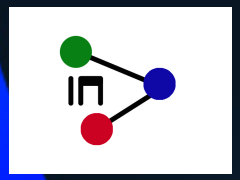
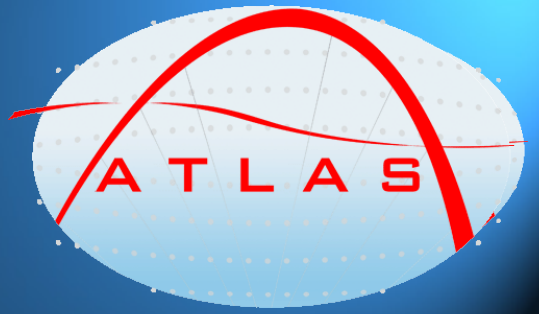




LINGUISTIC PROCESSING IN THE ATLAS PROJECT

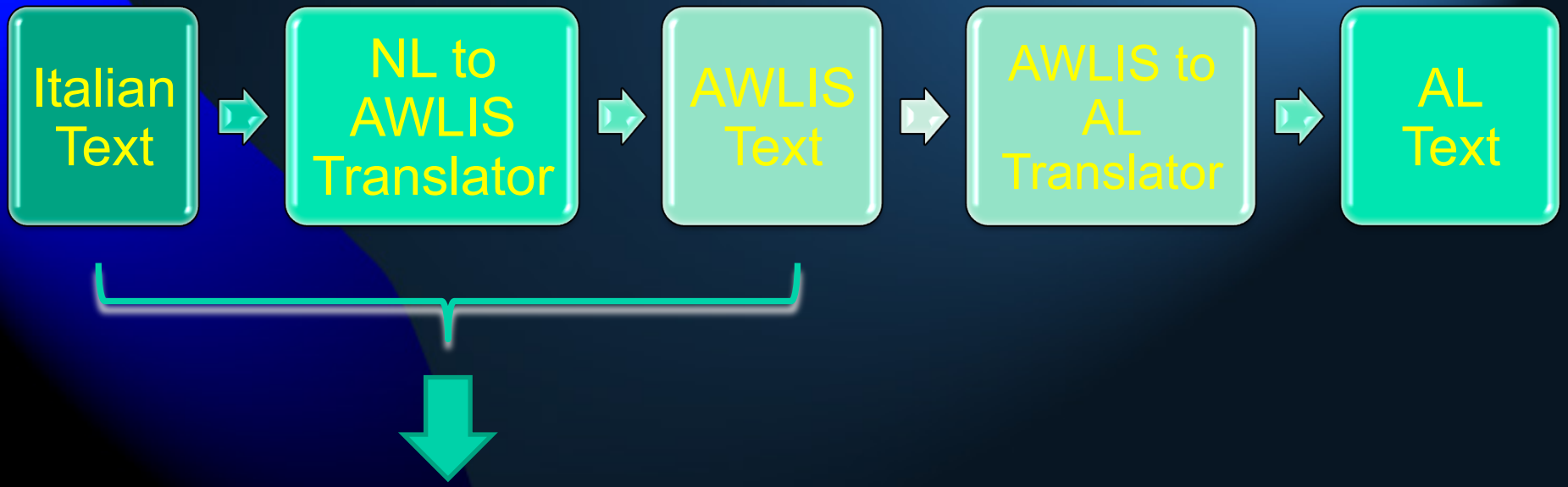
Leonardo LESMO, Alessandro Mazzei, Daniele Radicioni
Interaction Models Group
Dipartimento di Informatica e Centro di Scienze Cognitive
Università di Torino
{lesmo,mazzei,radicion}@di.unito.it

