Annotation format defined by the CADERIGE consortium (V1.51)

Previous versions of this document

V 1.0:   Meeting (02/20/2001) at LRI
V 1.1:   Meeting (04/10/2001) at LRI
V 1.25 :   Meeting (10/17/2001) at LRI (+ 11/20/2001 at LIMSI)
V 1.4:   Meeting (06/04/2001) at LRI
V 1.5:   Meeting (07/01/2001) at INRA Paris

Translation V 0,2 (Thierry Poibeau, 04/01/2005)

1   Description of the set of annotation tags

This document presents the set of annotation tags proposed to annotate sentences containing (or not) interaction between biological entities for *prokaryote* and *eukaryote* organisms. In a first section we present the format defined for the annotation of text and sentences containing interaction information. The second section presents the DTD. The third section contains some examples with explanations.

1.1   Overall format to annotate documents

<!DOCTYPE Caderige_Annotation " Caderige_Annotation-v1.5.dtd">

These tags indicate the origin of the document, its modification date and the authors of the annotations.

```
<ANNOTATED-DOCUMENT
        id              = "short name to identify the document "
        reference       = "reference to track the document, typically a URI"
        description     = "origin of the document and pre-processing steps applied to the document"
        date            = "date of last modification"
        author          = "name(s) of the author(s) of the annotation" >
```

The <ABSTRACT> tag includes sentences belonging to the same abstract or, more generally, to the same piece of text (see the <SENTENCE> tag that is described below).

```
<ABSTRACT
        id              = "identifier of the abstract "
        reference       = "reference to track the document, typically a URI">
```

In the current version, the document is made of a series of sentences bound by the <SENTENCE> tag. For each sentence, interactions must be annotated using one the following tags:

1

- GENIC-INTERACTION,
- NON-GENIC-AGENT-INTERACTION,
- NON-GENIC-TARGET-INTERACTION,
- EXPERIMENT
- NO-INTERACTION.

If a sentence simultaneously describes several interaction processes, the sentence must be duplicated and a new INTERACTION tag must be used for each new interaction inside the SENTENCE block. Note however that the two last tags EXPERIMENT and NO-INTERACTION are mutually incompatible.

<SENTENCE
      id   = "sentence number in the abstract after pre-processing"
      title   = {yes, no}>   /* if sentence is a title or not

The GENIC-INTERACTION tag surrounds a sentence describing a genic interaction, that is to say an interaction for which at least one of the agents and all the targets are proteins, genes or ARNm. Each interaction must be identified with a name or a number (default value can be an empty string if only one interaction is described). The TYPE attribute indicates the level of regulation, that is to say when (in the process) does the interaction happen. These values are organised using a taxonomy.
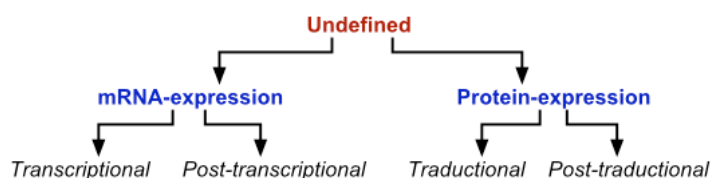
The ASSERTION and REGULATION attributes precise the nature of the interaction. REGULATION is used to say if the interaction is an activation or an inhibition (wrt. if it is an explicit information) and ASSERTION indicates possible modalities (for example if the REGULATION is negated, see below).

The UNCERTAINTY attribute indicates the certainty of the information supported by the text itself. The SELF attribute indicates whether the text in itself is enough to fully understand the interaction process or if it implies implicit information or external knowledge (note that the interaction annotation should be ground as much as possible on explicit information). Lastly, the CONFIDENCE attribute refers to the confidence of the annotator in his own annotation.

