ATLAS - Automatic Translation into Sign Languages

Title	Rule-based Translation
Sub-title	Semantic Interpretation
Document ID	D4.3#3
Release #	000.01

Responsibilities

Role	Name	Company	E-mail	Date
Prepared by	Leonardo Lesmo	Dip. Informatica – Università di Torino	lesmo@di.unito.it	08/04/2012
Controlled & Approved by				
Issued by				

ATLAS project related info

Involved Tasks	WP04	
Related Documents	WP03 – Languages Formal Definition	
Superseded Documents		
Document Type	Administrative	
	X Technical	
	Meeting Minute	
Document Status	Working Document	
	X Technical Report	
	Deliverable	
Document Classification	**X** Restricted to ATLAS partners	
	Public	

Revision history

Date	Release #	Modified Parts	Description of variations	Author

Table of content

REVISION HISTORY 2

1 INTRODUCTION 5

2 SKETCH OF THE MORPHOLOGICAL AND SYNTACTIC STEPS 7

3 ONTOLOGICAL FORMS 9

4 SEMANTIC INTERPRETATION 12

4.1 TREE ANNOTATION 12
4.1.1 ADDING WORD MEANINGS 12
4.1.2 ADDING PRAGMATIC INFORMATION 14
4.2 THE ONTOLOGY 14
4.2.1 THE TOP LEVEL 15
4.2.2 SITUATIONS 17
4.2.3 DESCRIPTIONS 17
4.2.4 RELATIONS 18
4.3 THE SHORTEST PATHS 19
4.4 THE PREPLATE TABLE 19
4.5 PATH COMPOSITION 20

5 CONCLUSIONS 24

List Of Figures

Figure 1 – Architecture of the Italian – AWLIS rule-based translator 5
Figure 2 – Syntactic structure of "Locali addensamenti potranno interessare il settore nord-orientale" 6
Figure 3 – Syntactic structure of "Ancora correnti settentrionali, molto agitato lo Ionio, …" 7
Figure 4 – Partial chunks built during the analysis of the sentence in fig.2 8
Figure 5 – The annotated syntactic tree of "Locali addensamenti potranno interessare …" 14
Figure 6 – The annotated syntactic tree of "Ancora correnti settentrionali, molto agitato lo Ionio, …" 15
Figure 7 – The top-level of the ontology 16
Figure 8 – The portion of the ontology describing situations 17
Figure 9 – The portion of the ontology concerning descriptions 18
Figure 10 – Use of relations for representing the features of *££sea-status-situation* 19
Figure 11 – [Copy of fig.5] The annotated syntactic tree of "Locali addensamenti potranno interessare …" .. 20

List Of Boxes

Box 1 – Ontological form of "Locali addensamenti potranno interessare il settore nord-orientale" 9
Box 2 – Grammar of the ontological forms 11
Box 3 – An excerpt of the table that maps words onto concepts 14
Box 4 – Organization of the semantic interpreter 21
Box 5 – Ontological form of "Locali addensamenti potranno interessare il settore nord-orientale" 22

1 INTRODUCTION

This document describes the semantic interpretation phase of the Atlas translator. The overall organization of the translator which is being developed at UnitoInf is sketched in fig.1. The final goal of the Translator is to extract from a sentence in written Italian the pieces of information useful for the weather forecasts. Such information must be encoded in a language whose expressions can be used as input for the phase of Signed Language generation. This interface is the AWLIS (ATLAS Written Italian Language of Sign), described in D3.2. In order to get this result, the Semantic Interpreter produces a (simplified) representation of the sentence in First Order Logics. This representation is then sent to the generator, which produces the FOL representation. This document describes the first two steps of the Semantic Interpreter, that produce the so-called ontological form. The third step (Translation into FOL) is the topic of D4.3#4, while two more documents (D4.3#5 and D4.3#6) describe the generation part of the Translator. The present deliverable also includes a short introduction to the syntactic parser, in order to present the syntactic structures that are the input of the interpreter.

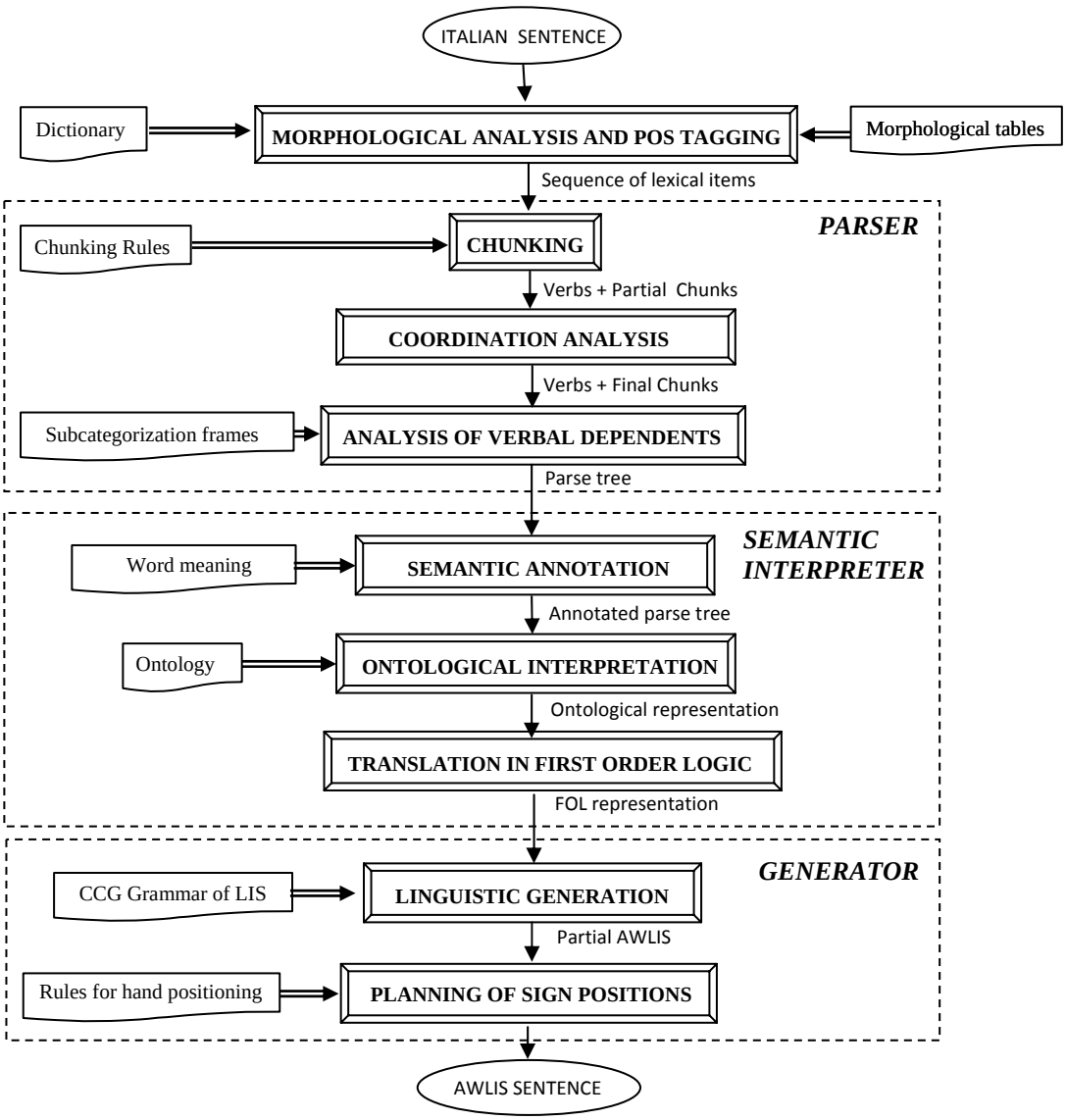

Figure 1 – Architecture of the Italian – AWLIS rule-based translator

The syntactic analysis includes two main steps:

1. Morphological analysis and dictionary access
2. Parsing

The basic idea is that in limited domains (as the one of weather forecasts) is possible to obtain a "deep understanding" of the meaning of texts. To get this result, we need the detailed syntactic structure of the input sentences and specific information about the meaning of the words appearing in the sentences. The syntactic structure is produced by the parser already available at UniToInf. It uses a morphological dictionary of Italian (about 25.000 lemmas) and a rule-based grammar that describes dependency structures. The final result is a "dependency syntactic tree", that makes clear the structural relationships occurring between the words of the sentence. Two examples are reported in fig.2 and 3. The first one is an example of a simple sentence, while the second includes some complex structures, as the ones associated with ellipsis and coordination. These two examples are used in this document to make clear the methodology. The two examples correspond to the sentences:

S-1: Locali addensamenti potranno interessare il settore nord-orientale
(Local cloudiness could concern the north-eastern sector)

S-2: Ancora correnti settentrionali, molto agitato lo Ionio, generalmente agitati gli altri mari.
(Again northern currents, very rough the Ionio, generally rough the other seas)

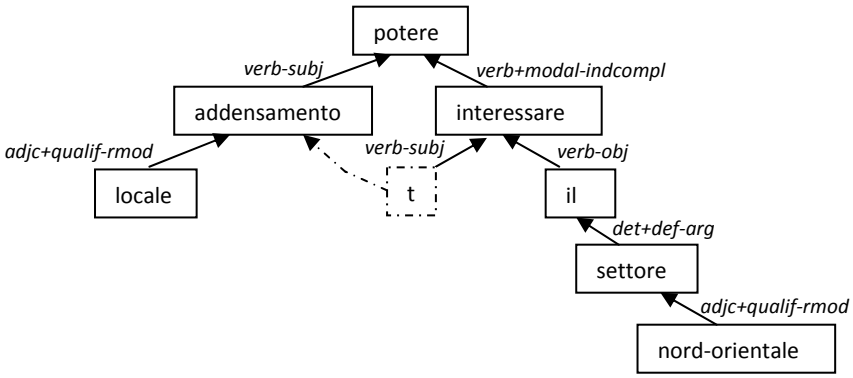

Figure 2 – Syntactic structure of "Locali addensamenti potranno interessare il settore nord-orientale"
(Local cloudiness could concern the north-eastern sector)

The basic features of the representation format appear in fig.2. Each word in the sentence is associated with a node of the tree. Actually, the nodes include other data (e.g. the gender and number for nouns and adjectives and the tense of the verbs) which do not appear in the figure because of space reasons. The nodes are linked via labeled arcs that specify the role of the dependents with respect to their governor (the parent). For instance, "addensamento" (cloudiness) is the subject (*verb-subj*) of the verb "potere" (to can), while "il" (actually, the whole subtree rooted in "il", "the") is the direct object (*verb*-obj) of "interessare" (to interest). In the figure, there is also a special node (t), which is a "trace". It specifies that the subject of "interessare" is "addensamento", although the node associated with it syntactically depends on "potere". In other words, this node, which does not correspond to any word in the sentence, enables us to specify that "addensamento" is a subject shared by "potere" and "interessare".