SLTAT 2011 - Berlin, January 10-11, 2011

ATLAS : Scientific challenges

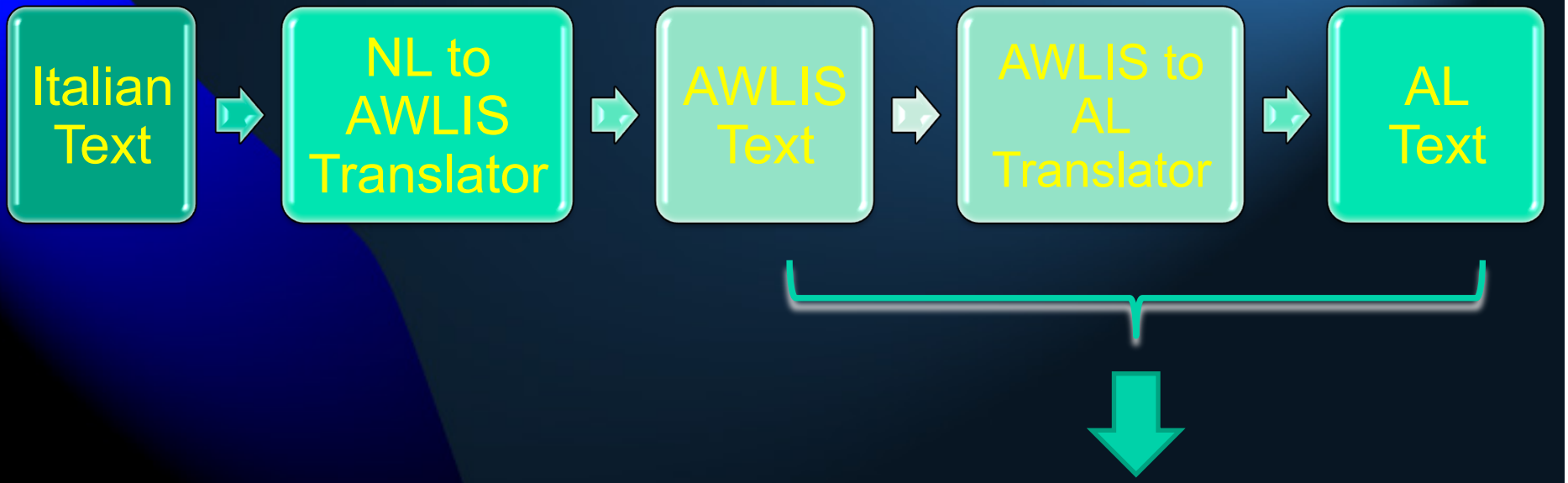
- Written representation of LIS (Linguaggio Italiano dei Segni: Italian Sign Language)
- Automatic translation from Italian to LIS
- Creation of linguistic corpora
- Definition of a virtual character with proper expressions
- Transmission/visualization on several media
- ,,,

ATLAS Translation Pipeline



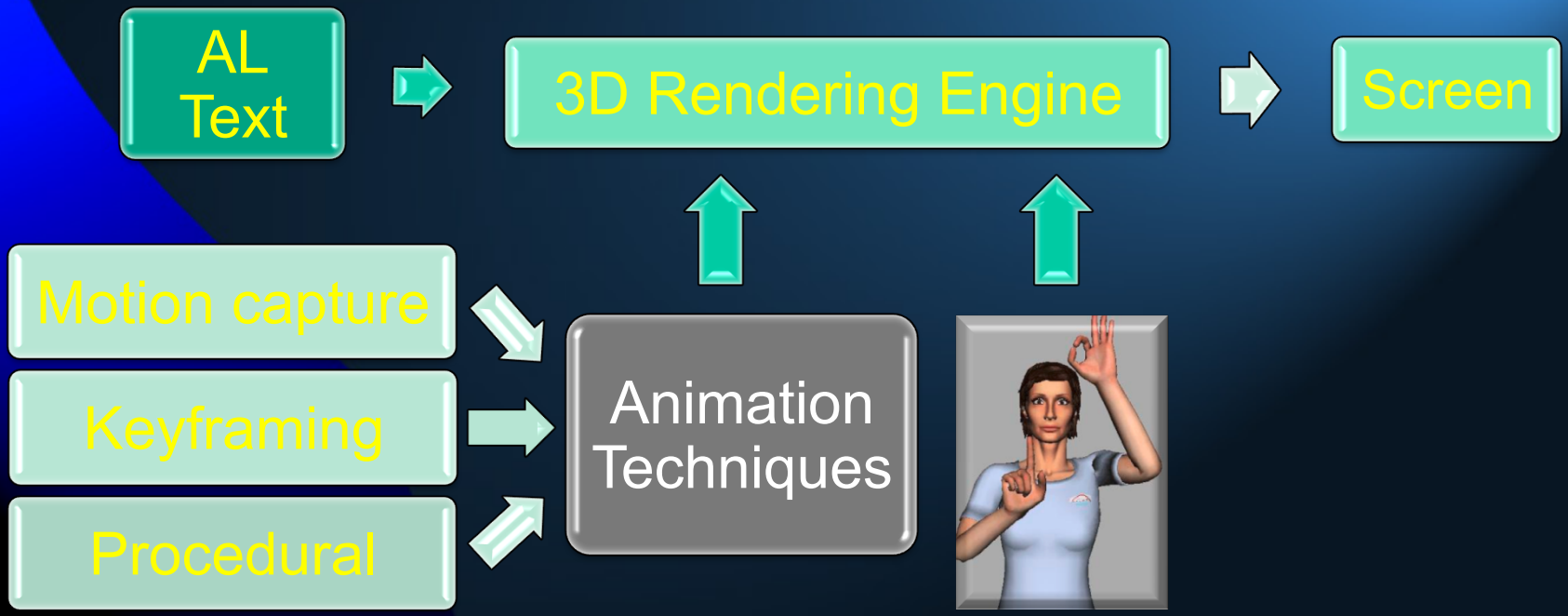
Input texts written in Italian will be translated into an intermediate representation, called ATLAS Written Italian Sign Language (AWLIS)