    < GENIC-INTERACTION
        id   = "interaction identifier "
        type   = {mRNA-expression, transcriptional, post-transcriptional,
               protein-expression, traductional, post-traductional, undefined}
        assertion   = {exist, non-exist}   /* sentence type
        regulation = {activate, inhibit}   /* interaction type
        uncertainty = {certain, probable, doubtful } >   /* certainty of the text
        self-contained = {yes, no}   /* implicit knowledge involved
        confidence = {good, medium, poor}

        Annotated part of the sentence (see section 2.2)

</ GENIC-INTERACTION >

The following table shows the values of the different attributes for the two interactions contained in the sentence. Note that the second attribute REGULATION may have no value (empty string). In such cases, the meaning depends on the value of the first TYPE attribute.

		REGULATION		
		« activate »	*« inhibit »*	*« »*
ASSERTION	*« exist »*	activation	inhibition	interaction mais de nature indéterminée
	« non-exist »	pas d'activation	pas d'inhibition	pas d'interaction du tout

OR

The NON-GENIC-AGENT-INTERACTION tag surrounds a sentence describing an interaction in which all the agents are non genic agents, that is to say that they are not proteins nor genes nor ARNm. The interaction must contain at least one genic target.

Attributes used for this tag are the same as the ones used for the GENIC-INTERACTION tag.

```
< NON-GENIC-AGENT-INTERACTION
        id      = "identificateur de l'interaction"
        type    = {mRNA-expression, transcriptional, post-transcriptional,
                  protein-expression, traductional, post-traductional, undefined}
        assertion    = {exist, non-exist}      /* sentence type
        regulation   = {activate, inhibit}       /* interaction type
        uncertainty = {certain, probable, doubtful } >      /* certainty
        self-contained = {yes, no}              /* implicit knowledge involved
        confidence = {good, medium, poor}

        Annotated part of the sentence  (see section 2.2)

</ NON-GENIC-AGENT-INTERACTION >
```

OR

The NON-GENIC-TARGET-INTERACTION tag surrounds a sentence describing an interaction in which all the targets are non genic agents, that is to say that they are not proteins nor genes nor ARNm. The interaction must contain at least one genic agent.

Attributes used for this tag are the same as the ones used for the GENIC-INTERACTION tag.

```
< NON-GENIC-TARGET-INTERACTION
```

```
id      = "identificateur de l'interaction"
type    = {mRNA-expression, transcriptional, post-transcriptional,
          protein-expression, traductional, post-traductional, undefined}
assertion   = {exist, non-exist}      /* sentence type
regulation  = {activate, inhibit}      /* interaction type
uncertainty = {certain, probable, doubtful } >      /* certainty
self-contained = {yes, no}            /* implicit knowledge involved
confidence = {good, medium, poor}
```

Annotated part of the sentence (see section 2.2)

</ NON-GENIC-TARGET-INTERACTION >

OR (exclusively)

The EXPERIMENT tag surrounds a non-annotated sentence describing an experimental framework that does not directly contain any explicit interaction but may have been elaborated to test an interaction (for example the action of an hormone on the expression of a gene). Contrary to previous cases, the EXPERIMENT tag should appear only once in the SENTENCE section.

<EXPERIMENT>

Non annotated part of the sentence

</ EXPERIMENT >

OR (exclusively)

The NO-INTERACTION tag contains a sentence that does not contain any interaction.

< NO-INTERACTION

Non annotated part of the sentence

</NO-INTERACTION>

For each abstract, the annotator is free to add some comments between the COMMENT tag. .

<COMMENT>

Comments (optional) on the possible annotations and the confidence in the annotation. The content of this tag will not be used for further processing stages.

</COMMENT>

</SENTENCE>

< /ABSTRACT >

</ANNOTATED-DOCUMENT>

1.2   Interaction annotation format

1.2.1   Overview

Each annotated sentence concerns only one interaction. If several interactions simultaneously appear in one sentence, the sentence must be duplicated as many times as necessary. For each interaction, the annotation process aims at highlighting the phrases of the sentence describing:

- agents (A)   : biological entities considered as agents of the interaction,
- targets (T)   : biological entities considered as targets of the interaction,
- interaction (I)   : type of control performed by the interaction,
- confidence (C)   : confidence in the described interaction.