In the next example, the role of traces is more relevant, since they are used to represent "understood" verbs (verbal gaps). The sentence, in fact, is interpreted as if it were "Avremo ancora correnti settentrionali, molto agitato sarà lo Ionio, generalmente agitati saranno gli altri mari" (*We will have again northern currents, the Ionio will be very rough, the other seas will be generically rough*). In this case, the traces are of different

type, since they do not "co-refer" with other existing words (as in the "addensamento" case), but are forced by the parser into the tree. The choice of the verb (to have or to be) is made on a contextual basis, but does not substantially affect the subsequent behavior of the interpreter. So, we have that some (generic) group of entities (e.g. the listeners of the weather forecasts) "will have" some northern sea currents, and that the Ionic sea "will be" very rough, and that the other seas "will be" generically rough. Note that the commas are interpreted (in this example) as coordination operators (as "and"), and that the coordination is represented as a right-branching structure where the conjuncts are linked via <*coord*, *coord-2nd*> pairs of arcs.

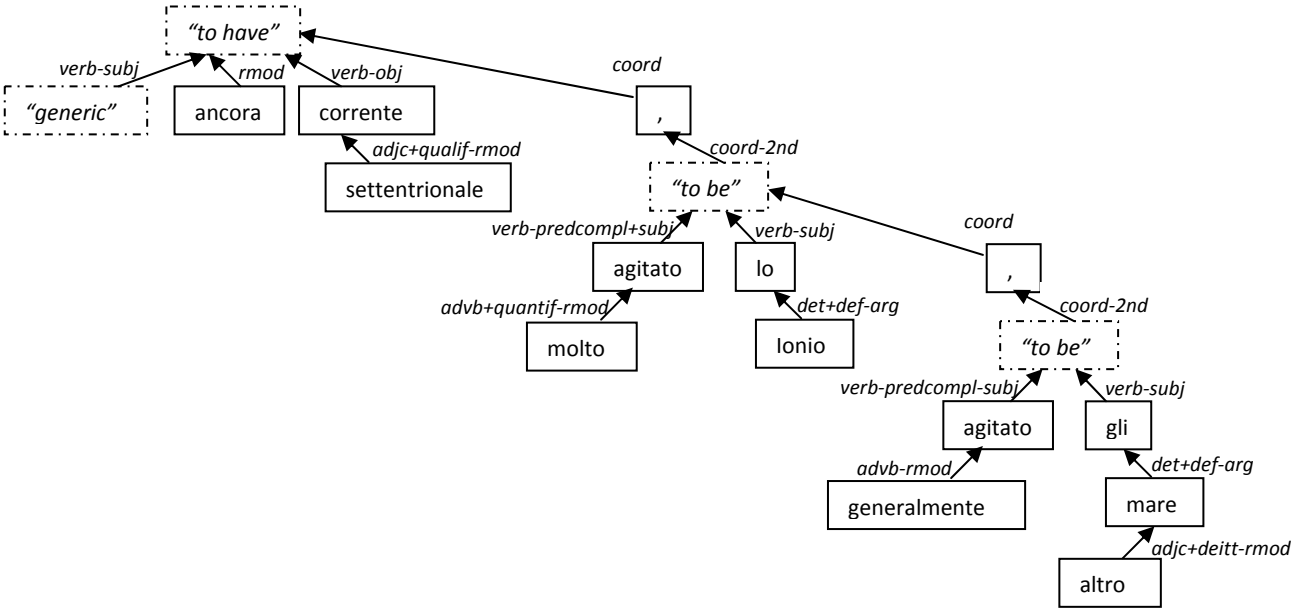

Figure 3 – Syntactic structure of "Ancora correnti settentrionali, molto agitato lo Ionio, generalmente agitati gli altri mari" *(Again northern currents, very rough the Ionio, generally rough the other seas)*

2 SKETCH OF THE MORPHOLOGICAL AND SYNTACTIC STEPS

The syntactic tree is built from the input by first tokenizing the sentence. This step identifies words, possible proper names or identifiers and punctuation marks. Standard words are then split in pairs <root, suffix>, where all different suffix possibilities are taken into account. For instance, the word "mente" would be associated with the following options:

1. mente+empty (in case "mente" is an invariable word, as "caos")
2. ment+e (for possible adjectives, nouns or verbs taking the suffix –e)
3. m+ente (in case "mere" were a verb, so that "mente" is its present participle, as for ardere→ardente)
4. empty+mente (for possible adverbials, as abile→abilmente)

Then, the various possibilities are matched with the dictionary. This step reveals that

a. "mente" does not exist as an invariable (option 1 discarded)
b. there exists a verb (ment+ire, to lie), that accepts, as third person of the present indicative tense, the suffix –e, and there exist two nouns (ment-a, mint, and ment-e, mind) which are compatible with the –e suffix, in the first case as feminine plural (mints), in the second as feminine singular (mind)
c. the verb "mere" does not exist (option 3 discarded)
d. the adverbial suffix is not compatible with an empty root (option 4 discarded)

The result of this step is as follows:

i. [lemma=mentire, category=verb, tense=present, person=3, mood=indicative]

ii. [lemma=menta, category=noun, gender=f, number=plural]
iii. [lemma=mente, category=noun, gender=f, number=singular]

This threefold ambiguity is submitted to a rule based Part Of Speech Tagger that, by inspecting the local context, chooses the most probable interpretation. For instance

- ... con le mente … (with the mints): interpretation ii [agreement with the article "le", feminine plural]
- … con la mente …(with the mind): interpretation iii [agreement with the article "la", feminine singular]
- … gli mente … (lies to him) : interpretation i [since there is no agreement between the article "gli", masculine plural, and any nominal interpretations, "gli" has been disambiguated as a dative clitic; consequently, "mente" has been disambiguated as a verb]

Note that also "le" is ambiguous (clitic dative pronoun feminine), so that in a context as "… non le mente spesso …" (… does not lie often to her …) the correct interpretation would the verbal one. Note also that the PoS tagger is based on a limited context, so that sometimes it makes the wrong choice. In a general context, the wrong analyses are around 3.8%, on the training Treebank they are around 2.1%, on the ATLAS texts, after careful manual refinement, they are around 0.5%.

After this step, the first example input sentence (Locali addensamenti potranno interessare il settore nord-orientale: *Local cloudiness could concern the north-eastern sector)* appears as follows:

Locali [lemma=locale, category=adjective, gender=m, number=plural]
addensamenti [lemma=addensamento, category=noun, gender=m, number=plural]
potranno [lemma=potere, category=verb, type=modal, mood=indicative, tense=future, person=3, number=plural]
interessare [lemma=interessare, category=verb, mood=infinite]
il [lemma=il, category=article, type=definite, gender=m, number=singular]
settore [lemma=settore, cetagory=noun, gender=m, number=singular]
nord-orientale [lemma=nord-orientale, category=adjective, gender=all, number=singular]
. [lemma=., category=punctuation]

This sequence of entries is submitted to the parser, which extracts the relations among words. It first looks for "chunks", i.e. subsequences of words usually concerning noun substructures. In this case, we get "Locali addensamenti" and "il settore nord-orientale", as shown in fig.4. Then, verbal subcategorization is used for attaching the chunks to the verbs and for inserting the standard traces. After a post-processing step that takes into account possible fragments still unattached, this leads to the final representation of fig.2.

This syntactic tree is then submitted to the semantic interpreter which, accessing the ATLAS ontology of the Meteo domain, builds the Ontological form, that is described in the next section.

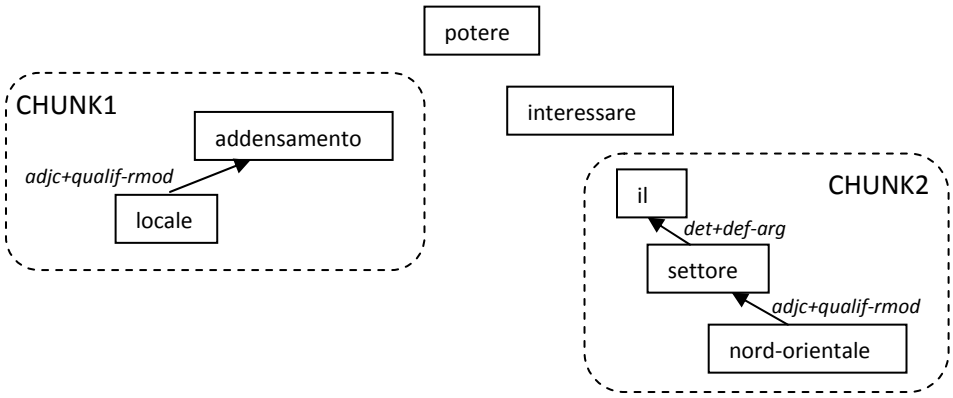

Figure 4 – Partial chunks built during the analysis of the sentence in fig.2

3 ONTOLOGICAL FORMS

In this section, we describe the semantic form which is obtained by the semantic interpreter. It is not the final result of the interpretation process, since it is converted into a First Order Formula before being sent to the generator of the LIS written form. However, it includes all the semantic pieces of information that have been extracted from the ontology starting from the input sentence. This section describes ontological forms, while the next one is concerned with the interpretation process.

