Semantic and Pragmatic Computing with GETARUNS

Rodolfo Delmonte

University of Venice "Ca' Foscari" (Italy)

email: `delmont@unive.it`

Abstract

We present a system for text understanding called GETARUNS, in its deep version applicable only to Closed Domains. We will present the low level component organized according to LFG theory. The system also does pronominal binding, quantifier raising and temporal interpretation. Then we will introduce the high level component where the Discourse Model is created from a text. Texts belonging to closed domains are characterized by the fact that their semantics is controlled or under command of the system; and most importantly, sentences making up the texts are fully parsed without failures. In practice, these texts are short and sentences are also below a certain threshold, typically less than 25 words. For longer sentences the system switches from the topdown to the bottomup system. In case of failure it will backoff to the partial system which produces a very lean and shallow semantics with no inference rules. The small text we will present contains what is called a "psychological statement" sentence which contributes an important bias as to the linking of the free pronominal expression contained in the last sentence.

1 The System GETARUNS

GETARUNS, the system for text understanding developed at the University of Venice, is equipped with three main modules: a lower module for parsing where sentence strategies are implemented; a middle module for semantic interpretation and discourse model construction which is cast into Situation Semantics; and a higher module where reasoning and generation takes place.

The system is based on LFG theoretical framework (Bresnan, 2001) and has a highly interconnected modular structure. The Closed Domain version of the system is a top-down depth-first DCG-based parser written in Prolog Horn Clauses, which uses a strong deterministic policy by means of a lookahead mechanism with a WFST to help recovery when failure is unavoidable due to strong attachment ambiguity.

It is divided up into a pipeline of sequential but independent modules which realize the subdivision of a parsing scheme as proposed in LFG theory where a c-structure is built before the f-structure can be projected by unification into a DAG (Direct Acyclic Graph). In this sense we try to apply in a given sequence phrase-structure rules as they are ordered in the grammar: whenever a syntactic constituent is successfully built, it is checked for semantic consistency. In case the governing predicate expects obligatory arguments to be lexically realized they will be searched and checked for uniqueness and coherence as LFG grammaticality principles require.

Syntactic and semantic information is accessed and used as soon as possible: in particular, both categorial and subcategorization information attached to predicates in the lexicon is extracted as soon as the main predicate is processed, be it adjective, noun or verb, and is used to subsequently restrict the number of possible structures to be built. Adjuncts are computed by semantic compatibility tests on the basis of selectional restrictions of main predicates and adjuncts heads.

The output of grammatical modules is fed then onto the Binding Module (BM) which activates an algorithm for anaphoric binding. Antecedents for pronouns are ranked according to grammatical function, semantic role, inherent features and their position at f-structure. Eventually, this information is added into the original f-structure graph and then passed on to the Discourse Module (DM).

The grammar is equipped with a core lexicon containing most frequent 5,000 fully specified inflected word forms where each entry is followed by its lemma and a list of morphological features, organised in the form of attribute-value pairs. However, morphological analysers for English are also available with big root dictionaries (25,000 for English) which only provide for syntactic subcategorization, though. In addition to that there are all lexical form provided by a fully revised version of COMLEX, and in order to take into account phrasal and adverbial verbal compound forms, we also use lexical entries made available by UPenn and TAG encoding. Their grammatical verbal syntactic codes have then been adapted to our formalism and are used to generate a subcategorization schemes with an aspectual and semantic class associated to it — however no restrictions can reasonably be formulated on arguments of predicates. Semantic inherent features for Out of Vocabulary Words, be they nouns, verbs, adjectives or adverbs, are provided by a fully revised version of WordNet (Fellbaum, 1998) — plus EuroWordnet, with a number of additions coming from computer, economics, and advertising semantic fields — in which we used 75 semantic classes similar to those provided by CoreLex (Buitelaar, 1998).