In order to be compliant with the XML standard, tags used to annotate these different kinds of information must be strictly hierarchically nested. Moreover, from the semantic point of view, we assume that none of the four tags describing the above information (agent, target, interaction, confidence) can be nested. More drastically, no nesting is generally authorized for the same kind of tag. These are examples of such constraints:

- Forbidden annotation: <AF1> ... <IF> ... </IF> ... </AF1>
- Correct annotation: <AF1> ... </AF1> <IF> ... </IF> <AF1> ... </AF1>

- Forbidden annotation: <AF1> ... <AF2> ... </AF2> ... </AF1>
- Correct annotation: <AF1> ... </AF1> <AF2> ... </AF2> <AF1> ... </AF1>

The sentence is thus strictly partitioned: each part of the sentence receives at most one tag or at most one kind of tags. On the other hand, as we see on the last example above, a same tag may appear more than once if it is used to annotate information disseminated in different parts of the sentence. In such cases, it is the concatenation of the different groups of tagged words that ultimately constitutes the tagged information. The main advantage of this strategy is to introduce constraints in the DTD and thus being able to control and validate the annotation process more accurately. From the "edit point of view", a "visual" annotation of texts (via colours, fonts …) is simpler since the defined framework implicates no overlap between different text zones.

1.2.2   Agent description

We distinguish two different sets of tags to describe agents involved in an interaction. The first one <AFn> (Agent Fragment n) surrounds the chunks of text describing and qualifying the agent (this is called *large fragment*). The second one <An> (Agent n) is always included in the former. It surrounds the limited chunk of text in which the agent is named and clearly identified (*short fragment*). The short fragment itself can be made of several non-contiguous strings in the text. Both tags are identified with a number between 1 and 9, to distinguish the different agents that could occur in a sentence and thus avoid ambiguity. However, no specific meaning is attached to these identifiers. When an AF zone exactly corresponds to a A zone, both tags occur simultaneously for coherence <AFn> <An> ... </An> </AFn>. Logically, the identifier used for a An tag must be the same as for the Afn tag surrounding it.

We distinguish two types of agents: genic agents (GA) and non-genic agents (NGA). Genic agents GAn (together with their associated GAFn tags) only occur with the GENIC-INTERACTION and NON-GENIC-TARGET-INTERACTION tags. Non-genic agents NGAn (together with their associated NGAFn tags) only occur with the GENIC-INTERACTION and NON-GENIC-AGENT-INTERACTION tags.

1.2.2.1 Genic agents <GAn>

All the genic agents in a sentence must be distinguished using a <GAn> tag (included in a large fragment <GAFn>).

The TYPE attribute indicates the nature of the agent and the ROLE attribute indicates its implication in the interaction. The DIRECT attribute indicates whether the agent directly interacts with the target or if another genic entity is required in a way or another during the interaction. The value of this attribute is "no" if other proteins, genes or ARNm are required in the interaction process.

<GAFn>

Surrounds the chunk of text concerning the agent and possibly a GA tag

```
<GAn
        type = {gene, protein, arn, undefined}        /* nature of the agent
        role = {required, modulate, undefined }       /* role in the interaction
        direct = {yes, no, undefined }                /* is it a direct interaction?
>
```
Chunk of text precisely identifying the agent.
```
</GAn>
</GAFn>
```

Note that in this version of the DTD, we do not mention the concentration of the agent that is required for the interaction (among: basal, increase, decrease).

1.2.2.2 Non genic agents <NGAn>

Non-genic agents are described using the same rules as for the genic agents. The TYPE tag indicates if the organism may produce the product or if it is an external agent. In this version of the DTD, we do not keep track of the influence of the agent in the interaction (possible value would be: neutral, strengthen, weaken).

<NGAFn>

Surrounds the chunk of text concerning the agent and possibly a NGA tag

```
<NGAn
        type = {endogenous, exogenous}                    /* type of the agent
>
```

Chunk of text precisely identifying the agent.

         </NGAn>

</NGAFn>

1.2.3   Target description

The set of tags used to annotate sentence fragments identifying the biological targets in the interaction are very similar to the ones used for the agents (note however that attributes are not the same). As for agents, tags contain an identifier (a number between 1 and 9 with no specific meaning) to distinguish targets in the case of an interaction impacting several targets.

