# Challenge LLL
# Syntactic Analysis Guidelines

Sophie Aubin

8th March 2005

## 1 Introduction

This document aims at describing the syntactic relations provided to the Challenge participants.
The syntactic parsing of the sentence is performed by the Link Parser[1], version 4.1. The Link Parser data (grammar and dictionaries) were adapted to biological texts at MIG lab. The analysis producted by the Link Parser being very rich and hardly understandable by non-specialists, it is filtered, interpreted and written in a more readable format. We created a set of syntactic relations that try to satisfy the following requirements :

- homogeneity
- expressivity
- conciseness

Section 2 describes the format we chose to express the different information given by the relations. Section 3 is an alphabetical list of the syntactic relations along with their definition and an example. Some help about english grammar can be found on http://www.ucl.ac.uk/internet-grammar/home.htm.

## 2 Relations format

Two major kinds of information are contained in a syntactic relation :

- the **function** of the relation that links 2 words (e.g. "subject")
- the **morpho-syntactic nature** of the 2 words linked by the relation (e.g. a "noun" and a "verb")

We then separated our relations names into 2 fields distinguished by the mark ":" that correspond to two different features :

**function:nature**

[1] Davy Temperley. *An Introduction to the Link Grammar Parser* Carnegie Mellon University, March 1999.
http://www.link.cs.cmu.edu/link/

The "**function**" feature is necessary while the "**nature**" feature is not.
The "**function**" feature always starts with a generic function that can be specified with an attribute, separated from the generic name by an underscore ("_"). The specification can be either particular function (see "MOD_ATT") or position (see "MOD_POST") or the preposition that participates in a ternary relation, e.g. for a complement introduced by the preposition "in", the function is "COMP_in".
The function feature is formed as follows (the square brackets mean that the element is optional) :

**function[_specification]**

The "**nature**" feature is composed of 2 elements corresponding to an artificial morpho-syntactic category (MSC) of the 2 words linked by the relation, separated by an hyphen ("-"). The first element is always the HEAD[2] of the relation, the second being the EXPANSION[3]. The MSC is not directly issued from the analysis and does not necessarily correspond to the real MSC of the words. For instance, the second element of a SUBJ:V-N relation can be a pronoun. This part of the relation name is more or less informative depending on the kind of the relation (more in a MOD_ATT:N-N or MOD:V-ADV relation, less in a OBJ:V-N where the elements are always a verb and a noun or pronoun).

**MSChead-MSCexpansion**

There are 5 different morpho-syntactic categories :

- **V** : verb,
- **V_PASS** : passive verb,
- **N** : noun,
- **ADJ** : adjective,
- **ADV** : adverb

The table 1 in section 4 shows the relations that can occur between pairs of syntactic elements.

# 3 List of syntactic relations

Note that in the examples below, the words appear in the relations for sake of readability, as opposed to the challenge dataset format.[4].

**APPOS** :

**Apposition** generally occurs between two nouns. This relation, however, does not give any information about the morpho-syntactic categories of the two elements involved.

[2]The HEAD is the governor of the relation, e.g. the verb in a verb-subject relation or the noun in an noun-adjective relation

[3]The EXPANSION is the governed element of the relation, e.g. the noun in a verb-subject relation or the adjective in a noun-adjective relation

[4]The FORMAT is syntactic_relation(word1,word2)(id_word1,id_word2) instead of relation('syntactic_relation',id_word1,id_word2)

*A sigmaW dependent* ***promoter ( PW )*** *precedes sigW, demonstrating that this transcription factor is positively autoregulated.*
APPOS(promoter,PW)

**COMP_prep:ADJ-N** :
**Prepositional complement** between an **adjective** and a noun.
*Dephosphorylation of SpoIIAA-P by SpoIIE is strictly* ***dependent on*** *the* ***presence*** *of the bivalent metal ions Mn2+ or Mg2+.*
COMP_on:ADJ-N(dependent,presence)