In Box 1, we report the ontological form obtained for our first example, i.e. "Locali addensamenti potranno interessare il settore nord-orientale" (Local cloudiness could concern the north-eastern sector). The first four lines refer to the general principle that a sentence is part of a dialogue (in some cases, as in the Meteo domain, one-way dialogues). The actual content starts from the fifth line, specifying that this dialogue part concerns a *££situation*. Among the subclasses of situations there is *££status* which, in turn, has, among its subclasses, *££to-affect-1* (referring to a meteorological phenomenon affecting a geographic area). The *(SYNT n concept)* expressions are links to the syntactic structure: *££to-affect-1* is related to the "interessare" (to concern) syntactic item, so that the link says that the access to *££to-affect-1* has been made because of the presence of 'interessare', which is the fourth word of the sentence. Analogously *(SYNT 2 £clouds)* refers to the second word, i.e. 'addensamenti', *(SYNT 1 ££local-phenomenon)* to the first word (the adjective 'local'), and so on. They have no effect on the semantic interpretation of the ontological form, but make it a kind of 'annotation' of the sentence, which keeps the whole syntactic-semantic structure connected.

```
(ABOUT ££dialogue
 WHERE ((SYNT 0 ££dialogue)
            DOMAIN-OF &has-dial-topic
               RANGE ££dialogue-topic
                  HAS-SUBCLASS ££situation
                     HAS-SUBCLASS ££status
                        HAS-SUBCLASS (SYNT 4 ££to-affect-1)
                           DOMAIN-OF
                  (AND ((&affecter
                              RANGE ££entity
                                 HAS-SUBCLASS ££situation
                                    HAS-SUBCLASS ££event
                                       HAS-SUBCLASS ££weather-event
                                          HAS-SUBCLASS (SYNT 2 ££clouds)
                                             SUBCLASS-OF ££weather-event
                                                DOMAIN-OF &has-event-width
                                                   RANGE ££weather-event-width
                                                      HAS-INSTANCE (SYNT 1 £local-phenomenon))
                        (&affectee
                           RANGE ££entity
                              HAS-SUBCLASS ££spatial-location
                                 HAS-SUBCLASS ££geographic-area
                                    HAS-SUBCLASS (SYNT 6 ££it-geogr-area)
                                       HAS-INSTANCE £it-northeastern-area
                                          ARGUMENT-OF &has-it-area7
                                             RELINSTANCE &has-it-area-spec
                                                RANGE ££it-area-spec
                                                   HAS-INSTANCE (SYNT 7 £northeastern))))))
```

Box 1 – Ontological form of "Locali addensamenti potranno interessare il settore nord-orientale"
(Local cloudiness could concern the north-eastern sector)

Note that, for the sake of readability, some conventions have been adopted for ontology names:

- Concepts (classes) have a ££ prefix
- Instances have a £ prefix
- Three special individuals have a § prefix (the system: *§myself*, the user: *§speaker* and an indefinite individual: *§indef-ref*)
- Relations and relation instances have a *&* prefix
- Datatypes have a $ prefix (*$string* and *$sound*).
- Pseudo-concepts, usually having a bridging role to syntax, have a – prefix

In general, an ontological form is a path in the ontology where the 'rest' of the path (all items except the first) express a restriction on first item. For instance,

HAS-SUBCLASS (SYNT 2 ££clouds)
SUBCLASS-OF ££weather-event
DOMAIN-OF &has-event-width
RANGE ££weather-event-width
HAS-INSTANCE (SYNT 1 £local-phenomenon))

Specifies that the sentence is not referring to "any cloud", but just to clouds whose *££weather-event-width* is *£local-phenomenon*. In this specific case, the path describes a movement upward in the taxonomic backbone of the ontology, which is needed because all *££weather-event*, and not only *££clouds*, have a width.[1] In some cases, however, more than one restriction is given. In these cases, the path 'branches', due to the insertion of the AND connective. So, the final form is a tree, where all branchings are associated with multiple restrictions. In our example, *££to-affect-1* has two restrictions, corresponding to the thematic roles of the verb "interessare": the *&affecter* (i.e. the 'agent') are the local clouds, while the affected entity is the Italian northeastern region. Note that, at this level of description, there is no difference between arguments and adjuncts, since all of them act as 'restrictions' of the situation description. However, this distinction is relevant during the semantic interpretation process, as will be explained below.

Although, at the current level of implementation, we do not cope with logical connectives as disjunction or implication, the AND operator that conjoins restrictions does not cover all cases. In the example above, we have seen that it is used to put constraints on a single situation, i.e. for specifying affecter and affectee. A different case is the one of sentences as "Domani avremo piogge al sud e cielo sereno al nord" (Tomorrow we will have rain in the South and clear sky in the North) . Here, two situations are being described, so that we cannot simply specify that the (and-ed) constraints are "(location North) (location South) (event-type rain) (event-type clear-sky) (time tomorrow)": in any reasonable interpretation, there is no model satisfying this bundle of constraints. Rather, what is needed is something as "(event ev1) (location ev1 North) (event-type ev1 rain) (time ev1 tomorrow) (event ev2) (location ev2 South) (event-type ev2 clear-sky) (time ev2 tomorrow)", plus a statement that both ev1 and ev2 are claimed to exist. Since the reification of events is a basic feature of the final FOL representation (see D4.3#5), the inclusion of event variables (e1 and e2) is not a problem. However, they have to be joined by a different type of AND, that we call EVENT-AND.

The general structure of the ontological form is given by the grammar appearing in Box 2. Apart from the initial four lines (referring to the inclusion of the sentence in a dialogue), a full ontological form is an **<a-concept-path>**. This a "concept description", which is either single or complex, where complex descriptions are associated with the AND and EVENT-AND operators (the double-plus superscript is a shorthand for 'at least two').

The actual "description part" (**<a-concept-descr>**) is a simple or AND-ed path starting with an arc linking the concept to another concept (**<conc-conc-op>**), to a relation (**<conc-rel-op>**), or to an instance (**<conc-inst-op>**).

[1] It can be argued that this is irrelevant for the semantics of the sentence, since it only encodes an ontology-internal application of inheritance, and that it could be avoided, for instance, by compiling the ontology. We repeat that the Ontological Form mirrors exactly the content of the ontological knowledge source. In any case, these movements are avoided in the Simplified Ontological Form, which is described in D4.3#4 (From Ontological form to FOL).

According to the starting arc of the description, the rest of the path (after the operator) is another "concept path, a "relation path" (which is the counterpart of a concept path) or an instance path. Relation paths can be simple or AND-ed, but they cannot exploit EVENT-AND, while instance paths can only be simple.

The description of relations is similar to the one of concepts, where <**rel-rel-op**> replaces <**conc-conc-op**> and <**rel-relinst-op**> replaces <**conc-inst-op**>. Instance and relinstance descriptions are analogous, but simpler.

<ontological-form> → (ABOUT ff dialogue
WHERE ((SYNT 0 ff dialogue)
DOMAIN-OF &has-dial-topic
RANGE ff dialogue-topic
HAS-SUBCLASS **<a-concept-path>**))

<a-concept-path> → **<concept-path>** | (and **<concept-path>**$^{++}$)) | (event-and (**<concept-path>**$^{++}$))

<a-relation-path> → **<relation-path>** | (and (**<relation-path>**$^{++}$))

<concept-path> → **<concept>** | **<concept>** **<a-concept-descr>**
<relation -path> → **<relation >** | **<relation>** **<a-relation-descr>**
<instance -path> → **<instance>** | **<instance>** **<instance-descr>**
<relinstance -path> → **<relinstance>** **<relinstance-descr>**

<a-concept-descr> → **<concept-descr>** | (and (**<concept-descr>**$^{++}$))

<a-relation-descr> → **<relation-descr>** | (and (**<relation-descr>**$^{++}$))

<concept-descr> → **<conc-conc-op>** **<a-concept-path>** |
<conc-rel-op> **<a-relation-path>** |
<conc-inst-op> **<instance-path>**
<relation-descr> → **<rel-conc-op>** **<a-concept-path>**
<rel-rel-op> **<a-relation-path>** |
<rel-relinst-op> **<relinstance-path>**
<instance-descr> → **<inst-relinst-op>** **<relinstance-path>**
<relinstance-descr> → **<relinst-inst-op>** **<instance-path>** | **<relinst-inst-op>** **<concept-path>**

<conc-conc-op> → SUBCLASS-OF | HAS-SUBCLASS
<conc-rel-op> → RANGE-OF | DOMAIN-OF
<conc-inst-op> → HAS-INSTANCE
<rel-conc-op> → RANGE | DOMAIN
<rel-rel-op> → RESTRICTS | RESTRICTED-BY
<rel-relinst-op> → RELINSTANCE
<inst-relinst-op> → ARGUMENT-OF | VALUE-OF
<relinst-inst-op> → ARGUMENT | VALUE

<concept> → *concept-id* | **(**SYNT *line-id concept-id***)**
<instance> → *instance-id* | **(**SYNT *line-id instance-id***)**
<relation> → *relation-id* | **(**SYNT *line-id relation-id****)***
<relinstance> → *relinstance-id*

Box 2 – Grammar of the ontological forms

4 SEMANTIC INTERPRETATION

This step includes two sub-steps. The first one concerns the annotation of the syntactic tree with the meaning of the words. The included meanings act as access points to the ontology and are the starting points for the search of a path connecting two words (semantic connection). The second sub-step builds the actual ontological representation in a quasi-compositional way, by putting together paths found in the ontology in a single representation which is a sub-graph (with possible redundancies) of the ontology itself. These two sub-steps are described in detail in the present section.

4.1 Tree Annotation

The syntactic tree is annotated in two ways: it inserts in each lexical node its semantic value and adds some non-lexical nodes. For content words, the semantic value is a concept of the ontology, or an identifier of an instance together with its immediate class. For non-content words, it depends on the category of the word.

4.1.1 Adding word meanings

The basic function of this step is to look in the dictionary for the meaning (ontology concept) of a word. However, this step takes care of anaphors in case of pronouns and traces.