ATLAS Translation Pipeline



Sentences expressed in AWLIS are then translated into a character's gestures Animation Language (AL) which describes the way the basic movements are produced and linked.

ATLAS Translation Pipeline



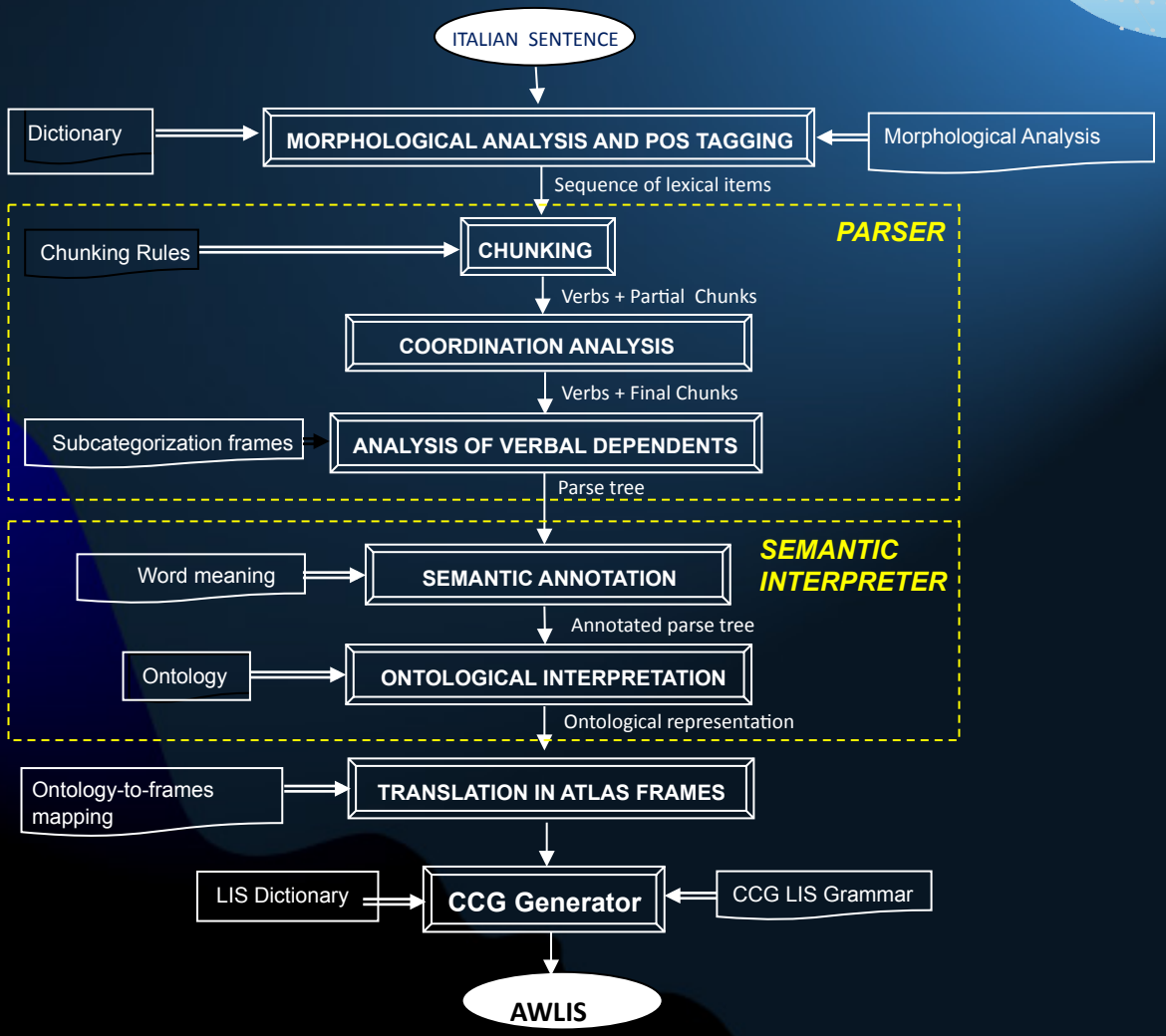
AL sequences will be rendered through a 3D rendering engine.
An automatic “animation blending” system will ensure smooth transitions between signs of the generated LIS animations