When each sentence is parsed, tense aspect and temporal adjuncts are accessed to build the basic temporal interpretation to be used by the temporal reasoner. Eventually two important modules are fired: Quantifier Raising and Pronominal Binding. QR is computed on f-structure which is represented internally as a DAG. It may introduce a pair of functional components: an operator where the quantifier can be raised, and a pool containing the associated variable where the quantifier is actually placed in the f-structure representation. This information may then be used by the following higher system to inspect quantifier scope. Pronominal binding is carried out at first at sentence internal level. DAGs will be searched for binding domains and antecedents matched to the pronouns if any to produce a list of possible bindings. Best candidates will then be chosen.

2   The Upper Module

GETARUNS has a highly sophisticated linguistically based semantic module which is used to build up the Discourse Model. Semantic processing is strongly modularized and distributed amongst a number of different submodules which take care of Spatio-Temporal Reasoning, Discourse Level Anaphora Resolution, and other subsidiary processes like Topic Hierarchy which cooperate to find the most probable antecedent of coreferring and cospecifying referential expressions when creating semantic individuals. These are then asserted in the Discourse Model (hence the DM), which is then the sole knowledge representation used to solve nominal coreference.

The system uses two resolution submodules which work in sequence: they constitute independent modules and allow no backtracking. The first one is fired whenever a free sentence external pronoun is spotted; the second one takes the results of the first submodule and checks for nominal anaphora. They have access to all data structures contemporarily and pass the resolved pair, anaphor-antecedent to the following modules.

Semantic Mapping is performed in two steps: at first a Logical Form is produced which is a structural mapping from DAGs onto unscoped well-formed formulas. These are then turned into situational semantics informational units, infons which may become facts or sits. Each unit has a relation, a list of arguments which in our case receive their semantic roles from lower processing — a polarity, a temporal and a spatial location index.

3   The Text

The text we present for the shared task (Bos, 2008) is a "psychological statement" text, i.e. it includes a sentence (namely sentence 4) that represents a psychological statement, i.e. it expresses the feelings and is viewed from the point of view of one of the participants in the story. The relevance of the sentence is its role in the assignment of the antecedent to the pronominal expressions contained in the following sentence. Without such a sentence the anaphora resolution module would have no way of computing "John" as the legitimate antecedent of "He/his". On the contrary, in a system like ours that computes Point of View and Discourse Domain on the basis of Informational Structure and Centering information, it will be possible to make available the appropriate antecedent to the anaphora resolution module.

We will discuss mainly semantic information processing. In so doing we shall have to devote some space to LFG grammatical representation, to Logical Form and eventually the Discourse Model. However, since this is meant to be a short paper, we will only be able to show some fragments of the overall representation, highlighting the most important features and disregarding the rest. So first of all, consider the sentences making up the text:

> **1. John went into a restaurant.**
> **2. There was a table in the corner.**
> **3. The waiter took the order.**
> **4. The atmosphere was warm and friendly.**
> **5. He began to read his book.**

We will be able to present an almost complete sequence of representations as produced by GETARUNS only for one sentence, and then we will comment on the rest.

```
1. John went into a restaurant

index:f1
pred:go
lex_form:[np/subj/agente/[human, object],
        sp/obl/locat/[to, in, into]/[object, place]]
voice:active; mood:ind; tense:pres; cat:risultato
subj/agent:index:sn4
        cat:[human]; pred:'John'
        gen:mas; num:sing; pers:3; spec:def:'0'
        tab_ref:[+ref, -pro, -ana, -class]
obl/locat:index:sn5
        cat:[place]; pred:restaurant
        num:sing; pers:3; spec:def:-
        tab_ref:[+ref, -pro, -ana, +class]; qmark:q1
aspect:achiev_tr
rel1:[td(f1_res2)=tr(f1_res2)]
rel2:[included(tr(f1_res2), tes(f1_res2))]
specificity:-; ref_int:[tr(f1_res2)]
qops:qop:q(q1, indefinite)
```