The meaning (*SEM* feature in the tree) is determined by accessing the dictionary. This is the standard way of finding the meaning, which applies to the so-called "content words". So, if the lemma associated with the current node of the syntactic tree is "pioggia", then the SEM of that node is set to *££rain*. Note that *££rain* is a node in the ontology, so that meanings are basically expressed in terms of ontology nodes.

However, this method applies also to most non-content words, as articles and prepositions. Most of them have an associated meaning (as *--indef-art*, for the indefinite articles), which is not used in the subsequent steps. An exception are prepositions, where the meaning refers to an entry in a "preplate" (prepositional templates) table. Since this table is used in the search for shortest paths (see §4.3), it will be described there. Here, what we have is that a node whose lemma is "di", "in", etc. will be annotated with *--di-relation*, *--in-relation*, which will act as a key for accessing the preplate table.

Although these meanings are conceptually part of the dictionary, they are stored in a separate file, whose entries are pairs <lemma meaning> (as, for instance <pioggia *££rain*>). Some example entries are reported in Box 3.

With respect to relations, it is worth noting that, currently, no word refers to ontology relations. This is due to the fact that most actual relations are reified, i.e. they are represented by classes, while actual ontology relations are associated with roles (arguments) of relation. This will be discussed in the section on ontology. What is important in the present context is that verbs are associated with a "thematic grid", specifying the correspondence between the syntactic label of the verbal arguments and the ontology relations associated with the thematic role of the argument (for a brief introduction to the use of relations, see §4.2.4; for more details D4.3#2). An example is the verb 'interessare' (in general 'to interest', but in the present context 'to affect'), appearing in Box 3.

Of course, it is possible to express the fact that a word has multiple meanings. Although this is not common in a simple domain as the weather one, this feature can be useful for further extensions. There are three ways for expressing ambiguity:

- Simple ambiguity

The corresponding entry has the form <*lemma* <*meaning1 meaning2* ....>>. In this case, all meanings are stored in the tree. In the query construction phase, one of the meanings will be selected, on the basis of the ontological distance from the surrounding words.

addensamento	££clouds
corrente	££sea-current
mare	££sea
settore	££it-geogr-area
potere	££modal-can
interessare	££to-affect-1 (thematic-grid (verb-subj &affecter) (verb-obj &affectee))
locale	£local-phenomenon
domani	£tomorrow
stasera	£today-evening
molto	--intensifier-adv
generalmente	££empty-conc
di	--di-relation
in	--in-relation
altro	--deictic-specif-other
poco	--q-little
il	--def-art
quale	(adj --q-art) (pron --q-pron)
Sardegna	£Sardegna
Ionio	£Ionio
sud	£south

Box 3 – An excerpt of the table that maps words onto concepts

- Category-controlled ambiguity

This happens in case different lemmas have the same surface form. An example is the word "quale" (what), which could be an adjective ("What movie have you seen?") or a pronoun ("What have you done?"). This would not be a problem if the meanings were actually stored in the dictionary entries (which are distinct). The notation currently used is <*lemma* <*category1 meaning1*> <*category2 meaning2*> ...>.

- Ambiguity with disambiguators

These are special cases, where one of the dependents of the word, if present, can disambiguate the meaning. An example is "number": if the phrase is "phone number", then it has a specific ontological interpretation *(££phone-number*), otherwise the reference is to the standard *££number* concept (this is not an example from the ATLAS domain). This is expressed in the form <*lemma meaning modifier-lemma modified-meaning*>.

The found meaning, either non-ambiguous or ambiguous (just in case of simple ambiguities, since the other two cases should have been disambiguated) is stored in the *lexmean* of the *SEM* feature of the tree node.

A different type of treatment is applied in case of instances (as the one referred to via proper names, as "Ionio"). In these cases, the annotator carries out a double work. First, it looks for the meaning in the table (as above). Then it checks if the word meaning concerns an instance (it is only at this stage that instances are recognized, not on the basis of the syntactic features of the tree node). In this case, it gets from the ontology the immediate class of the instance. The two data (instance identifier and immediate class identifier) are used to annotate the tree: the class name is stored in the *lexmean* subfeature of *SEM* (exactly as above), while the instance identifier is stored in the *ident* subfeature of *SEM*. So, for instance, for Ionio, we get:

<*SEM* < <*lexmean ££it-sea*> <*ident £Ionio*> > >

Where *££it-sea* is the concept associated with Italian seas. Note that, while the *ident* value is obtained from the table of word meanings, its class is found by inspecting the ontology.

4.1.2 Adding pragmatic information

The second type of annotation consists in adding some extra nodes that make explicit the communicative role of the sentence in the dialogue. Although this does not seem to be very relevant here, this paves the way for the application to other domains. Actually, weather forecasts (as all TV news) are one-way, i.e. they are not a dialogue, but all sentences include information flowing from the TV speaker to the audience. In general, however, a dialogue could include requests for information, replies, orders, and so on. Since this information concerns a different domain (i.e. pragmatics), the choice of expressing it as tree annotation is questionable. However, it simplifies the processing, and we accept critiques concerning this point.

In particular, in this domain, the syntactic tree is extended by inserting a *££give-info* node as the root of the tree. This is associated with the action of providing information and the agent is here *§myself*, i.e. the agent playing the role of the TV speaker. This new piece of tree could be paraphrased as "I, the system, tell you that …", where the original tree is linked to this new part via a node whose meaning is given as *–about-relation*. The annotated syntactic trees associated with the two example sentences are reported in fig. 5 and 6.

In the figures, we have used the same notation for "pragmatic" nodes and traces, since both of them correspond to items that are lexically empty. In square parentheses there is the ontological concept associated with the node lemma. In case it carries also information about an instance, this is included in curled brackets.

Note that the words "molto" (very) and "altro" (other) refer to special ontological items; in particular, they are used, in the next phase of access to the ontology, to carry out special operations, in order to get the interpretation of "molto agitato" (very rough) and "altri mari" (other seas).

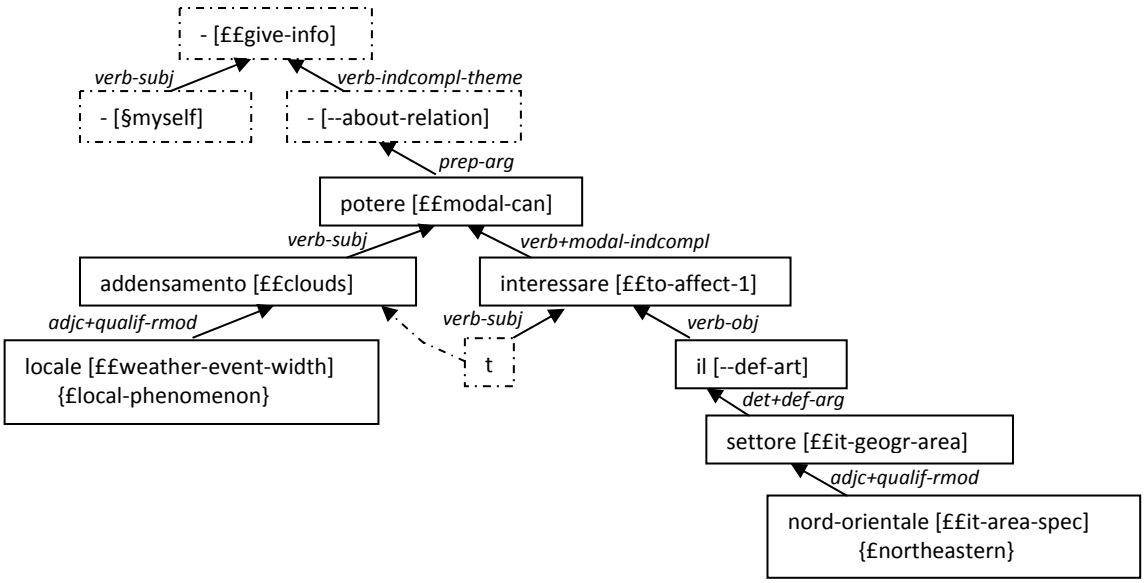

Figure 5 – The annotated syntactic tree of "Locali addensamenti potranno interessare il settore nord-orientale"
(Local cloudiness could concern the north-eastern sector)

4.2 The Ontology

A full description of the ontological knowledge source used in ATLAS for semantic interpretation can be found in deliverable D4.3#2. Here, we provide some basic notions, needed to understand how it is applied in the process of semantic interpretation.

The ATLAS ontological knowledge base is a formal (partial) description of the domain of application. It is formal, in the sense that its primitives are formally defined. It is partial, since it does not include all axioms that provide details about the relationships between the involved concepts.