1.2.3.1   *Genic targets <GTn>*

<GTFn>

    Surrounds the chunk of text concerning the agent and possibly a GT tag

    <GTn type = {gene, protein, arn , undefined} >

        Chunk of text precisely identifying the target.

    </GTn>

</GTFn>

1.2.3.2   *Cibles non géniques <NGTn>*

<NGTFn>

    Surrounds the chunk of text concerning the agent and possibly a NGT tag

    <NGTn>

        Chunk of text precisely identifying the target.

    </NGTn>

</NGTFn>

1.2.4   Interaction description

The annotation of the interaction linking the agents and the targets is done through the <IF> and <I> tags respectively surrounding large and short fragments related to interaction information in the sentence. Contrary to what is done concerning agents and targets, these tags has no identification number since we suppose that each description only contains one interaction (if there are several interactions in a sentence, the sentence is duplicated).

    <IF>

    Surrounds the chunk of text concerning the agent and possibly a I tag

```
<I>
        Chunk of text precisely identifying the target.
</I>
</IF>
```

1.2.5 Confidence description (optional)

These last tags allows to optionally annotate words and phrases that ensure that the process is effective (confidence in the effectiveness of the interaction)

```
<CF>

        Surrounds the chunk of text concerning the agent and possibly a I tag

<C>
        Chunk of text precisely identifying the target.
</C>
</IF>
```