'Centering and Topic Hierarchy'
state(1, retaining)   topic(1, main, id5)   topic(1, potential, id1)

```
INFORMATIONAL STRUCTURE

CLAUSE IDENTIFIER:         2-n1
CLAUSE TYPE:               main/prop
FACTUALITY:                factive
CHANGE IN THE WORLD:       null
RELEVANCE:                 background
TEMP_RELATION:             undef(tes(f1_res2), nil)
DISCOURSE FOCUS:           tes(f1_res2)
DISCOURSE RELATION:        narration
DISCOURSE DOMAIN:          objective
POINT OF VIEW:             narrator
```

```
LOGICAL FORM
wff(situation,
    wff(go,
        < entity : sn4 : wff(isa, sn4,  John) >,
        < indefinite : sn5 : wff(isa, sn5, restaurant) >,
        < event : f1 :
            wff(and, wff(isa,  f1,  ev),
                     wff(time, f1, < definite : t2 :
                          wff(and, wff(isa, t2, tloc),
                                   wff(pres, t2)) >)) >))
```

DISCOURSE MODEL 2

```
/*** There was a table in the corner. ***/
loc(infon13, id4, [arg:main_sloc, arg:restaurant])
ind(infon14, id5)
fact(infon15, inst_of, [ind:id5, class:man], 1, univ, univ)
fact(infon16, name, ['John', id5], 1, univ, univ)
fact(id6, go, [agente:id5, locat:id1], 1, tes(f1_res2), id4)
fact(infon19, isa, [arg:id6, arg:ev], 1, tes(f1_res2), id4)
fact(infon20, isa, [arg:id7, arg:tloc], 1, tes(f1_res2), id4)
fact(infon21, pres, [arg:id7], 1, tes(f1_res2), id4)
fact(infon22, time, [arg:id6, arg:id7], 1, tes(f1_res2), id4)
includes(tr(f1_res2), univ)
```

Sentence 2, is a presentational structure, where the subject form "there" is recovered as being part of the meaning of the main predicate in the semantics. The location "in the corner" is computed as a adjunct and it is understood as a entertaining a meronimic relation with the main location, "the restaurant", again in the semantics. When building the Discourse Model it is possible to fire inferences to recover pragmatic unexpressed implicatures, as for instance, the fact that introducing a "table" with a presentational structure and an indefinite NP but accompanied by a definite location induces the reader to produce such implicit information as indicated below, i.e, the fact that the main topic and only current participant to the discourse is supposed to be sitting at the table in the corner. This inference is fired by inferential rules that look for relations intevening between main location and current location; also presentational structure contributes by introducing an indefinite "table" which is the trigger of the SITTING event.

DISCOURSE MODEL 3

```
/*** The waiter took the order. ***/
loc(infon26, id8, [arg:main_tloc, arg:tes(f1_res2)])
ent(infon27, id9)
fact(infon28, inst_of, [ind:id9, class:place], 1, univ, univ)
fact(infon29, isa, [ind:id9, class:table], 1, id8, id4)
in(infon30, id9, id4)
fact(id10, sit, [actor:id5, locat:id9], 1, tes(f5_id10), id4)
fact(infon31, isa, [arg:id10, arg:ev], 1, tes(f5_id10), id4)
fact(infon32, isa, [arg:id11, arg:tloc], 1, tes(f5_id10), id4)
fact(infon33, isa, [arg:id11], 1, tes(f5_id10), id4)
ind(infon34, id12)
fact(infon35, inst_of, [ind:id12, class:place], 1, univ, univ)
```

```
fact(infon36, isa, [ind:id12, class:corner], 1, id8, id4)
fact(infon37, part_of, [restaurant, id12, id1], 1, id8, id4)
fact(id13, there_be, [prop:id9], 1, tes(f4_res3), id4)
```