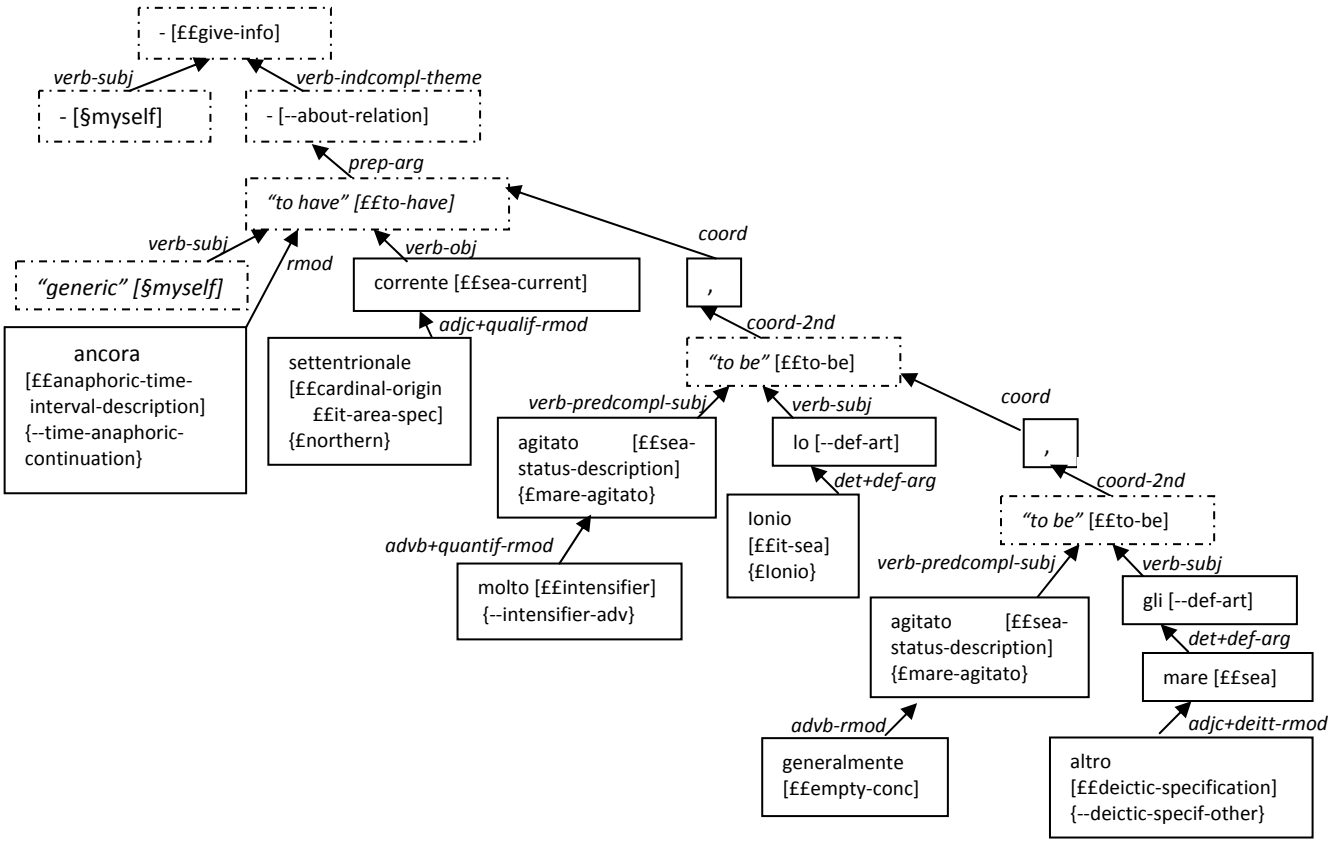

Figure 6 – The annotated syntactic tree of "Ancora correnti settentrionali, molto agitato lo Ionio, generalmente agitati gli altri mari" *(Again northern currents, very rough the Ionio, generally rough the other seas)*

4.2.1 The top level

The ontology currently includes 176 concepts. Its top level is depicted in fig.7.

The root of the hierarchy (*ff entity*) is the direct superclass of:

- *ff description*: we keep apart the actual situation and its description. For instance, if today is April 28, then "today" is a *ff deictic-description* of a particular instance of a day. "April 20, 2012" is another description (absolute) of the same instance.
- *ff geogr-part-selection-criterium*: in descriptions, a particular instance (or group of instances) can be identified by a general class term (e.g. area) and a selector (e.g. northern). This concept refers to the parts of the reality that can act as selectors. For instance, the cardinal direction can be such a criterion for geographic parts, while a date is not.
- *ff spatial-location* and, in particular, *ff geographic-area*: any weather situation holds in a specific place. A geographic area can be an Italian region, a group of regions, a sea, or may be identified by specifying a cardinal direction (North, South, …).
- *ff situation*. This is the most relevant subclass in the present context, since it, among other things, refers to the possible weather situations that are described in the news (*ff meteo-status-situation*). It is shown in fig.8. It may concern the status of the seas, a generic weather status (in particular, if it is stable or not) or possible atmospheric events (e.g. rain or clouds).
- *ff reified-relation*: this refers to most actual 'world' relations holding among individuals.

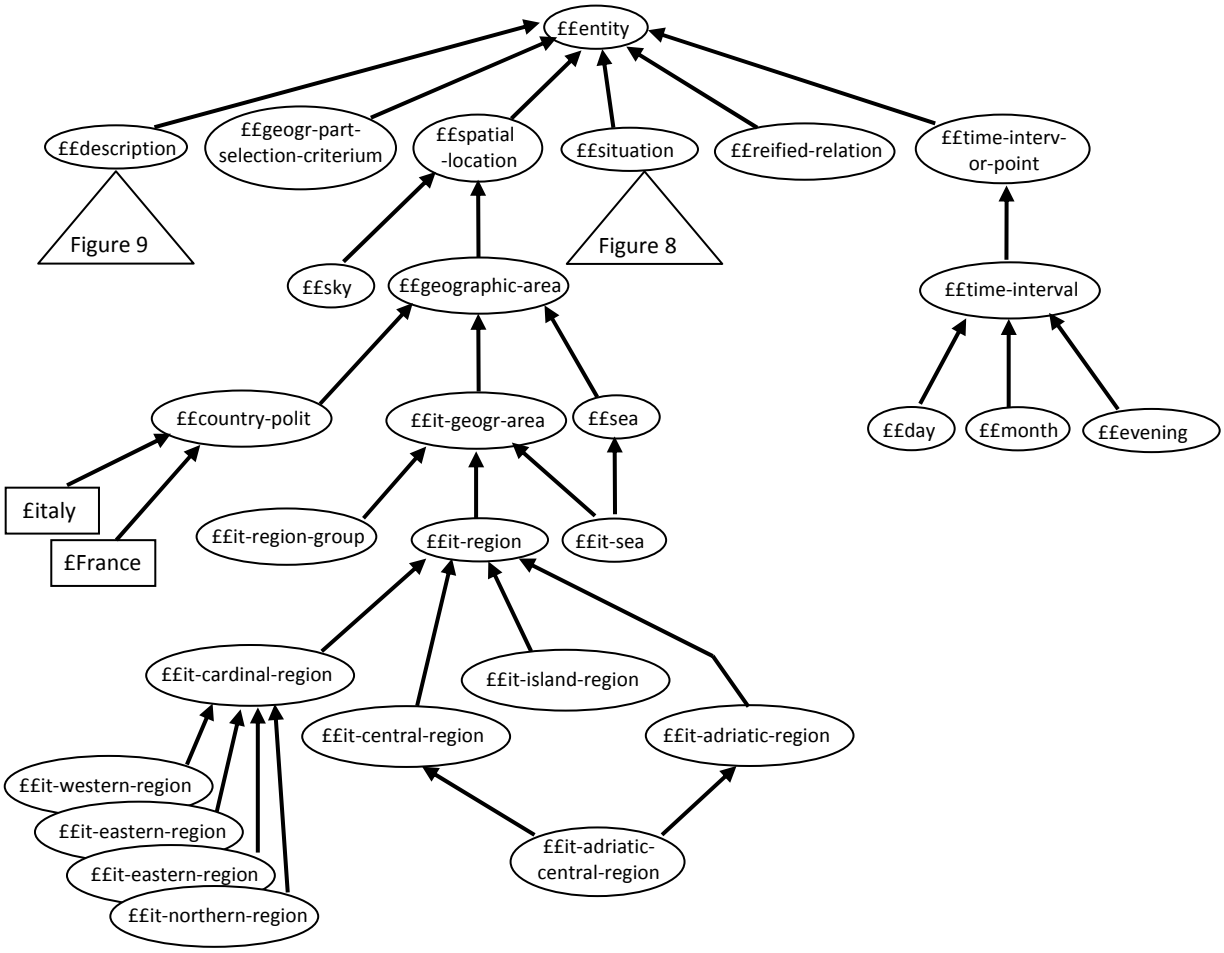

Figure 7 – The top-level of the ontology. Triangles refer to parts depicted in other figures. Rectangles are instances

- *ff time-interv-or-point:* any weather situation holds in a specific temporal interval (time points are not currently used in ATLAS). This could last one or more days or a part of a day. Currently, we do not handle dates (e.g. "on May 3 it will rain"), but only deictic terms (as today and tomorrow). Expression as "in the evening" are interpreted anaphorically, i.e. on the basis of the current context: if the context is referring to "today", then it is interpreted as "today evening", for "tomorrow" as "tomorrow evening", etc.

Other subclasses of *ff entity* non depicted in the figure are:

- *ff datatype*: this has just two instances *$string* (written text) and *$sound* (for speech).
- *ff degree*: this enables one to specify, for instance, that the weather is more or less stable
- *ff television-program*: just to enable for reference to the "weather forecast" program.
- *ff measureunit* (e.g. degrees, for the temperature)
- *ff comparison-operator* (e.g. greater than)
- *ff math-function* (e.g. average)
- *ff evaluable-entity* (e.g. evenings, for interpreting greetings as "good evening")
- *ff image* (for images coming from satellites)
- *ff space* (the location of satellites)
- *ff ownable-entity* (idiosyncratic for phrases as "tomorrow we will have rain")
- *ff property-value* (the specific value of temperature, or degree of humidity, etc.)
- *ff math-value* (any value coming from the application of a math function, as average)
- *ff property* (properties of entities, as the temperature of the air)
- *ff sequenceable-entity* (entities that can be put in a sequence, as the days of a month)
- *ff entity-sequence* (the specific sequence of the days of a given month)

The most relevant of them are perhaps in the group including *££property*, *££property-value*, *££measureunit*, *£evaluable-entity*, *££degree*, *££math-function*, all of which are concerned with evaluation of weather conditions (humidity, coldness, rain strength, etc.).

For all details about these classes and other parts of the ontology, refer to the ATLAS deliverable D4.3#2. Here, we only spend a few more words about situations, since they include meteorological states and events and about descriptions.