2   Description of the DTD

Below is the complete text of the DTD described in the first section of this document.

```
<!-- DTD Caderige V1.51 -->

<!-- le code pourrait etre simplifie en definissant des ENTITIES
     voir http://ctdp.tripod.com/independent/web/dtd/index.html pour un
     tutorial sur l'écriture des DTD -->

<!ELEMENT annotated-document (abstract+) >

<!ATTLIST annotated-document    id           CDATA #REQUIRED>
<!ATTLIST annotated-document    reference    CDATA #REQUIRED>
<!ATTLIST annotated-document    description  CDATA #REQUIRED>
<!ATTLIST annotated-document    date         CDATA #REQUIRED>
<!ATTLIST annotated-document    author       CDATA #REQUIRED>

<!ELEMENT abstract (sentence+) >

<!ATTLIST abstract  id        CDATA #REQUIRED>
<!ATTLIST abstract  reference CDATA #REQUIRED>

<!ELEMENT sentence ((genic-interaction+|
                     non-genic-agent-interaction+|
                     non-genic-target-interaction+|
                     experiment|no-interaction),comment?) >

<!ATTLIST sentence id     CDATA     #REQUIRED>
<!ATTLIST sentence title  (yes|no)  "yes">


<!--  il faut avoir au moins un agent genique et une cible genique apres on peut avoir
une combinaison d'agents non geniques et cibles quelconques ( ? VOIR DISCUSSION ) -->

<!ELEMENT genic-interaction
              (gaf1+|gtf1+|if+|(#PCDATA|gaf2|gaf3|gaf4|gtf2|gtf3|gtf4|cf|
              ngaf1|ngaf2|ngaf3|ngaf4|ngtf1|ngtf2|ngtf3|ngtf4)*) >

<!--  il faut avoir au moins un agent non genique et une cible genique apres
on peut avoir une combinaison d'agents non geniques et cibles quelconques -->

<!ELEMENT non-genic-agent-interaction
              (gtf1+|ngaf1+|if+|(#PCDATA|gtf2|gtf3|gtf4|cf|
              ngaf2|ngaf3|ngaf4|ngtf1|ngtf2|ngtf3|ngtf4)*) >

<!--  il faut avoir au moins un agent genique et une cible non genique apres on peut
avoir une combinaison d'agents geniques ou non et cibles non geniques -->

<!ELEMENT non-genic-target-interaction
              (gaf1+|ngtf1+|if+|(#PCDATA|gaf2|gaf3|gaf4|cf|
              ngaf1|ngaf2|ngaf3|ngaf4|ngtf2|ngtf3|ngtf4)*) >

<!ELEMENT experiment       (#CDATA) >
<!ELEMENT no-interaction   (#CDATA) >
```

```
<!-- attributs de genic-interaction -->

<!ATTLIST genic-interaction        id    CDATA #REQUIRED>
<!ATTLIST genic-interaction        type (mRNA-expression|transcriptional|
                                   post-transcriptional|protein-expression|
                                   traductional|post-traductional|undefined)
                                    "mRNA-expression"  >
<!ATTLIST genic-interaction        assertion    (exist|non-exist)   "exist" >
<!ATTLIST genic-interaction        regulation   (activate|inhibit)  "activate" >
<!ATTLIST genic-interaction        uncertainty (certain|probable|doubtful) "certain">
<!ATTLIST genic-interaction        self-contained     (yes|no)          "yes" >
<!ATTLIST genic-interaction        confidence   (good|medium|poor)  "good" >

<!-- attributs de non-genic-agent-interaction -->

<!ATTLIST non-genic-agent-interactionid         CDATA #REQUIRED>
<!ATTLIST non-genic-agent-interactiontype
                                   (mRNA-expression|transcriptional|
                                   post-transcriptional|protein-expression|
                                   traductional|post-traductional|undefined)
                                    "mRNA-expression"  >
<!ATTLIST non-genic-agent-interaction  assertion (exist|non-exist) "exist" >
<!ATTLIST non-genic-agent-interaction  regulation (activate|inhibit) "activate" >
<!ATTLIST non-genic-agent-interaction  uncertainty
                                   (certain|probable|doubtful) "certain">
<!ATTLIST non-genic-agent-interactionself-contained      (yes|no)     "yes" >
<!ATTLIST non-genic-agent-interactionconfidence (good|medium|poor)      "good" >

<!-- attributs de non-genic-target-interaction -->

<!ATTLIST non-genic-target-interactionid         CDATA #REQUIRED>
<!ATTLIST non-genic-target-interactiontype   (to-be-determined)
                                   "to-be-determined">
<!ATTLIST non-genic-target-interaction   assertion (exist|non-exist) "exist" >
<!ATTLIST non-genic-target-interaction   regulation (activate|inhibit) "activate" >
<!ATTLIST non-genic-target-interaction   uncertainty
                                   (certain|probable|doubtful) "certain">
<!ATTLIST non-genic-target-interaction       self-contained     (yes|no)      "yes" >
<!ATTLIST non-genic-target-interactionconfidence    (good|medium|poor)  "good" >

<!-- définition des zone « larges », pour le moment on ne gère que les numéro 1-4 -->

<!ELEMENT gaf1 (#PCDATA|ga1)* >
<!ELEMENT gaf2 (#PCDATA|ga2)* >
<!ELEMENT gaf3 (#PCDATA|ga3)* >
<!ELEMENT gaf4 (#PCDATA|ga4)* >

<!ELEMENT ngaf1 (#PCDATA|nga1)* >
<!ELEMENT ngaf2 (#PCDATA|nga2)* >
<!ELEMENT ngaf3 (#PCDATA|nga3)* >
<!ELEMENT ngaf4 (#PCDATA|nga4)* >

<!ELEMENT gtf1 (#PCDATA|gt1)* >
<!ELEMENT gtf2 (#PCDATA|gt2)* >
<!ELEMENT gtf3 (#PCDATA|gt3)* >
<!ELEMENT gtf4 (#PCDATA|gt4)* >

<!ELEMENT ngtf1 (#PCDATA|ngt1)* >
<!ELEMENT ngtf2 (#PCDATA|ngt2)* >
```

```
<!ELEMENT ngtf3 (#PCDATA|ngt3)* >
<!ELEMENT ngtf4 (#PCDATA|ngt4)* >

<!ELEMENT if (#PCDATA|i)* >

<!ELEMENT cf (#PCDATA|c)* >
```