**COMP_prep:ADV-N** :
**Prepositional complement** between an **adverb** and a noun
*Primer extension experiments and Northern blot analysis show that an active sigmaA dependent promoter precedes kdgR and transcription is terminated at the putative p independent terminator* ***downstream of kdgT.***
COMP_of:ADV-N(downstream,kdgT)

**COMP_prep:N-N** :
**Prepositional complement** between a **noun** and a noun
*Dephosphorylation of SpoIIAA-P by SpoIIE is strictly dependent on the* ***presence of*** *the bivalent metal ions* ***Mn2+*** *or Mg2+.*
COMP_of:N-N(presence,Mn2+)

**COMP_prep:N-V** :
**Prepositional complement** between a **noun** and a verb
*Evidence based on the use of modified and mutant forms of the phosphatase protein indicates that SpoIIE blocks the* ***capacity*** *of unphosphorylated SpoIIAA* ***to activate*** *sigmaF until formation of the polar septum is completed.*
COMP_to:N-V(capacity,activate)

**COMP_prep:N-V_PASS** :
**Prepositional complement** between a **noun** and a passive verb
*The amino domain retains* ***ability to*** *be* ***phosphorylated*** *by the phosphorelay.*
COMP_to:N-V_PASS(ability,phosphorylated)

**COMP_prep:V-N** :
**Prepositional complement** between a **verb** and a noun
*These results suggest that YfhP may* ***act as*** *a negative* ***regulator*** *for the transcription of yfhQ, yfhR, sspE and yfhP.*
COMP_as:V-N (acts,regulator)

**COMP_prep:V-V** :
**Prepositional complement** between a **verb** and another verb
*These results demonstrate that sigmaK dependent transcription of gerE initiates a negative feedback loop in which GerE* ***acts*** *as a repressor* ***to limit*** *production of sigmaK.*
COMP_to:V-V (acts,limit)

**COMP_prep:V-V_PASS** :
**Prepositional complement** between a **verb** and a passive verb. This relation does not occur in the results because the head verb (almost always

a verb of perception like "seem") is informative in the present task.
*The product of the codY gene **proved to** be **required** for this repression.*
COMP_to:V-V_PASS (proved,required)

**COMP_prep:V_PASS-N** :
**Prepositional complement** between a **passive verbe** and a noun.
*Northern blot and primer extension analyses indicated that yfhS is **transcribed by E sigma E** during sporulation.*
COMP_by:V_PASS-N(transcribed,E sigma E)

**COMP_prep:V_PASS-V** :
**Prepositional complement** between a **passive verb** and a verb
*Selective 2H-labeling, 13C-labeling and isotopic heterodimers were **used to distinguish** contacts between and within monomers of the dimeric protein.*
COMP_to:V_PASS-V (used,distinguish)

**COMP_prep:V_PASS-V_PASS** :
**Prepositional complement** between a **passive verb** and another passive verb. This relation does not occur in the results because the head verb (almost always a verb of opinion or perception like "believe" or "appear") is informative in the present task.
*The signal peptide was **considered to** be **consisted** of 38 amino acids.*
COMP_to:V_PASS-V_PASS (considered,consisted)

**MOD:ADJ-ADV** :
**Modifier** between an **adjective** and an adverb
*Dephosphorylation of SpoIIAA-P by SpoIIE is **strictly dependent** on the presence of the bivalent metal ions Mn2+ or Mg2+.*
MOD:ADJ-ADV(dependent,strictly)

**MOD:ADJ-N** :
**Modifier** between an **adjective** and a noun
*sspG transcription also requires the **DNA binding** protein GerE.*
MOD:ADJ-N(binding,DNA)

**MOD:ADV-ADV** :
**Modifier** between an **adverb** and an other adverb
*R factors indicate the structures fit the experimental NOE data **very well.***
MOD:ADV-ADV(well,very)

**MOD:V-ADV** :
**Modifier** between a **verb** and an adverb
*From these results we conclude that ComK **negatively regulates** degR expression by preventing sigmaD-driven transcription of degR, possibly through interaction with the control region.*
MOD:V-ADV(regulates,negatively)