4.2.2 Situations

££situation is (partially) depicted in Fig.8; we will focus on *££meteo-status-situation*, which has four subclasses. The figure should be self-readable, apart from a couple of observations. First, the choice to have *££sea-current* as a situation of the sea is questionable. However, this seems to be a reasonable choice for handling phrases as "correnti settentrionali" (northern currents), where the presence of a current characterizes the status of the sea. Second, and more important, *££weather-event* is both a subclass of *££event* (reasonable) and of *££meteo-status-situation* (apparently less reasonable). This enables us to characterize the double perspective under which 'rain' (say) is seen. The standard one is that it is something happening (an event); the second one concerns the view of a person that exits her house and finds herself in a 'raining' situation. This double view seems acceptable and is the way we talk about it.[2]

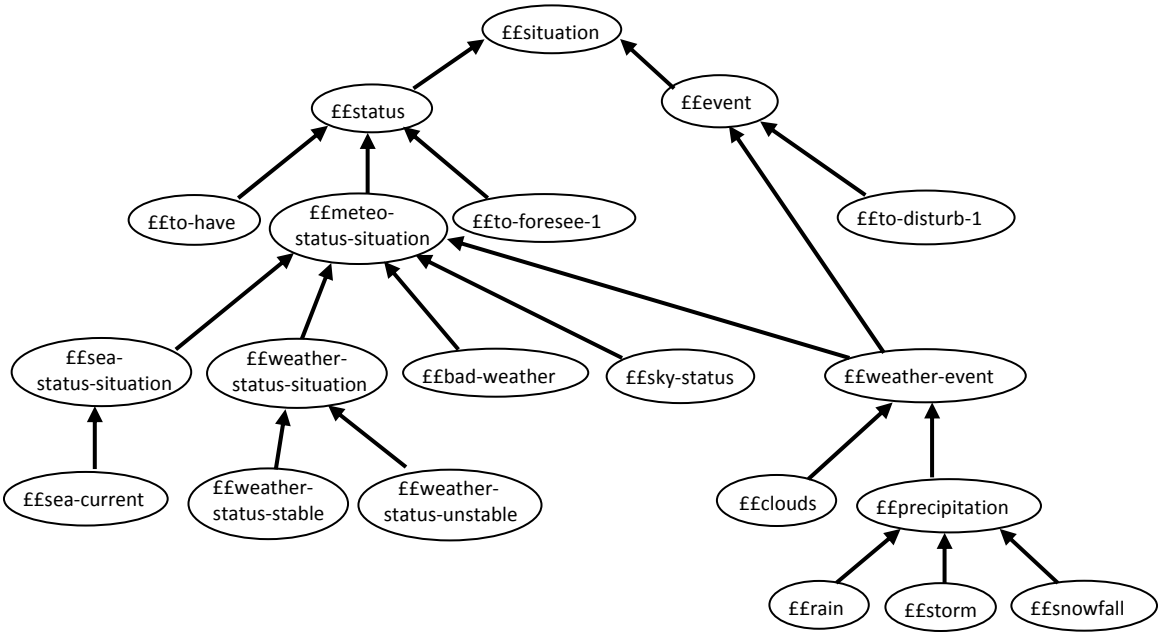

Figure 8 – The portion of the ontology describing situations

4.2.3 Descriptions

In fig.9, we have the part of the ontology concerning descriptions. Particular relevance have the deictic descriptions, since most temporal descriptions (today, tomorrow, but also the weekday names, as Monday, Tuesday, …) are deictic in nature. These can easily be handled by means of procedures (not implemented), that obtain the current day through calls to system functions. Since the current application of the interpreter is to translation, we are not interested in knowing which is the 'actual' day, since 'today' has a corresponding deictic expression in the target language. This could be relevant for languages that do not have deictics.

[2] Remember that it is a deliberate choice to make the ontology largely dependent on language: concepts are in our minds but their organization partially depends on the way we talk about them.

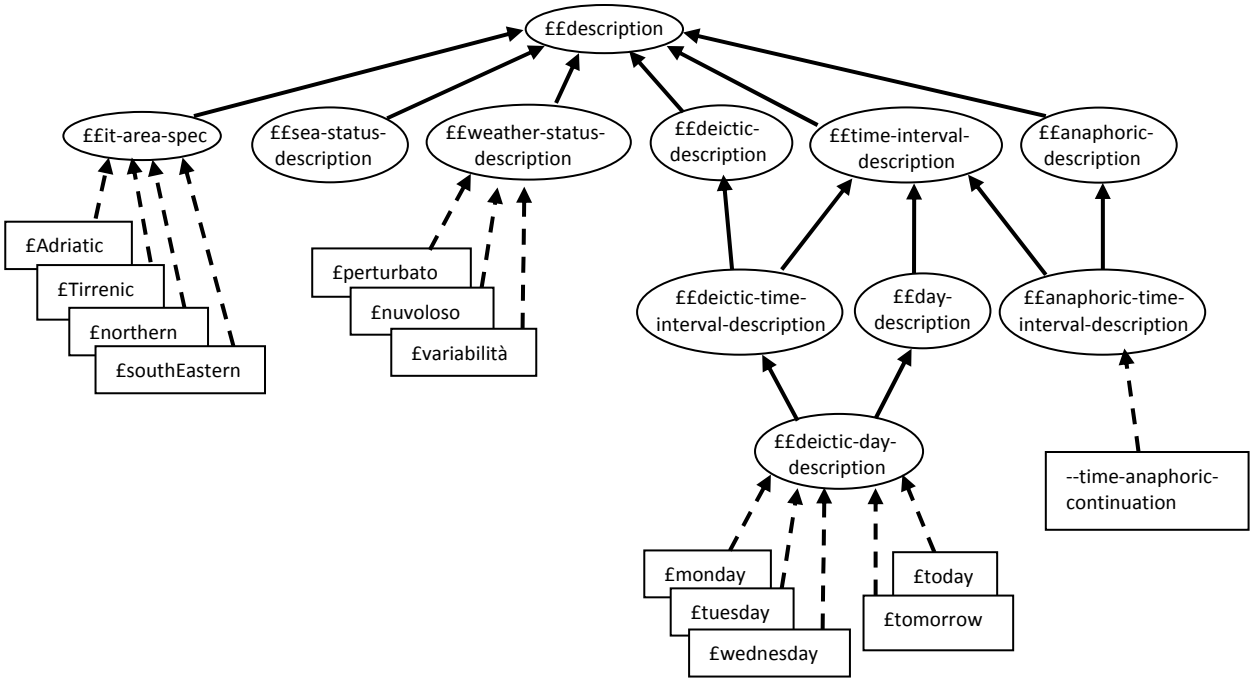

Figure 9 – The portion of the ontology concerning descriptions (with some example instances, depicted as boxes)

4.2.4 Relations

The last relevant portion of the ontology concerns relations. Although the ontology has no axioms, the (class) concepts are connected via relevant relations, and the relations are the basic passages to form paths. All relations in the ontology are binary, so that the representation of relations of arity greater than 2 requires that they be reified. For instance, almost all relations associated with natural language verbs undergo this treatment. This amounts to say that, for instance, the verb "to eat" refers to a class of events that can have common features. The fact that "to eat" is a transitive verb (so that it could be associated with a binary relation) is only partially relevant in the present context, since the associated actions have many features that do not appear as obligatory arguments of the verb. For example, each eating action occurs in some place, at some time, is made with some instrument(s), and so on.

Each true (binary) relation is characterized by some features, listed below:

- Domain (a class): Obligatory
- Range (a class): Obligatory
- Funct (1:1, 1:N, N:1, N:M): Obligatory, but not currently used by the procedures
- Restricts (a relation): Optional

In fig.10, we report two example relations that occur in the weather forecast domain. In fig.10, relations are represented as arrows with small boxes. The *domain* of the relation is the node that the arrow leaves, while the *range* is the node that the arrow enters. The name of the relation is depicted near the small box. The functionality infos have the usual meaning: 1:1 means that both the relation and its inverse are functional; 1:N means that each individual of the domain can be associated with N individuals of the range, but not viceversa. The converse for N:1. N:M refers to the absence of functionality constraints.

The dashed link connecting *&has-meteo-time* and *&has-sea-status-time* specifies that the second *restricts* the first. The notation is a bit imprecise here, since the *restricts* relation is oriented, but the involved concepts should make clear the direction; this is made just for space reasons.

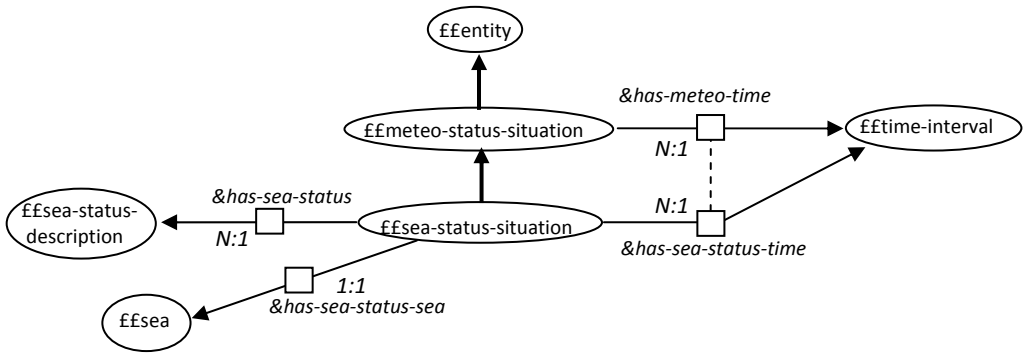

Figure 10 – Use of relations for representing the features of *ff sea-status-situation*

4.3 The shortest paths

The actual semantic interpretation of the input involves a match between the items appearing as annotations in the syntactic trees and nodes in the ontology. In order to build the final representation, two types of operations are required; the first consists in the search for a semantic connection between pairs of concepts, the second one in the integration of the various connections in a single integrated structure. In this subsection, we describe the first of these operations.

The search for connections is based on the idea that the shortest path that can be found in the ontology between two nodes represents the stronger semantic connection between them; consequently, it is such a path that must be used to build the semantic representation. We can refer to Fig.5, where the leftmost subtree describes the syntactic structure of "locali addensamenti" (local cloudiness). By paraphrasing this expression, we could say that it refers to "a cloudiness whose area of presence (or width) is local". By inspecting the figure, we see that "addensamenti" has been annotated with *ff clouds*, and that "locali" has an annotation whose concept of reference (*lexmean*) is *ff weather-event-width*, and whose instance (*ident*) is *f local-phenomenon*.