<!-- agents geniques -->

```
<!ELEMENT ga1      (#CDATA) >
<!ATTLIST ga1type  (gene|protein|arn|undefined)      "protein" >
<!ATTLIST ga1role  (required|modulate|undefined)     "modulate" >
<!ATTLIST ga1direct (yes|no|undefined)                       "yes" >


<!ELEMENT ga2      (#CDATA) >
<!ATTLIST ga2type  (gene|protein|arn|undefined)      "protein" >
<!ATTLIST ga2role  (required|modulate|undefined)     "modulate" >
<!ATTLIST ga2direct (yes|no|undefined)                       "yes" >


<!ELEMENT ga3      (#CDATA) >
<!ATTLIST ga3type  (gene|protein|arn|undefined)      "protein" >
<!ATTLIST ga3role  (required|modulate|undefined)     "modulate" >
<!ATTLIST ga3direct (yes|no|undefined)                       "yes" >


<!ELEMENT ga4      (#CDATA) >
<!ATTLIST ga4type  (gene|protein|arn|undefined)      "protein" >
<!ATTLIST ga4role  (required|modulate|undefined)     "modulate" >
<!ATTLIST ga4direct (yes|no|undefined)                       "yes" >
```

<!-- agents non geniques -->

```
<!ELEMENT nga1     (#CDATA) >
<!ATTLIST nga1     type  (endogenous|exogenous)           "endogenous">


<!ELEMENT nga2     (#CDATA) >
<!ATTLIST nga2     type  (endogenous|exogenous)           "endogenous">


<!ELEMENT nga3     (#CDATA) >
<!ATTLIST nga3     type  (endogenous|exogenous)           "endogenous">


<!ELEMENT nga4     (#CDATA) >
<!ATTLIST nga4     type  (endogenous|exogenous)           "endogenous">
```

<!-- cibles geniques -->

```
<!ELEMENT gt1      (#CDATA) >
<!ATTLIST gt1type  (gene|protein|arn|undefined)      "protein" >


<!ELEMENT gt2      (#CDATA) >
<!ATTLIST gt2type  (gene|protein|arn|undefined)      "protein" >


<!ELEMENT gt3      (#CDATA) >
<!ATTLIST gt3type  (gene|protein|arn|undefined)      "protein" >


<!ELEMENT gt4      (#CDATA) >
<!ATTLIST gt4type  (gene|protein|arn|undefined)      "protein" >
```

<!-- cibles non geniques -->

```
<!ELEMENT ngt1      (#CDATA) >
<!ELEMENT ngt2      (#CDATA) >
<!ELEMENT ngt3      (#CDATA) >
<!ELEMENT ngt4      (#CDATA) >
```

<!-- autres zones étroites et divers -->

```
<!ELEMENT i (#CDATA) >
```

```
<!ELEMENT c (#CDATA) >
```

```
<!ELEMENT comment (#CDATA) >
```

3   Examples of document

Below is an example based on a subset of the sentences proposed by Claire Nedellec to develop version 1.0. To simplify the example, we assume that all the sentences in this example come from the same abstract. Colour should also facilitate the reading of these examples.

	Agents	Targets	Interaction	Confidence
Large fragment	<AFn> .. </AFn>	<TFn> .. </TFn>	<IF> .. </IF>	<CF> .. </CF>
Short fragment	<An> .. </An>	<Tn> .. </Tn>	<I> .. <I>	<C> .. </C>

<!DOCTYPE Caderige_Annotation " Caderige_Annotation-v1.5.dtd">

<ANNOTATED-DOCUMENT
    id   = "Exemple d'annotation DTD"
    reference   = "http://medline"
    description   = "Requête effectuée sur MedLine avec *bacillus subtilis transcription*"
    date   = "1/07/02"
    author   = "S. Lagarrigue, P. Bessière, A. Nazarenko, G. Bisson" >

<ABSTRACT   id = "UI 1212" reference = "" >

<SENTENCE   id = "4"   title= no >

    <EXPERIMENT>
      The existence of the feedback loop was demonstrated by using antibodies prepared against SpoIIID to measure the level of spoIIID during sporulation of wild-type cells, mutants defective in sigma K production, and a mutant engineered to produce sigma K earlier than normal.
    </EXPERIMENT>

    <COMMENT>
      The experiment is aiming at testing an interaction between SpoIIID and spoIIID in a retroaction loop implicating sigma K. Next sentences may describe the result of the experiment.

