12'*, and two separate frames with essentially the same content can be annotated in this request.



Some questions can be represented from multiple perspectives. For example, in the request *'I have suffered since mid-20's with hair loss... I have now become aware, through a friend, that castor oil has been promulgated as the cure. To me it sounds like old wife's tale as looking it up on the web it cures everything from hair loss to arthritis and auto-immune deficiency. My thoughts would be NHS would use this if it worked. Any comments would be appreciated?'* can be thought of a TREATMENT question about *hair loss* (with the Keyword *castor oil*) or an INDICATION question about *castor oil*. The annotator should strive to annotate only a single frame in such cases, unless there is enough context to justify the annotation of the additional frame. What justifies the second INDICATION frame in this case are the sentences *'To me it sounds like old wife's tale as looking it up on the web it cures everything from hair loss to arthritis and auto-immune deficiency. My thoughts would be NHS would use this if it worked.'*, which indicate that the person asking the question may be interested in knowing more about castor oil, independent from its effect on hair loss.