

Annotation guidelines

This document includes guidelines on the annotation of speculative and negative scopes in biomedical texts. The most typical keywords are listed in the document and their scopes are illustrated with examples.

Rules for annotation

Speculative annotation

Only sentences with some instance of speculative language are considered. If a sentence is a statement, that is, it does not include any speculative element or an element that refers to uncertainty, just leave it as it is. Questions inherently include uncertainty – that is why they are asked –, however, they are neglected and need not be annotated.

Marking

Speculative elements are marked with **angle brackets**: <or>, <suggests> etc. Their scope is denoted by parentheses. The speculative element is always included in its scope:

This result (<suggests> that the valency of Bi in the material is smaller than + 3).

Keywords

The most typical instances of speculative language, that is, keywords can be grouped as follows:

1. auxiliaries: *may, might, can, would, should, could* etc.
2. verbs of hedging or verbs with speculative content: *suggest, question, presume, suspect, indicate, suppose, seem, appear, favor* etc.
3. adjectives or adverbs: *probable, likely, possible, unsure* etc.
4. conjunctions: *or, and/or, either...or, versus, vs.* etc.

Sometimes hedge can be expressed through a phrase rather than a single word. **Complex keywords** are phrases that express uncertainty together, however, they cannot do that separately. An instance of a complex keyword can be seen in the following sentence:

Mild bladder wall thickening (<raises the question of> cystitis).

Complex keywords are not to be confused with the sequence of two or more keywords because they can express hedge on their own, that is, without the other keyword as well. An example for the latter:

Slightly increased perihilar lung markings (<may> (<indicate> early reactive airways disease)).

When **marking the keywords**, a **strategy of minimalism** is followed: the minimal unit that expresses hedging is marked as keyword. In other words, a sequence of words cannot be marked as a complex keyword if it is only one of those words that expresses the speculative

content (even without the other word). Thus, prepositions, determiners, adverbs etc. are not to be annotated as parts of the complex keyword if the keyword can have a speculative content on its own:

The picture most (<likely> reflects airways disease).

On the other hand, if the sequence of words expresses hedging only together, that is, the preposition, determiner or adverb modifies the meaning of the other word, then they form a complex keyword:

8-year-10-month - old with cough, (<evaluate for> pneumonia).

vs.

Evaluate kidneys.

In the first case, the patient probably suffers from pneumonia, that is, this is a case of hedging. On the contrary, the patient surely has kidneys in the second case, so this is not hedging at all.

However, it is important to note that speculation is a matter of the keyword and the syntactic structure of the sentence. That is, some of the above-mentioned keywords **do not denote speculation in all their occurrences**. The following sentence is a statement and it is the degree of probability that is precisely determined: it is not a case of hedging.

*The planar amide groups in which is still digging nylon splay around 30 less **probable** event.*

Another example where *either ... or* functions similarly to *both* is provided here, thus, the sentence is simply a statement without any speculative content.

*Here we show that the response of the HIV-1 LTR may be governed by two independent sequences located 5' to the site of transcription initiation sequences that bind **either** NFAT-1 **or** NF kappa B.*

Thus, some precaution is needed when marking the keywords and their scopes in the texts.

Scopes

The scope of a speculative element can be determined on the basis of syntax. Generally, the scope extends to the biggest unit possible, that is, **annotated scopes always have the maximal size**.

The scope of **verbs, auxiliaries, adjectives and adverbs** usually starts right with the keyword. In the case of verbal elements, i.e. verbs and auxiliaries, it ends at the end of the clause (if the verbal element is within a relative clause or a coordinated clause) or sentence, thus, all complements and adjuncts are included.

The presence of urothelial thickening and mild dilatation of the left ureter (<suggest> that the patient may have continued vesicoureteral reflux).

The presence of urothelial thickening and mild dilatation of the left ureter suggest that the patient (<may> have continued vesicoureteral reflux).

These findings that (<may> be from an acute pneumonia) include minimal bronchiectasis as well.

These findings (<might> be chronic) and (<may> represent reactive airways disease).

The scope of **attributive adjectives** generally extends to the following noun phrase, whereas the scope of **predicative adjectives** includes the whole sentence.

This is a 3 month old patient who had (<possible> pyelonephritis) with elevated fever.

(The demonstration of hormone receptor proteins in cells from malignant effusions is <possible>).

Sentential adverbs have scope over the entire sentence, while the scope of **other adverbs** usually ends at the end of the clause or sentence.

(The chimaeric oncoprotein <probably> affects cell survival rather than cell growth).

Right upper lobe volume loss and (<probably> pneumonia).

Conjunctions generally have scope over the syntactic unit whose members they coordinate.

A small amount of adenopathy cannot be completely excluded although there are no other findings of (adenopathy <or> pleural effusion).

Findings consistent with (viral <vs> reactive) airways disease with subsegmental atelectasis in the right lower lobe.

Determiners, adjectives or any modifiers – as they modify the following noun, that is, belong to the noun phrase – are also included in the scope of the conjunction.

Small focal opacity in right lower lobe, which represents (a small focus of pneumonia <or> atelectasis).

The main exception that changes the original scope of the keyword is the **passive voice**. The subject of the passive sentence was originally the object of the verb, that is, it would be within its scope. That is why the subject also must be marked within the scope of the verb or auxiliary:

(A small amount of adenopathy <cannot be> completely <excluded>).

Doctors (<cannot> completely <exclude> a small amount of adenopathy).

In a similar manner, **relative pronouns** such as *which* also have to be included in the scope of the auxiliary in the case of passive voice:

These findings further support our previous observations on the distinct regulation of expression of the human HLA-DQ class II subset, (which <may> be thus controlled at the posttranscriptional level by a CIITA-independent mechanism).