      L'expérience a pour but de tester une interaction entre SpoIIID et spoIIID dans une boucle de rétroaction impliquant sigma K. On peut s'attendre à ce que les phrases suivantes décrivent les résultats de l'expérience.
    </COMMENT>

</SENTENCE>

<SENTENCE   id   = "15" title= no >

    <NO-INTERACTION>
      The sigma H factor, on the other hand, is dispensable for the switch in the position of the ftsZ assembly site.
    </NO-INTERACTION>

</SENTENCE>

<SENTENCE id = "17" title= no >

    <GENIC-INTERACTION
id = "1"
type = transcriptional
assertion = exist
regulation = activate
uncertainty = certain
self-contained = yes
confidence = good

    <IF> <I> Induction of </I> </IF> the <TF1> Bacillus subtilis <T1 type=gene> kinA gene </T1>, which codes for a major kinase of the phosphorelay pathway </TF1>, <IF> required </IF> the <AF1> <A1 type=gene role=undefined direct=undefined> spo0H gene </A1>, coding for the sigma H protein </AF1>

    </GENIC-INTERACTION >

</SENTENCE>

<SENTENCE id = "2" title=no >

    <GENIC-INTERACTION
id = "1"
type = transcriptional
assertion = exist
regulation = activate
uncertainty = certain
self-contained = yes
text-clarity = good

    <CF> <C> Previous studies showed </C> </CF> that <AF1> <A1 type=gene role=activate direct=undefined > spoIIID </A1> </AF1> <IF> <I> is needed to produce </I> </IF> <TF1> <T1 type=protein> sigma K</T1> </TF1>, but suggested that spoIIID represses sigma K directed transcription of genes encoding spore coat proteins.

    </GENIC-INTERACTION >

    <GENIC-INTERACTION
id = "2"
type = transcriptional
assertion = exist
regulation = inhibit
uncertainty = probable
self-contained = yes

```
text-clarity          = good
```

Previous studies showed that spoIIID is needed to producesigma K, <CF> but <C> suggested </C> </CF> that <AF1> <A1 type="gene" role="modulate" direct= undefined> spoIIID </A1> </AF1> <IF> <I> represses </I> </IF> <TF1> sigma K directed transcription of <T1 type="gene"> genes encoding spore coat proteins </T1> </TF1>.

</GENIC-INTERACTION >

```
<GENIC-INTERACTION
    id                = "3"
    type              = transcriptional
    assertion         = exist
    regulation        = inhibit
    uncertainty       = probable
    self-contained    = yes
    text-clarity      = good
```

Previous studies showed that spoIIID is needed to produce sigma K, <CF> but <C> suggested </C> </CF> that spoIIID represses <AF1> <A1 type="protein" role="required" direct=undefined> sigma K</A1> </AF1> <IF> <I> directed transcription of </I> </IF> <TF1> <T1 type="gene"> genes encoding spore coat proteins </T1> </TF1>.

</GENIC-INTERACTION >

</SENTENCE>

<SENTENCE   id   = "1" >

```
<GENIC-INTERACTION
    id                = "1"
    type              = mRNA-expression
    assertion         = exist
    regulation        = inhibit
    uncertainty       = certain
    self-contained    = yes
    text-clarity      = good
```

Rather, <AF1> <A1 type=protein role=modulate direct=undefined> Sigma K </A1> </AF1> <CF> <C> appears to </C> </CF> <IF> <I> negatively regulate </I> the synthesis </IF> of <TF1> <T1 type=arn> spoIIID mRNA </T1> </TF1> by accelerating the disappearance of sigmaE RNA polymerase, which transcribes spoIIID.