SL Translation Issues

Translating from a spoken language to a SL is a complex undertaking:

- The translation must transform a symbolic representation into another symbolic representation
- But SLs incorporate gestures, facial expressions, head movements, body language and even the space around the speaker
- So, we need a way to express in symbolic form the LIS sentence, i.e. a written LIS

Linguistic Analysis



Example

- **Locali addensamenti potranno interessare il settore nord-orientale (local clouding could affect the North-eastern sector)**

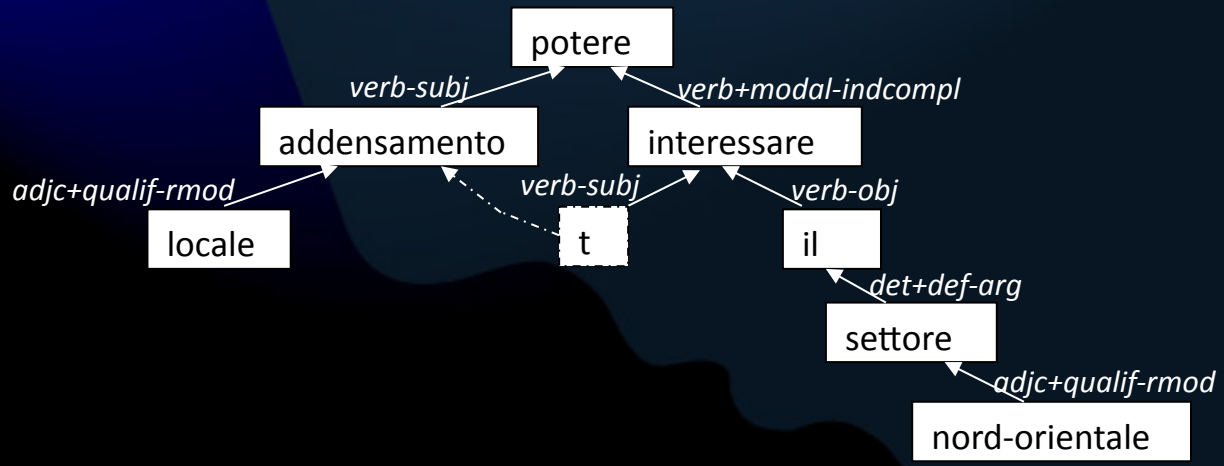
Result of the Morphologic-lexical analysis+ POS Tagging):

1 Locali (LOCALE ADJ QUALIF ALLVAL PL)
2 addensamenti (ADDENSAMENTO NOUN COMMON M PL ADDENSARE TRANS)
3 potranno (POTERE VERB MOD IND FUT INTRANS 3 PL)
4 interessare (INTERESSARE VERB MAIN INFINITE PRES TRANS)
5 il (IL ART DEF M SING)
6 settore (SETTORE NOUN COMMON M SING)
7 nord-orientale (NORD-ORIENTALE ADJ QUALIF ALLVAL SING)
8 . (#\. PUNCT)