This sentence is computed as containing an idiomatic predicate "take_order" which in turn has a BENEFICIARY/GOAL of the same event. In turn the Goal is computed as if it were an obligatory semantic role like the missing Agent of passivized structures. The semantics is then responsible for checking consistency of predicate-argument structures. The Goal induces the presence of an Oblique which is filled with an "exist" dummy predicate. This predicate is then linked to the only other available participant in the topic structure organized by the Centering Algorithm, John with semantic Id = id5.

```
index:f1
pred:take
lex_form:[np/subj/agent/[human], idioms/obj/form/[order],
          pp/obl/goal/from/[human]]
voice:active;  mood:ind; tense:pres; cat:activity
subj/agent:index:sn3
          cat:[human, social]; pred:waiter
          gen:mas; num:sing;  pers:3; spec:def:+
          tab_ref:[+ref, -pro, -ana, +class]
ogg/form:index:sn4
        cat:[activity, event]; pred:order
        num:sing; pers:3; spec:def:+
        tab_ref:[+ref, -pro, -ana, +class]
obj2/goal:index:sn5
          cat:[human, animate]; pred:exist
          spec:def:-; part:+
          tab_ref:[+ref, -pro, -ana, +me]
aspect:activity
rel1:[td(f1_res4)=tr(f1_res4)]
rel2:[included(tr(f1_res4), tes(f1_res4))]
specificity:+; ref_int:[tr(f1_res4)]
```

'Centering and Topic Hierarchy'
state(4, continue)   topic(4, main, id5)   topic(4, potential, id16)

DISCOURSE MODEL 4

```
/*** The atmosphere was warm and friendly. ***/
loc(infon49, id15, [arg:main_tloc, arg:tes(f4_res3)])
ind(infon50, id16)
fact(infon51, inst_of, [ind:id16, class:social_role], 1, univ, univ)
fact(infon52, isa, [ind:id16, class:waiter], 1, id15, id4)
fact(infon53, role, [waiter, id4, id16], 1, id15, id4)
fact(infon55, isa, [arg:id5, arg:exist], 1, id15, id4)
fact(id18, take_order, [agent:id16, goal:id5], 1, tes(f1_res4), id4)
```

Sentence 4, is the psychological statement, where the Centering Algorithm uses the information made available by the computational called Informational Structure that we report here below.

'Centering and Topic Hierarchy'
state(4, continue)   topic(4, main, id5)   topic(4, potential, id21)

```
INFORMATIONAL STRUCTURE

CLAUSE IDENTIFIER:          5-n1
CLAUSE TYPE:                main/prop
FACTUALITY:                 factive
CHANGE IN THE WORLD:        null
RELEVANCE:                  background
TEMP_RELATION:              during(tes(f1_res5), tes(f1_res4))
DISCOURSE FOCUS:            tes(f1_res5)
DISCOURSE RELATION:         explanation
DISCOURSE DOMAIN:           subjective
POINT OF VIEW:              John
```

As can be noticed, the system has computed the Discourse Domain as "subjective", and the Point of View as belonging to one of the participants, the one referred by with a proper name. In fact, it is just the use of a definite expression "the waiter" that tells the system to underrate the importance in the Topic Hierarchy automatically built by the Centering Algorithm.

DISCOURSE MODEL

```
loc(infon77, id24, [arg:main_tloc, arg:tes(f1_res5)])
fact(infon78, poss, ['John', id5, id25], 1, id24, id4)
ind(infon79, id25)
fact(infon80, inst_of, [ind:id25, class:thing], 1, univ, univ)
fact(infon81, isa, [ind:id25, class:book], 1, id24, id4)
fact(id26, read, [agent:id5, theme_aff:id25], 1, tes(finf1_res6), id4)
fact(infon85, isa, [arg:id26, arg:ev], 1, tes(finf1_res6), id4)
fact(infon86, isa, [arg:id27, arg:tloc], 1, tes(finf1_res6), id4)
fact(infon87, pres, [arg:id27], 1, tes(finf1_res6), id4)
fact(infon88, time, [arg:id26, arg:id27], 1, tes(finf1_res6), id4)
fact(id28, begin, [actor:id5, prop:id26], 1, tes(f1_res6), id4).
```

4 Performance on the Shared Task Texts

If we try to grade the seven texts of the shared task (Bos, 2008), from the point of view of their intrinsic semantic complexity we should get the following picture:

(a) Texts 6, 7 (scientific texts)
(b) Texts 4, 5 (newswire articles)
(c) Texts 1, 2, 3 (made up texts, schoolbook texts)

Overall, the system performed better with category (c). texts and worse with scientific texts, category (a). I take Text 6 and 7 to be in need of a specific domain ontology in order to have semantic inferences fired when needed. In addition, in our case, these two texts have sentences exceeding the maximum length for topdown parsing, which is the modality that better guarantees a full parse. Text 6 has sentences respectively 31, 38 and 49. In fact Text 1 represents an easy to understand scientific text and is much easier to parse — even though there are mistakes in Adjuncts attachment.