Another example of scope change is the case of **raising verbs** (*seem, appear, be expected, be likely* etc.). They can show two different syntactic patterns:

It seems that the treatment is successful.

The treatment seems to be successful.

In the first case, the scope of *seems* starts right after the verb. If this were the case in the second pattern, *the treatment* would not be included in the scope, however, it should be as it is indicated by the first pattern. Thus, in the second sentence, the scope must be extended to the subject as well.

It <seems> (that the treatment is successful).

(The treatment <seems> to be successful).

When the subject is not expressed in the second one of two coordinated clauses, the scope of the raising predicate starts with the main verb:

This phenomenon, which is independent of tumour necrosis factor, is associated with HIV replication, and (is thus <likely> to explain at least in part the perpetuation of HIV infection in monocytes).

The scope of **conjunctions** extends to all the members of the coordination. That is, it usually extends to both left and right. Complex keywords such as *either ... or* have got one scope. For the sake of simplicity, the following sentences are annotated only for conjunctions:

Symptoms may include (fever, cough <or> itches).

Mild perihilar bronchial wall thickening may represent (<either> viral infection <or> reactive airways disease).

Negative annotation

Sentences containing any kind of negation are considered. Negation is understood as the implication of the non-existence of something. However, the presence of a word with negative content does not imply that the sentence should be annotated as negative, since there are sentences that include grammatically negative words but they have got a speculative content. Thus, special attention must be paid to sentences including such words (see below).

Marking

Negative elements are marked with **brackets**: [no], [without] etc. Their scope is denoted by parentheses:

Stable appearance the right kidney ([without] hydronephrosis).

Keywords

The most typical instances of negative keywords are listed here:

1. auxiliary: *cannot*
2. adjectives or adverbs: *impossible, impossibly* etc.
3. conjunctions: *neither...nor* etc.
4. negation words: *no, not*
5. preposition: *without*

As it was mentioned before, **sentences including a negative keyword** are not necessarily to be annotated for negation. They **can**, however, **have speculative content** as well. The following sentence contains *cannot*, which is a negative keyword on its own, but not in this case:

(A small amount of adenopathy <cannot be> completely <excluded>).

Another example containing *no*:

(<No clear evidence> of hydronephrosis).

Some other sentences containing a negative keyword are not to be annotated either for speculation or for negation. In the following examples, the negative keyword is accompanied by an adverb and their meaning is neither speculative nor negative. The sequence of the negative keyword and the adverb can be easily substituted by another adverb or adjective having the same (or similar) meaning, which is by no means negative – this is shown within the examples. In this way, the sentences are considered to be statements and not cases of negation.

*Tumor necrosis factor induced slightly c-fos and had **almost no** (=little) effect on c-jun and API.*

*Thus, signaling in NK3.3 cells is **not always** (=sometimes) identical with that in primary NK cells.*

*However, NAC did **not significantly** (=lightly) affect the spontaneous production of IgE by atopic B cells.*

Thus, sentences containing negative keywords need thorough consideration before being annotated.

Scopes

Negative scopes can be determined on a syntactic base, similarly to speculative scopes (see above).

The scope of **negative auxiliaries, adjectives and adverbs** usually starts right with the keyword and ends at the end of the phrase, clause or sentence.

Single AP of the abdomen does ([not] reveal the etiology of the patient's pain).

When the **subject** of the – passive or active – sentence contains the negative determiner *no*, its scope extends to the entire sentence:

Surprisingly, however, ([neither] of these proteins bound in vitro to EBS1 or EBS2).

([No] instance of microsatellite instability was observed in a total of 880 comparisons of granulocyte and T-cell DNA).

Passive voice again changes the scope of the negative word:

(The previously demonstrated focus within the mid pole of the right kidney is [not] seen on the current study).

Negative conjunctions generally have scope over the syntactic unit whose members it coordinates. Complex keywords have got one scope:

In contrast, sodium salicylate (1 mM) inhibited ([neither] adhesion [nor] expression of these adhesion molecules).

However, if the complex negative keyword occurs within the subject of the sentence, its scope is extended to the whole sentence:

Unlike classical NFkappaB sites, ([neither] the SNE [nor] the SEE motif responded to phosphatase inhibition by okadaic acid).

Prepositions have scope over the following (noun) phrase:

Mildly hyperinflated lungs ([without] focal opacity).

Sometimes a **negative keyword** is present in the text apparently **without a scope**: it obviously expresses negation, however, the negated fact is not part of the sentence. In such cases, the keyword must be marked and the scope is neglected.

*[Negative] chest radiograph.
Otherwise [negative].*

In the case of **elliptic sentences**, the previous strategy is followed: the keyword is marked and its scope is neglected since the verbal phrase, that is, the scope of *not* is not repeated in the sentence.

This decrease was seen in patients who responded to the therapy as well as in those who did [not].

vs.

This decrease was seen in patients who responded to the therapy as well as in those who did ([not] respond to the therapy).

A special case

Verbs like *rule out* and *exclude* behave in an interesting way. If they occur in an imperative sentence, they are to be annotated as instances of hedging:

(<Rule out> pneumonia).

In this sentence, the possibility of suffering from pneumonia is still present (that is why the examination is requested), thus, it functions as a speculative keyword.

However, when these verbs are found in passive sentences, they express negation:

(The possibility of lung cancer is [ruled out]).

(Myelomeningocele is [excluded]).

On the other hand, if the passive form of these verbs is negated, the sentence again expresses hedging:

(It is <not ruled out> that the ureterocele opens into the vagina).

(Early infiltrate is <not excluded>).

Thus, these verbs need special attention when being annotated.