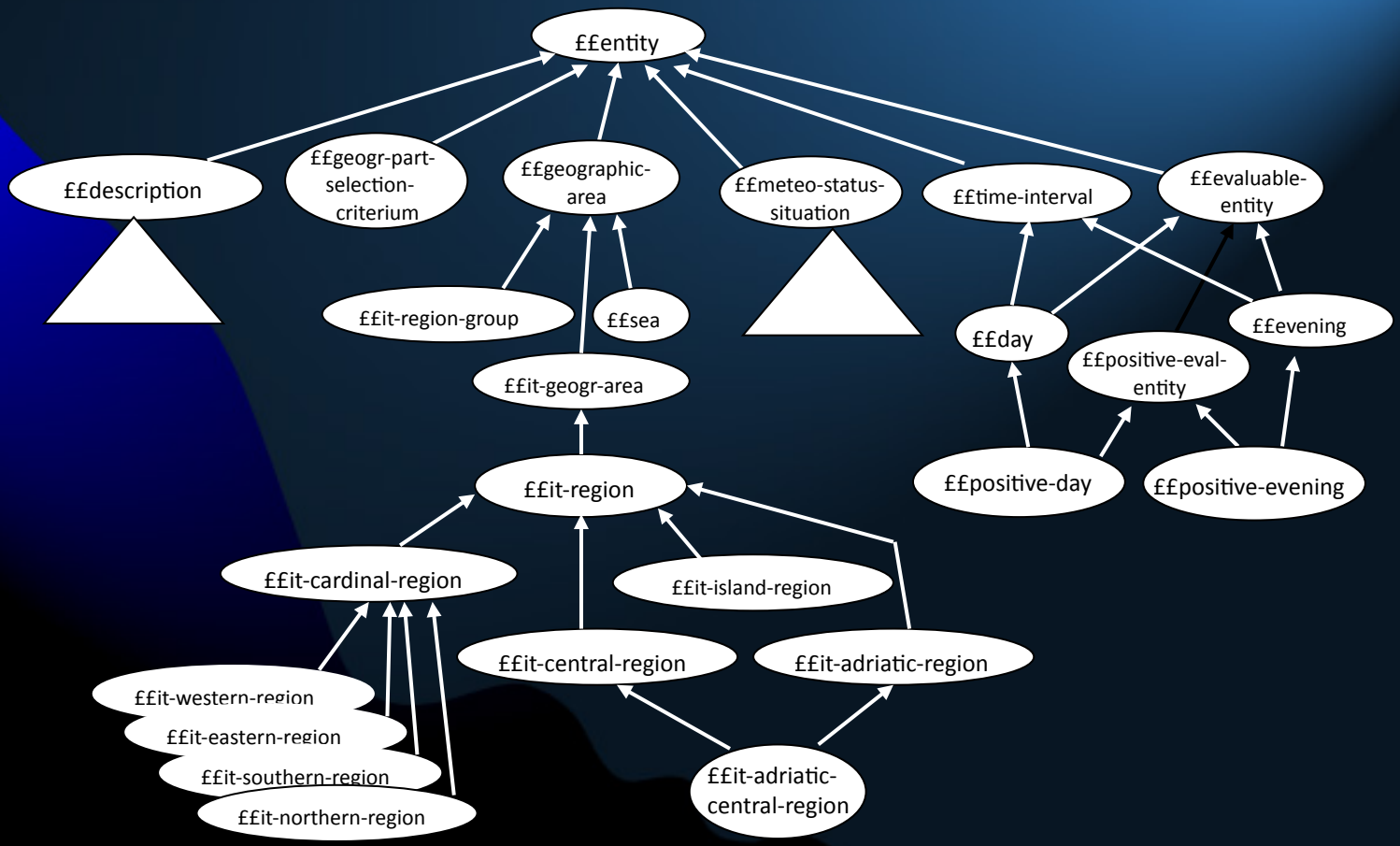
Example (2)

Result of syntactic analysis

1 Locali (LOCALE ADJ QUALIF ALLVAL PL) **[2;ADJC+QUALIF-RMOD]**
2 addensamenti (ADDENSAMENTO NOUN COMMON M PL ADDENSARE TRANS) **[3;VERB-SUBJ]**
3 potranno (POTERE VERB MOD IND FUT INTRANS 3 PL) **[0;TOP-VERB]**
4 interessare (INTERESSARE VERB MAIN INFINITE PRES TRANS) **[3;VERB+MODAL-INDCOMPL]**
4.10 t [2f] (ADDENSAMENTO NOUN COMMON M PL ADDENSARE TRANS) **[4;VERB-SUBJ]**
5 il (IL ART DEF M SING) **[4;VERB-OBJ]**
6 settore (SETTORE NOUN COMMON M SING) **[5;DET+DEF-ARG]**
7 nord-orientale (NORD-ORIENTALE ADJ QUALIF ALLVAL SING) **[6;ADJC+QUALIF-RMOD]**
8 . (#\. PUNCT) **[3;END]**



Semantics: a fragment of the ontology



Example (3)

Semantics: an ontological expression

££meteo-status-situation has-subclass
(and
(££weather-event has-subclass
££clouds subclass-of ££weather-event
domain-of &has-event-width
range ££weather-event-width (eq £local-phenomenon)
(££weather-status-situation domain-of
&has-weather-location range ££geographic-area has-subclass
££it-geogr-area has-instance £it-northeastern-area arg-of
&has-it-area7 relinstance &has-it-area-spec range ££it-area-spec
(eq £northeastern)))

Example (4)

Semantics: First Order Predicate Logic

∀ e (possible (e) ∀ ££clouds (e) ∀
&has-event-width (e, £local-phenomenon) ∀
&has-weather-location (e, £it-northeastern-area))