Apart from Texts 6 and 7, which lack in semantic relations due to the lack of semantic information, the remaining texts abound in semantically relevant syntactic information which can be used to assert facts in the Discourse Model which create a

network of meaningful associations. PAs, that is Predicate Argument structures, together with implicit optional and obligatory arguments are mostly recovered — more on this in the following sections.

The system has failed in finding antecedents for the pronoun IT. The current version of the complete system is not equipped with an algorithm that tells expletive IT cases from referential ones. On the contrary, one such algorithm has been successfully experimented with the partial system. Other pronouns are almost all correctly bound. As for nominal expressions, problems arise with scientific texts in case a different linguistic description is used to corefer or cospecify to the same entity.

For every text we will list pieces of what we call the Discourse Model World of Entities participating in the events described in the text. This file is produced at the end of the analysis and contains all entities recorded with a semantic Identifier by the system during the analysis of the text. The file is produced by a procedure that recursively searches the dynamic database of FACTS or Infons in Situation Semantics terms, associated to each entity semantic identifier. These Infons may register properties, attributes or participation in events. Eventually, Infons may also be inherited in case one of the entity is semantically included in another entity — see the case of CANCER being included in the more general notion of CANCERS at the end of Text 2.

The procedure produces a score that is derived from the relevance in terms of topichood — being Main, Secondary or Potential Topic — as asserted by the Centering algorithm. Entities and their associated infons are thus graded according to relevance. They are listed on the basis of their ontological status: INDividuals, SETs, CLASSes.

4.1 Text One

The main topic is the OBJECT. As can be gathered from the question posed to the system at the end of the parse, the main relations are all captured throughout the text. They can also be recovered from the Inherited Discourse World of Entities:

```
entity(ind,id2,9,facts([
  fact(infon111, coincide, [arg:id24, arg:id29], 1, tes(sn59_t13), id20),
  fact(infon4, isa, [ind:id2, class:object], 1, id1, univ),
  fact(infon5, inst_of, [ind:id2, class:thing], 1, univ, univ),
  fact(id9, throw, [tema_nonaff:id2, agente:id8], 1, tes(sn42_t11), univ),
  fact(id17, fall, [actor:id2, modale:id16], 1, tes(f1_t12), univ),
  fact(id29, take, [actor:id26, theme_aff:id2], 1, tes(finf1_t13), id20)])).
```

THROW is understood as being an event that takes place from a CLIFF and with a SPEED. However the SPEED is HORIZONTAL but the CLIFF is not HIGH — this relation has been missed. The OBJECT falls from a height of the same CLIFF. The one but last sentence is only partially represented. On the contrary, the final question is perfectly understood.

4.2 Text Two

The main topic is CANCER. From the Discourse World we know that:

```
entity(class,id3,2,facts([
  fact(infon7, inst_of, [ind:id3, class:stato], 1, univ, univ),
  fact(infon8, isa, [ind:id3, class:cancer], 1, id1, univ),
```

```
      fact(id4, cause, [theme_aff:id3, agent:id2], 1, tes(f2_t21), univ),
      fact(infon81, isa, [arg:id3, arg:cancer], 1, id25, id26),
      fact(id31, look, [actor:id27, locat:id3], 1, tes(f3_t23), id26)])).
```

CANCER is CAUSED by a VIRUS and that RESEARCHERs have been LOOKing for other CANCERs which receive a different semantic identifier but inherit all the properties:

```
    entity(class,id28,2,facts([   in(infon79, id28, id3),
      fact(infon75, cause, [ind:id28], 1, id25, id26),
      fact(infon76, of, [arg:id28, specif:id28], 1, univ, univ),
      fact(infon77, inst_of, [ind:id28, class:stato], 1, univ, univ),
      fact(infon78, isa, [ind:id28, class:cancer], 1, id25, id26),
      fact(*, inst_of, [ind:id28, class:stato], 1, univ, univ),
      fact(*, isa, [ind:id28, class:cancer], 1, id1, univ),
      fact(*, cause, [theme_aff:id28, agent:id2], 1, tes(f2_t21), univ),
      fact(*, isa, [arg:id28, arg:cancer], 1, id25, id26),
      fact(*, look, [actor:id27, locat:id28], 1, tes(f3_t23), id26)])).
```