**MOD:V_PASS-ADV** :
**Modifier** between a **passive verb** and an adverb
*In addition to the typical sigmaB dependent, stress- and starvation inducible pattern, yvyD is **also induced** in response to amino acid depletion.*
MOD:V_PASS-ADV(induced,also)

**MOD_ATT:N-ADJ** :

**Attributive modifier** between a **noun** and an adjective

*Transcription of ydhD was dependent on SigE, and the mRNA was detectable from 2 h after the cessation of* ***logarithmic growth*** *( T2 of sporulation ).*

MOD_ATT:N-ADJ(growth,logarithmic)

**MOD_ATT:N-N** :

**Modifier** between a **noun** and a noun

*Northern blot and primer* ***extension analyses*** *indicated that yfhS is transcribed by E sigma E during sporulation.*

MOD_ATT:N-N(analyses,extension)

**MOD_POST:N-ADJ** :

**Post-posed modifier** between a **noun** and an adjective

*It binds to a palindromic* ***sequence****, very* ***similar*** *to an Escherichia coli Crp binding site, located upstream from arcA.*

MOD_POST:N-ADJ(sequence,similar)

**MOD_PRED:N-ADJ** :

**Predicative modifier** between a **noun** and an adjective

*Therefore, the physiological* ***role*** *of sigmaB dependent katX expression remains* ***obscure.***

MOD_PRED:N-ADJ(role,obscure)

**MOD_PRED:N-N** :

**Predicative modifier** between a **noun** and another noun

*In that respect sigB is similar to the previously described gene gsiB which is also a member of the sigmaB regulon.*

MOD_PRED:N-N(gsiB,member)

**NEG** :

**Negation** between any morpho-syntactic category and the negation word. The morpho-syntactic categories of the two elements involved are not specified in the relation.

*DNase I footprinting showed that SpoIIID binds strongly to two sites in the cotC promoter region, binds weakly to one site in the cotX promoter, and does* ***not bind*** *specifically to cotB.*

NEG(bind,not)

**OBJ:V-N** :

**Object** between a verb and a noun

*DnaK, a general regulator of the heat shock response, which in bacteria* ***inhibits*** *the heat shock sigma* ***factor*** *sigma 32.*

OBJ:V-N(inhibits,factor)

**OBJ:V_PASS-N** :

**Object** between a passive verb and a noun

*Transcription of the cotB, cotC, and cotX genes by final sigma(K) RNA polymerase is activated by a small, DNA-binding protein* ***called GerE.***

OBJ:V_PASS-N(called,GerE)

**SUBJ:V-N :**

**Subject** between a verb and a noun
***Expression*** *of the sigma(K)-dependent cwlH gene* ***depended*** *on gerE.*
SUBJ:V-N(depended,Expression)

**SUBJ:V_PASS-N :**

**Subject** between a **passive verb** and a noun. This stands for the patient of a passive verb.
*Northern blot and primer extension analyses indicated that* ***yfhS*** *is* ***transcribed*** *by E sigma E during sporulation.*
SUBJ:V_PASS-N(transcribed,yfhS)

# 4 Relations between different kinds of Heads and Expansions

Not any relation can be found between two syntactic elements. The table 1 is a synthesis of the possible associations of different syntactic elements through relations. The vertical entries stand for the head of a relation and the horizontal entries stand for the expansion of a relation. For instance, an OBJ:V-N relation (object relation between a verb (head) and a noun (expansion)) is to be found on line 1 (V), column 3 (N).

| Exp. / **Head** | V | V_PASS | N | ADJ | ADV |
|---|---|---|---|---|---|
| **V** | COMP_prep | COMP_prep | COMP_prep<br>SUBJ<br>OBJ | | MOD |
| **V_PASS** | COMP_prep | COMP_prep | COMP_prep<br>SUBJ<br>OBJ | | MOD |
| **N** | COMP_prep | COMP_prep | COMP_prep<br>MOD_ATT<br>MOD_PRED | MOD_ATT<br>MOD_POST<br>MOD_PRED | |
| **ADJ** | | | MOD<br>COMP_prep | | MOD |
| **ADV** | | | COMP_prep | | MOD |

Table 1: Relations between different elements