Notes

1) *Modal logics: ?; currently reification*
2) *The sequence*
   "£it-northeastern-area arg-of &has-it-area7 relinstance &has-it-area-spec range ££it-area-spec (eq £northeastern)"
   specifies that, in this context, "northeastern" is a "linguistic" description (specification) of the involved area

Generation

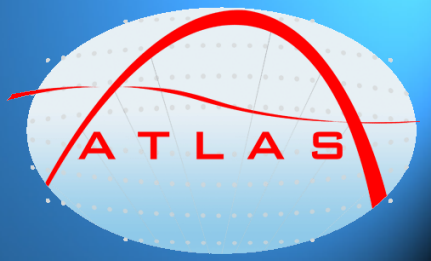
1. Starting from an ATLAS frame we compute a "semantic tree" constituting an intermediate semantic representation, by using IF-THEN rules
2. Starting from the semantic tree we produce the sequence of LIS glosses, by using a CCG for LIS

Example (4)

- *Per domani quindi nubi in aumento al nord.*

- $\exists$ (ev_1 t_1 l_1 x_1) (**cloud-increase**(ev1) $\wedge$ **time**(ev_1, t_1) $\wedge$ **deictic-descr**(t_1, t_2) $\wedge$ **id**(t_2,tomorrow) $\wedge$ **location**(ev_1, l_1) $\wedge$ **id**(l_1, nord) $\wedge$ **agent**(ev_1, x_1) $\wedge$ **cloud**(x_1)) --→

```
<?xml version="1.0" encoding="UTF-8"?>
<xml>
  <lf>
    <satop nom="c1:meteo-status-situation">
      <prop name="cloud-increase"/>
      <diamond mode="AGENT">
        <nom name="c2:clouds"/>
        <prop name="cloud"/>
      </diamond>
      <diamond mode="LOCATION">
        <nom name="n1:it-northern-region"/>
        <prop name="north"/>
      </diamond>
      <diamond mode="TIME">
        <nom name="t1:deictic-day-description"/>
        <prop name="tomorrow"/>
      </diamond>
    </satop>
  </lf>
</xml>
```

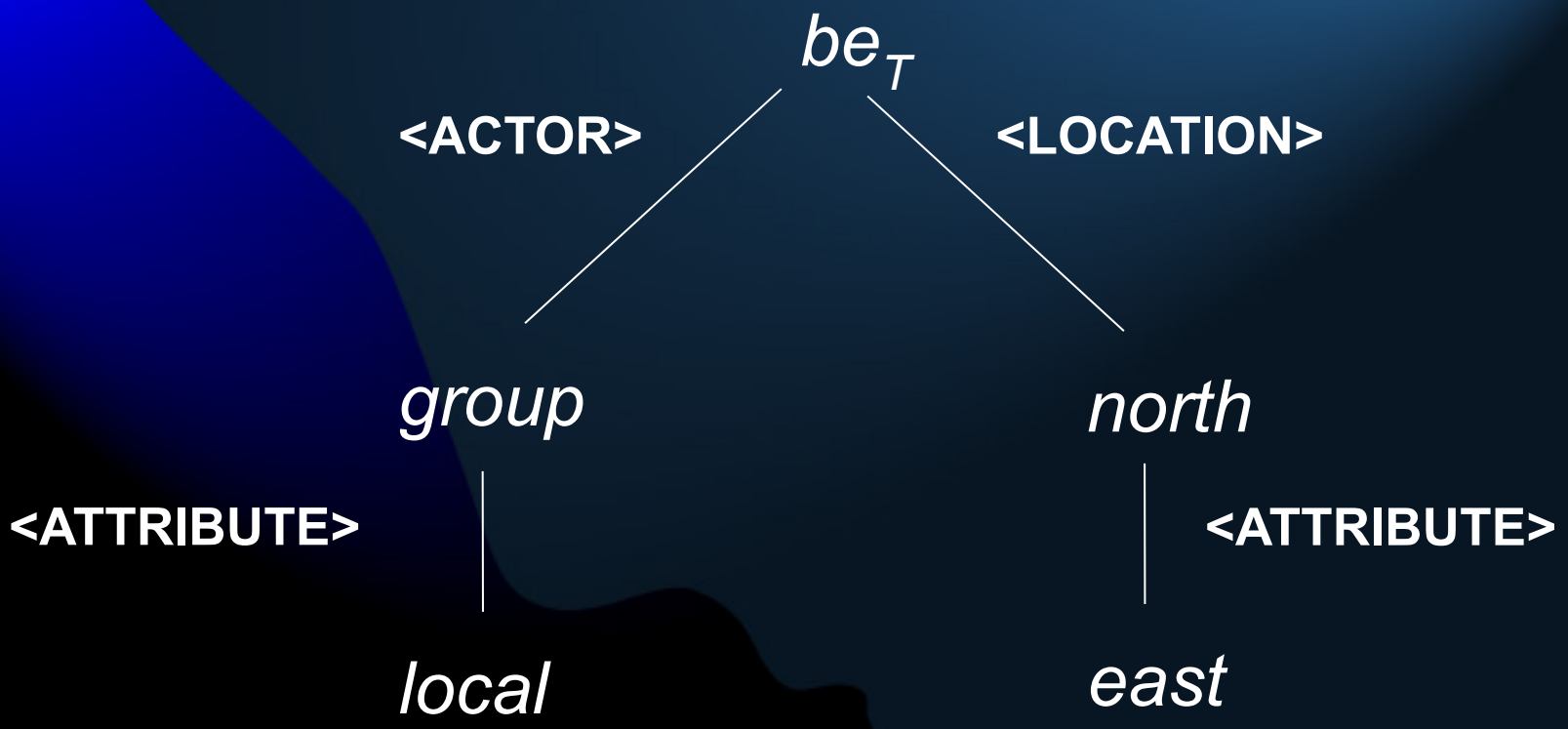



```
<?xml version="1.0" encoding="UTF-8"?>
<xml>
 <lf>
   <satop nom="c1:meteo-status-situation">
     <prop name="cloud-increase"/>
     <diamond mode="AGENT">
       <nom name="c2:clouds"/>
       <prop name="cloud"/>
     </diamond>
     <diamond mode="LOCATION">
       <nom name="n1:it-northern-region"/>
       <prop name="north"/>
     </diamond>
     <diamond mode="TIME">
       <nom name="t1:deictic-day-description"/>
       <prop name="tomorrow"/>
     </diamond>
   </satop>
 </lf>
</xml>
```