The VIRUS is understood as the AGENT.

```
    entity(ind,id2,11,facts([
      fact(infon4, isa, [ind:id2, class:virus], 1, id1, univ),
      fact(infon5, inst_of, [ind:id2, class:animal], 1, univ, univ),
      fact(id4, cause, [theme_aff:id3, agent:id2], 1, tes(f2_t21), univ),
      fact(infon82, isa, [arg:id2, arg:virus], 1, id25, id26),
      fact(id29, cause, [agent:id2], 1, tes(f2_t23), id26)])).
```

The system also understands that those EVENTs, were KNOWn for some time, as shown by the ID8 which is bound in the discourse by means of THAT to the event id4 listed above,

```
    entity(ind,id8,1,facts([
      fact(infon21, prop, [arg:id8,
                   disc_set:[id4:cause:
                             [theme_aff:id3, agent:id2]]],
                   1, id6, id7),
      fact(infon31, isa, [arg:id8, arg:that], 1, id6, id7),
      fact(id12, know, [tema_nonaff:id8, actor:id11], 1, tes(f2_t22), id7)])).
```

However the system has not bound IT to THAT so we do not know what LEADs to a vaccine, nor do we know what prevents from what. All IT are unbound.

4.3   Text Three

This is the text that we proposed for the shared task and is already completely and consistently semantically and pragmatically represented. It has already been presented above.

4.4   Text Four

The text is not completely and consistently represented but most of the relations are fully understood. In particular consider THEY in the third sentence which is rightly bound to the SET of two trainers asserted in the Discourse World. The school is always coindixed. The last sentence contains a first plural pronoun WE which is interpreted as being coindexed with the narrator, but also wrongly with the location of the text.

4.5   Text Five

The text is not completely and consistently represented but most of the relations are fully understood. We still know a lot about the main Entities, the PROPELLANT and NITROCELLULOSE which is composed in CHUNKs.

```
entity(ind,id19,8,facts([
  fact(infon42, inst_of, [ind:id19, class:sub], 1, univ, univ),
  fact(infon43, isa, [ind:id19, class:propellant], 1, id18, nil),
  fact(infon44, isa, [arg:id19, arg:propellant], 1, id18, univ),
  fact(id20, explode, [agent:id19], 1, tes(f1_t53), univ),
  fact(infon108, isa, [arg:id19, arg:propellant], 1, id30, univ),
  fact(id38, use, [theme_aff:id19, actor:id37], 1, tes(f2_t55), univ),
  fact(id41, make, [theme_aff:id19, actor:id40, loc_origin:id31],
                    1, tes(sn32_t55), univ),
  fact(id20, explode, [agent:id19], 1, tes(f1_t53), univ),
  fact(infon50, sub, [prop:id20], 1, id18, univ)])).

entity(ind,id32,1.2,facts([  in(infon91, id32, id31),
  fact(infon89, inst_of, [ind:id32, class:sub], 1, univ, univ),
  fact(infon90, isa, [ind:id32, class:nitrocellulose], 1, id30, nil),
  fact(*, nitrocellulose, [ind:id32], 1, id30, nil),
  fact(*, produce, [ind:id32], 1, id30, nil),
  fact(*, repackage, [ind:id32], 1, id30, nil),
  fact(*, of, [arg:id32, specif:id31], 1, univ, univ),
  fact(*, of, [arg:id32, specif:id31], 1, univ, univ),
  fact(*, of, [arg:id32, specif:id31], 1, univ, univ),
  fact(*, inst_of, [ind:id32, class:col], 1, univ, univ),
  fact(*, isa, [ind:id32, class:chunk], 1, id30, nil),
  fact(*, make, [theme_aff:id19, actor:id40, loc_origin:id32],
                  1, tes(sn32_t55), univ)])).

entity(set,id31,1,facts([  card(infon79, id31, 5),
  fact(infon80, nitrocellulose, [ind:id31], 1, id30, nil),
  fact(infon81, produce, [ind:id31], 1, id30, nil),
  fact(infon82, repackage, [ind:id31], 1, id30, nil),
  fact(infon83, of, [arg:id31, specif:id31], 1, univ, univ),
  fact(infon86, inst_of, [ind:id31, class:col], 1, univ, univ),
  fact(infon87, isa, [ind:id31, class:chunk], 1, id30, nil),
  fact(id41, make, [theme_aff:id19, actor:id40, loc_origin:id31],
                    1, tes(sn32_t55), univ)])).
```