</GENIC-INTERACTION >

<GENIC-INTERACTION

```
id              = "2"
type            = undefined
assertion       = exist
regulation      = inhibit
uncertainty     = certain
self-contained  = yes
confidence      = good
```

Rather, <AF1> <A1 type=protein role=modulate direct=undefined> Sigma K </A1> </AF1> appears to negatively regulate the synthesis of spoIIID mRNA <IF> by <I>accelerating the disappearance </I> </IF> <TF1> <T1 type=protein> sigmaE </T1> RNA polymerase, which transcribes spoIIID </TF1>

</GENIC-INTERACTION >

<COMMENT>
R (claire) : Si on laisse tel que le TF1, on ne voit pas que Sigma K a un effet négatif indirect sur SpoIIID *parce qu'il concourt à faire disparaître* SigmaE RNA polymerase qui transcrit SpoIIID. On a ici le problème de la gestion des relations ternaires, on le traite en divisant les phrases par couples d'interaction, la disparition de SigmaE RNA polymerase est relatée ci-dessous.

If TF1 is annotated is such a way, we do not see the fact that Sigma K has an indirect negative effect on SpoIIID because it is involved in a process that makes SigmaE RNA polymerase transcribing SpoIIID disappear. This sentence is typical of problems caused by ternary interactions. Thus the sentence must be duplicated: the disappearance of SigmaE RNA polymerase is annotated below.
</COMMENT>

<GENIC-INTERACTION
```
id              = "3"
type            = transcriptional
assertion       = exist
regulation      = activate
uncertainty     = certain
self-contained  = yes
confidence      = good
```

Rather, Sigma K appears to negatively regulate the synthesis of spoIIID appears to negatively regulate the synthesis of spoIIID mRNA > by accelerating the disappearance of <AF1> <A1 type=protein role=required direct=yes> SigmaE RNA polymerase </A1> </AF1>, which <IF> <I>transcribes</I> </IF> <TF1> <T1 type=protein> spoIIID </T1> </TF1>

</GENIC-INTERACTION >

<COMMENT>
Le découpage en 2 interactions résout le problème soulevé dans V1.0. Il y a en fait trois interactions (et non 2) qui sont décrites dans cette phrase ...

The duplication of the sentence solves the problem examined above. There are in reality three different interaction in this sentence, and not only two…
</COMMENT>

</SENTENCE>

<SENTENCE   id   = "?" >

    <GENIC-INTERACTION
      id   = "1"
      type   = transcriptional
      assertion   = exist
      regulation   = activate
      uncertainty   = certain
      self-contained   = yes
      text-clarity   = good

    <IF> A <I> low level </I> of </IF>   <AF1>   <A1 type=protein role=modulate direct=yes> GerE </A1> </AF1>, <IF> <I> activated </I> transcription of </IF>   <TF1> <T1 type=protein> CotD </T1> </TF1> by   <AF2>   <A2 type=protein role=required> GerE RNA polymerase </A2> </AF2>, but <CF> <C> in vitro </C> </CF>

    </GENIC-INTERACTION >

    <COMMENT>
    On peut mettre la valeur Required pour le ROLE de l'agent 2 car il est possible d'utiliser une règle systèmatique associant la notion de RNA polymerase avec Required.

    The "required" value may be used for the ROLE attribute concerning agent 2 because a systematic rule can be defined associating RNA polymerase with "required".
    </COMMENT>

</SENTENCE>

<SENTENCE   id   = "?" >

    <GENIC-INTERACTION
      id   = "1"
      type   = transcriptional
      assertion   = exist
      regulation   = inhibit
      uncertainty   = probable
      self-contained   = yes
      text-clarity   = good

    These results <CF> suggest that </CF> <AF1>   <A1 type=protein role=modulate direct=yes> yfhP </A1> </AF1>, <CF> <C> may act </C> </CF> as a <IF> <I> negative regulator </I> for the transcription of </IF> <TF1> <T1 type=gene> yfhQ

</T1> </TF1>, <TF2> <T2 type=gene> yfhR </T2> </TF2>, <TF3> <T3 type=gene> sspE </T3> </TF3> and <TF4> <T4 type=gene> yfhP </T4> </TF4>.

</GENIC-INTERACTION >

</SENTENCE>

</ABSTRACT>

</ANNOTATED-DOCUMENT>