The paraphrase mentioned above is obtained by looking for the shortest path between the ontology nodes *ff clouds* and *ff weather-event-width*. In fact, the ontology specifies that *ff clouds* is a subclass of *ff weather-event* and that all *ff weather-event* are linked via the relation *&has-event-width* to the class *ff weather-event-width*. Since *f local-phenomenon* is an instance of *ff weather-event-width*, we can thus obtain the representation of this fragment:

ff clouds SUBCLASS-OF ff weather-event
DOMAIN-OF &has-event-width
RANGE ff weather-event-width
HAS-INSTANCE f local-phenomenon

Of course, the interpretation process is not always so simple. For instance, for "settore nord-orientale" we must get in some way a reference to the single ontology instance *f it-northeastern-area*, so that the shortest path on the ontology connecting *f northeastern* (which is an instance of *ff it-area-spec*) and *ff it-geogr-area* must involve *f it-northeastern-area*. Since this process involves some technical issues, we report here only the final result, only noting that *&has-it-area-7* is the *relinstance* concerning "northeastern area".

ff it-geogr-area HAS-INSTANCE f it-northeastern-area
ARGUMENT-OF&has-it-area7
RELINSTANCE &has-it-area-spec
RANGE ff it-area-spec
HAS-INSTANCE f northeastern

4.4 The preplate table

In this paragraph, we present some basic ideas about the treatment of prepositions. Compare two expressions as "the first chapter of the report" and "the language of the report". In this case, the search for shortest path

could suffice for determining that in the first expression there is an implicit reference to the *££part-of* relation, while in the second one, the relevant relation is something that could be *££written-in*. On the other hand, in cases as "the present of Mary" vs. "the present for Mary", it is the preposition that enables one to distinguish the "giver" from the "receiver". In order to cope with these cases, a specific data structure, called *preplate* (prepositional template) table is used. For correctly analyzing the last example above, the table must include two entries, the first one associated with *–of-relation*, specifying that if "of" connects *££present* (or any subclass or instance of its), with *££person* (or any subclass or instance of its), then it has to be interpreted as a reference to *&giver*, while the entry associated with *–for-relation* will specify that, in the same context, it has to be interpreted as *&receiver*.[3]

With respect to the standard analysis (search for shortest paths), there is just a small difference, i.e. that, in these cases, the search is based on three parameters instead than two: in fact what is looked for is the shortest path connecting *concept1* and *concept2* (as before), but constrained to pass through *concept3*, which is the concept got from the preplate table in the given context. Consequently, in "the present for her brother", we will have that the search will be for the shortest connection between *££present* and *££brother* (a subclass of *££person*), constrained to pass through *&receiver*. The result of the search is a path that, as the ones described before, will be used as one of the inputs for the next step.

4.5 Path Composition

The syntactic tree is recursively inspected, starting from the root, launching the search for shortest paths between pairs of nodes and merging the results.

The goal of the entire process is to build a representation that expresses in semantic terms the description of an item, which, in the present context, usually is a weather situation. In order to do this, the interpreter analyses the substructures starting from the root of the tree and collecting its 'restrictions'. For instance, in "Domani avremo annuvolamenti nel settore occidentale" (Tomorrow, we will have cloudiness on the western area), we have that the pivot is "avremo", while "domani", "annuvolamenti" and "nel settore occidentale" are the restrictions. Note that the interpretation is recursive, so that we first build the semantic representation for the restrictions, and then merge them: the actual sequence of partial interpretations starts from the leaves of the tree and moves upward.

The overall organization of this process is shown in Box 4. We note that the scheme depicted in the figure refers only to the ATLAS domain, where the top "pragmatic" node is always *&give-info*. In general, the top-most procedure is able to take into account also other cases (in particular, queries), but the presentation of this type of processing is omitted here. When the top meaning is *££give-info*, the starting point of the analysis is easily found by traversing (downward) the arc labeled as *verb-indcompl-theme* and *prep-arg* (see figg.5 and 6). In this context, this amounts to retrieving the annotated tree prior to the insertion of pragmatic nodes.

The actual interpretation is carried on assuming that the root of the tree can be semantically ambiguous (i.e. its meaning – *SEM* – corresponds to more than one concept in the ontology). In this case, different interpretations are built independently of each other. Since this situation does not occur in the ATLAS corpus, just one interpretation is built.

For each root meaning, the pair <meaning (unambiguous) + restrictions> is interpreted by analyzing separately the various restrictions (function *build-sing-r-sem*)[4], so that we have

$$<meaning, <restr_1, restr_2, \ldots, restr_m>> \rightarrow <interp_1, interp_2, \ldots, interp_m>$$

[3] This is not accurate. Actually "the present of Mary" is ambiguous between *&giver* and *&receiver*. In this case, the preplate table reports both alternatives, and the choice (which is strongly context-dependent) is currently made at random. However, this is not the case for "present for Mary", so the preplate mechanism is useful in many cases.

[4] Build-sing-r-sem carries out different operations according to the root category of the tree being interpreter. The dashed rectangles in the figure list some of these possibilities.

build-restr-sem (tree restrictions) →

build-sing-restr-sem (root-single-meaning restrictions) →

BREADTH-COMPOSE

(build-sing-r-sem (root-single-meaning single-restriction) **1st restriction**

build-adj-sem (root-single-meaning single-restriction)
DEPTH-COMPOSE
(*find-shortest-path (root-single-meaning adj-meaning) +*
build-restr-sem (adj-tree adj-restrictions)) **OR**
build-noun-sem (root-single-meaning single-restriction)
DEPTH-COMPOSE
(*find-shortest-path (root-single-meaning noun-meaning) +*
build-restr-sem (noun-tree noun-restrictions) **OR**
build-verb-sem (root-single-meaning single-restriction)
DEPTH-COMPOSE
(*find-shortest-path (root-single-meaning verb-meaning) +*
build-restr-sem (verb-tree noun-restrictions) **OR**
…..

build-sing-r-sem (root-single-meaning single-restriction) **2nd restriction**

build-adj-sem (root-single-meaning single-restriction)
DEPTH-COMPOSE
(*find-shortest-path (root-single-meaning adj-meaning) +*
build-restr-sem (adj-tree adj-restrictions)) **OR**
build-noun-sem (root-single-meaning single-restriction)
DEPTH-COMPOSE
(*find-shortest-path (root-single-meaning noun-meaning) +*
build-restr-sem (noun-tree noun-restrictions) **OR**
build-verb-sem (root-single-meaning single-restriction)
DEPTH-COMPOSE
(*find-shortest-path (root-single-meaning verb-meaning) +*
build-restr-sem (verb-tree noun-restrictions) **OR**
…..

…..

Box 4 – Organization of the semantic interpreter

Where each $interp_i$ is the result of the interpretation of <*meaning*, $restr_i$> and is a path (in general a tree structure) extracted from the ontology, on the basis of the "shortest paths" starting from *meaning*. In its simplest form, such a path (the result of an application of *find-shortest-path* of Box 4) is:

$<meaning=concept_0, concept_{i1}, concept_{i2}, …., concept_{in_i}>$

So what we have now (at the end of all recursive calls) is a set of paths such as:

$<meaning=concept_0, concept_{11}, concept_{12}, …., concept_{1n_1}>$

$<meaning=concept_0, concept_{21}, concept_{22}, …., concept_{2n_2}>$

……

$<meaning=concept_0, concept_{m1}, concept_{m2}, …., concept_{mn_m}>$

In order to "merge" these paths (BREADTH-COMPOSE, in Box 4), we observe that all restrictions are currently considered as conjoined (with a logical "and"). In order to simplify the expression, we check if they have some common prefixes, so, in case, for example m=3, $concept_{11} = concept_{21} = concept_{31}$ and $concept_{22} = concept_{32}$ we obtain, as final result:

$<meaning=concept_0, concept_{11},$
$<and <concept_{12}, concept_{13} \ldots., concept_{1n_1}>$
$<concept_{22},$
$<and <concept_{23}, \ldots., concept_{2n_2}>$
$<concept_{33}, \ldots., concept_{3n_3}>>>>>$

In our first example (whose annotated parse tree, presented as fig.5, is repeated below, for the convenience of the reader, as fig.11) , *build-restr-sem* is activated on the main verb ("potranno": *££modal-can*), but at the current stage of development of the system, modals are ignored, so that the interpretation is actually carried out on the *verb+modal-indcompl* dependent, i.e. the verb "interessare" (*££to-affect-1*). In the first activation of that box, the default top-connection *££dialogue* is provided, and the result is that the shortest path between *££dialogue* and *££to-affect-1* is looked for. The result is:

(1) ££dialogue RANGE-OF &has-dial-topic
DOMAIN ££dialogue-topic
HAS-SUBCLASS ££situation
HAS-SUBCLASS ££status
HAS-SUBCLASS ££to-affect-1

This says that *££to-affect-1*, being a status, i.e. a situation, is one of the possible *££dialogue-topic*. The concepts associated with the management of the dialogue have not been described in the presentation of the ontology, but their use should be rather clear in this example, a use that is the same in all cases.