- **domani_N nord_U nuvola_U nuvola-aumentare_U**
- **...**

Example (5)

Generation: Semantic Tree



Example (6)

Generation: AWLIS

INVECE NORDORIENTALE L ZONA L NUVOLE

invece
both
std
...
...

nordorientale
sx
std
...
...

zona
dx
std
...
...

nuvole
dx
std
...
...

A note on the dictionary: The ATLAS Sign set

The set of signs used within the ATLAS project includes:

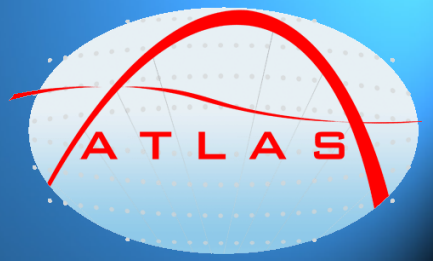
- Signs associated with “Lemmas” included in the Radutzky dictionary
- Signs widely used within the various Italian deaf communities, but not included in the Radutzky dictionary
- Signs newly defined within the ATLAS project, previously not existing, and needed to represent peculiar concepts, mainly linked with meteorological events and situations

From AWLIS to AEWLIS: Information Tracks

A sentence is represented, in AEWLIS, resorting to a set on concurrent synchronized "*Tracks*".

The following 4 Tracks (TRs) are used:

- ***Lemma***:
  Used to store information items related to the Lemmas, including, among the others, the lemma's ID, *Identifier,* its *Part of Speech,* and the set of *Modifiers* for the Lemma.



Information Tracks (2)

- ***Sentence Structure:***

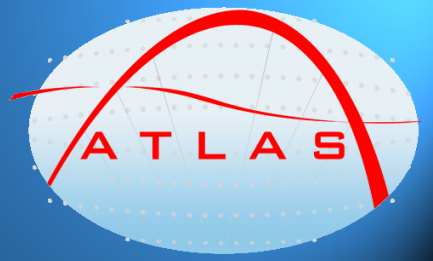
  Used to store information items related to the *Structure of the Sentence*, including, among the others, proper pointer to the related syntax tree, the sentence's syntactic and semantic roles, and its speech acts.