Then, *build-sing-r-sem* is applied to the restrictions of the verb (i.e. its dependents). The first of them is "addensamento" (*££clouds*)[5], and the second is "settore" (*££it-geogr-area*). Here, the thematic grid of the verb enters into play. It specifies that the verbal subject of "interessare" is the *&affecter* of the *££to-affect-1*

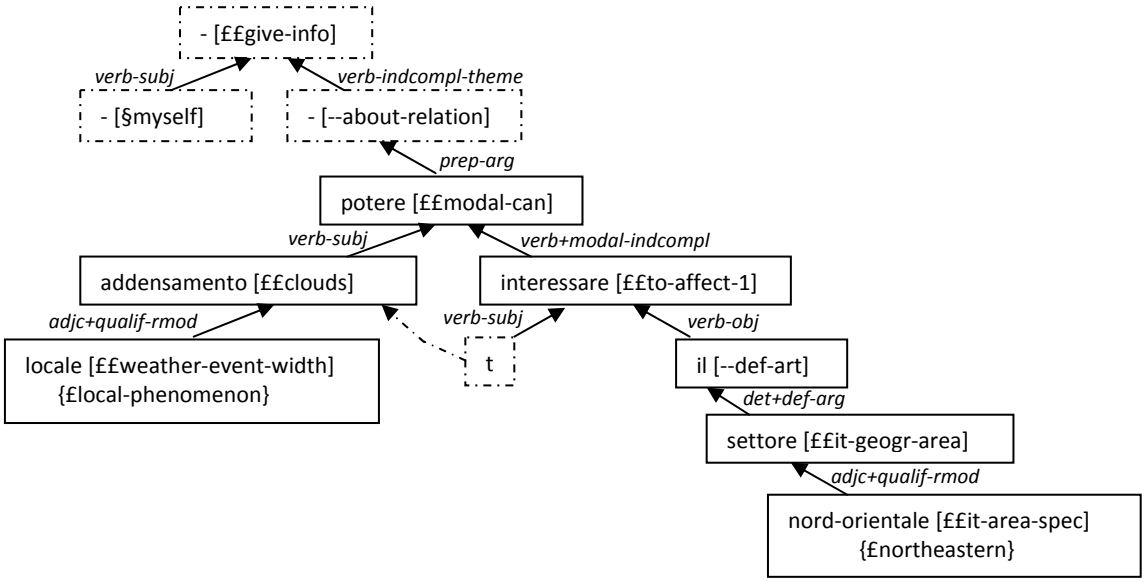

Figure 11 – [Copy of fig.5] The annotated syntactic tree of "Locali addensamenti potranno interessare il settore nord-orientale *(Local cloudiness could concern the north-eastern sector)*

[5] Actually, the subject is the trace, but it includes a reference to the subject of the governing modal, as explained above, so that "addensamento" is retrieved and used as the subject.

reified relation, while the direct object is the *&affectee*. Consequently, the two nested applications of *build-sing-r-sem* involve two invocations of *find-shortest-path*, the first (for connecting *££to-affect-1* to *££clouds*) is constrained to pass through *&affecter*, the other (for connecting *££to-affect-1* to *££it-geogr-area*) is constrained to pass through *&affectee*.

The first one finds (see fig.8):

(1.1) ££to-affect-1 DOMAIN-OF ££affecter
RANGE ££entity
HAS-SUBCLASS ££situation
HAS-SUBCLASS ££event
HAS-SUBCLASS ££weather-event
HAS-SUBCLASS ££clouds

Before looking for the second path, the construction of the first subtree must be completed, so that the interpreter is recursively called on "locali", i.e. the dependent of "addensamento". This produces the path already shown in §4.3, that we repeat here:

(1.1.1) ££clouds SUBCLASS-OF ££weather-event
DOMAIN-OF &has-event-width
RANGE ££weather-event-width
HAS-INSTANCE £local-phenomenon

Since "locali" is the only dependent (restriction) of "addensamento", the two paths are simply composed (DEPTH-COMPOSE, in Box 4), thus obtaining the final representation (marked with *) of "locali addensamenti" (local cloudiness):

(1.1) ££to-affect-1 DOMAIN-OF ££affecter*
RANGE ££entity
HAS-SUBCLASS ££situation
HAS-SUBCLASS ££event
HAS-SUBCLASS ££weather-event
HAS-SUBCLASS ££clouds
SUBCLASS-OF ££weather-event
DOMAIN-OF &has-event-width
RANGE ££weather-event-width
HAS-INSTANCE £local-phenomenon)

Now, the second dependent of the verb: first the shortest path from *££to-affect-1* to *££it-geogr-area*:

(1.2) ££to-affect-1 DOMAIN-OF &affectee
RANGE ££entity
HAS-SUBCLASS ££spatial-location
HAS-SUBCLASS ££geographic-area
HAS-SUBCLASS ££it-geogr-area

Then the recursive call for the interpretation of "nord-orientale" (northeastern), seen in §4.3, repeated here:

(1.2.1) ££it-geogr-area HAS-INSTANCE £it-northeastern-area
ARGUMENT-OF &has-it-area7
RELINSTANCE &has-it-area-spec
RANGE ££it-area-spec
HAS-INSTANCE £northeastern

The composition with (1.2) produces the final representation for the second argument of the verb "interessare":

(1.2) ££to-affect-1 DOMAIN-OF &affectee*
RANGE ££entity
HAS-SUBCLASS ££spatial-location

HAS-SUBCLASS ££geographic-area
HAS-SUBCLASS ££it-geogr-area
HAS-INSTANCE £it-northeastern-area
ARGUMENT-OF &has-it-area7
RELINSTANCE &has-it-area-spec
RANGE ££it-area-spec
HAS-INSTANCE £northeastern

Now, (1.1*) and (1.2*) are "merged" (BREADTH-COMPOSE), and the result is:

(1.1) ££to-affect-1 DOMAIN-OF*
(AND ((££affecter
RANGE ££entity
HAS-SUBCLASS ££situation
HAS-SUBCLASS ££event
HAS-SUBCLASS ££weather-event
HAS-SUBCLASS ££clouds
SUBCLASS-OF ££weather-event
DOMAIN-OF &has-event-width
RANGE ££weather-event-width
HAS-INSTANCE £local-phenomenon)
(&affectee
RANGE ££entity
HAS-SUBCLASS ££spatial-location
HAS-SUBCLASS ££geographic-area
HAS-SUBCLASS ££it-geogr-area
HAS-INSTANCE £it-northeastern-area
ARGUMENT-OF &has-it-area7
RELINSTANCE &has-it-area-spec
RANGE ££it-area-spec
HAS-INSTANCE £northeastern

which, after being DEPTH-COMPOSEd with (1), produces the final result (see Box 1).

We do not repeat step by step the interpretation of the second example, but we report in Box 5 the final result of this interpretation step, i.e. the ontological representation. In this representation, we have outlined in boldface the concepts that constitute the connection between the annotation of the parse tree and the ontological paths. Note also the presence *of –deictic-specif-other*, which is associated with "altri" in "altri mari". Also in this case, some special processing is required.

5 CONCLUSIONS

In this report, we have described the organization of the semantic interpreter developed by the Unit at the Department of Informatica of the University of Torino. The interpreter is currently devoted to the Meteo sentences of the ATLAS project, but its organization is general enough to make it applicable to other domains with limited effort, apart from the need to develop a suitable domain ontology. Also, is not related to any target language, since the semantic representation obtained by the interpreter is assumed to be largely language-independent. The answer to the question "how largely" will be left to the subsequent research activity.

A possible objection to the approach described in this Deliverable concerns the apparent complexity of the resulting ontological representation (see also Footnote 1). We partially agree, but we must repeat that the representation undergoes a simplification step in the passage to the FOL representation that eliminates a lot of redundancy (see D4.3#4).

```
….. ££dialogue-topic HAS-SUBCLASS
      (AND ((£££to-have-meteo-status
            SUBCLASS-OF £££to-have
               HAS-SUBCLASS ££to-have-meteo-status
                  DOMAIN-OF &has-meteo-status
                     RANGE ££meteo-status-situation
                        (AND ((HAS-SUBCLASS ££sea-status-situation
                                 HAS-SUBCLASS £££sea-current
                                    DOMAIN-OF &has-current-origin
                                     RANGE £££cardinal-origin
                                       HAS-INSTANCE £northern)
                              (DOMAIN-OF &has-meteo-time
                                RANGE ££time-interval
                                  DOMAIN-OF &has-time-interv-descr
                                    RANGE ££time-interval-description
                                      HAS-SUBCLASS £££anaphoric-time-interval-description
                                        HAS-INSTANCE –time-anaphoric-continuation))))
         (££situation
            HAS-SUBCLASS ££meteo-status-situation
               HAS-SUBCLASS ££sea-status-situation
                  DOMAIN-OF  &has-sea-status-sea
                     RANGE ££sea
                        HAS-SUBCLASS £££it-sea
                           (AND ((HAS-INSTANCE ionio)
                                 (SUBCLASS-OF ££sea
                                   RANGE-OF &has-sea-status-sea
                                     DOMAIN  ££sea-status-situation
                                       DOMAIN-OF &has-sea-status
                                         RANGE £££sea-status-description
                                           HAS-INSTANCE £mare-agitato
                                             VALUE-OF &moved-sea-intens-1
                                               ARGUMENT £intens-rel-1
                                                 (AND ((ARGUMENT-OF &moved-sea-intens-3
                                                        VALUE –intensifier-adv)
                                                       (ARGUMENT-OF &moved-sea-intens-2
                                                        VALUE £very-rough-sea)))))))
         (££situation
            HAS-SUBCLASS ££meteo-status-situation
               HAS-SUBCLASS ££sea-status-situation
                  DOMAIN-OF  &has-sea-status-sea
                     RANGE £££sea
                        (AND ((HAS-INSTANCE  –deictic-specif-other)
                              (RANGE-OF &has-sea-status-sea
                                DOMAIN ££sea-status-situation
                                  DOMAIN-OF &has-sea-status
                                    RANGE ££sea-status
                                      DOMAIN-OF &has-sea-status-descr
                                        RANGE  £££sea-status-description
                                          HAS-INSTANCE £rough-sea))))))
```

. Box 5 – Ontological form of "Locali addensamenti potranno interessare il settore nord-orientale"
(Local cloudiness could concern the north-eastern sector).

One of the main extensions needed to extend the coverage regards the proper treatment of quantification. This is currently limited to some existential quantifiers that are included in the representation in the next step (see D4.3#4), but it is clear that some effort is required here. Also, extensive tests need be done on other weather forecast news in order to check that all relevant phenomena are correctly caught.

All features described in the present report have been implemented in the current version of the interpreter, which has been implemented in LISP and is currently available (D4.2, software) to all partners of the project. For all details about the other components of the translator, the reader is addressed to the related parts of this report, i.e D4.3#1 (Organization), D4.3#2, (Ontology), D4.3#4 (From ontological forms to FOL), D4.3#5 (Linguistic generation), D4.3#6 (Planning for hand positioning).