Information Tracks (3)

- ***Channel Modifiers:***

  Its usage is twofold:

  - To record the actual values of the following sign parameters:
    - *Sign Spatial Location*, for “Relocatable” signs
    - *Action performed* by the not-signing hand, when the sign requires just one hand.
  - To record the differences on the various *Communication Channels* between the actual values of the performed sign and the default ones stored in the ATLAS MultiMedia Archive (AMMA)

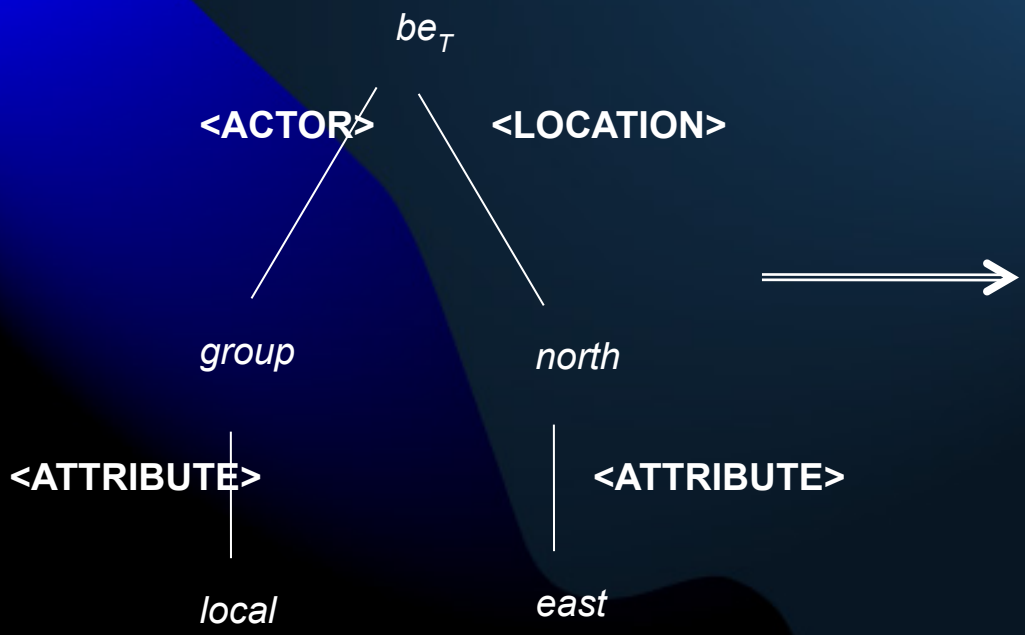


Information Tracks (4)

- ***Time Stamps***:

  Used to store info related to the Time Stamps associated with the sentence *Time Slices*

Example (7)



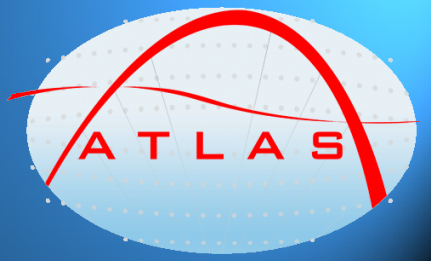
(lemma de00 nuvola)
(lemma de01 zona)
(lemma de02 nordorientale)
(lemma p00 in)
(isTopic de01)
(into de00 de01)
(rel de00 de01)
(right ndom)
(left dom)
(free dom)
(free ndom)
(is_relocatable de00)
(is_relocatable de01)
(not_relocatable de02)
(default_position de00 0 0 0)
(default_position de01 0 0 0)
(default_position de02 0 0 0)
(position de01 1 1 0)

Example (6)

invece nordorientale zona nuvole

(lemma de00 nuvola)
(lemma de01 zona)
(lemma de02 nordorientale)
(lemma p00 in)
(isTopic de01)
(into de00 de01)
(rel de00 de01)
(right ndom)
(left dom)
(free dom)
(free ndom)
(is_relocatable de00)
(is_relocatable de01)
(not_relocatable de02)
(default_position de00 0 0 0)
(default_position de01 0 0 0)
(default_position de02 0 0 0)
(position de01 1 1 0)

(!!handlocation both 0.0 0.0 0.0)
(!!handanimation both invece)
(!!handlocation both 0.0 0.0 0.0)
(!!handanimation both nordorientale)
(!!handlocation dom 1.0 1.0 0.0)
(!!handanimation dom zona)
(!!handlocation both 1.0 1.0 0.0)
(!!handanimation both nuvola)
(!removeentity (entity de00))



CONCLUSIONS

- Linguistic Analysis is a hard task
- It involves both grammatical knowledge and knowledge of the world (ontology)
- Though the problem is not solvable today in open-ended domains, good results can be obtained for restricted domains (as Weather Forecast)
- The adopted solutions can be extended to wider domains provided that the needed semantic knowledge is made available