The relation intervening between CHUNKS and NITROCELLULOSE endows transitivity to the EVENTS taking place so that both are involved in REPACKAGE, PRODUCE, MAKE. We also know that a CREWMAN was OPERATING at a center and that the GUN CREW was KILLed, by an unknown AGENT, id26.

```
entity(class,id23,6,facts([
  fact(infon55, of, [arg:id23, specif:id8], 1, univ, univ),
  fact(infon56, inst_of, [ind:id23, class:institution], 1, univ, univ),
  fact(infon57, isa, [ind:id23, class:crew], 1, id22, nil),
  fact(id27, kill, [theme_aff:id23, agent:id26], 1, tes(f2_t54), univ)])).
```

We know that EVENTS happened during WORLD_WAR_II. Also notice that IT SUBJect of SUSPECT is correctly computed as an expletive.

4.6 Text Six

Two of the sentences are parsed by the partial system, but the main relations are well understood. The FARM and the COMMUNITY provide FOOD and EARNs a REVENUE.

```
entity(ind,id13,3,facts([
  fact(infon30, inst_of, [ind:id13, class:informa], 1, univ, univ),
  fact(infon31, isa, [ind:id13, class:farm], 1, univ, univ),
  fact(id17, provide, [goal:id8,tema_nonaff:id7,actor:id13],1,univ,univ),
  fact(infon85, isa, [arg:id13, arg:farm], 1, id41, univ),
  fact(id43, earn, [agent:id13, theme_aff:id42], 1, tes(sn59_t63),univ)])).

entity(ind,id7,0,facts([
  fact(infon10, inst_of, [ind:id7, class:any], 1, univ, univ),
  fact(infon11, isa, [ind:id7, class:food], 1, univ, univ),
  fact(id17, provide,[goal:id8,tema_nonaff:id7,actor:id13],1,univ,univ)])).

entity(ind,id42,2,facts([
  fact(infon83, inst_of, [ind:id42, class:legal], 1, univ, univ),
  fact(infon84, isa, [ind:id42, class:revenue], 1, id41, nil),
  fact(id43, earn, [agent:id13, theme_aff:id42], 1, tes(sn59_t63), univ)])).
```

The COMMUNITY LACK the FOOD

```
entity(ind,id8,0,facts([
  fact(infon13, inst_of, [ind:id8, class:luogo], 1, univ, univ),
  fact(infon14, isa, [ind:id8, class:community], 1, univ, univ),
  fact(id17, provide, [goal:id8,tema_nonaff:id7,actor:id13],1,univ,univ),
  fact(id14, lack, [theme_aff:id9, actor:id8, purpose:cl5, result:id14],
                   1, univ, univ)])).
```

Most of the sentences are parsed by the partial system. However questions can be asked and get a reply, even though the generator does not handle uncountable nouns like MONEY properly.

4.7 Text Seven

The most difficult text is fully parsed but not satisfactorily semantically represented. We only know few things, and they are all unrelated. There is no way to related WIND to TURBINE and to ENERGY in a continuous way.

```
entity(set,id61,4,facts([  card(infon253, id61, 5),
  fact(infon254, power, [nil:id61], 1, id60, id20),
  fact(infon255, maximum, [ind:id61], 1, id60, id20),
  fact(infon256, of, [arg:id61, specif:id61], 1, univ, univ),
  fact(infon257, wind_turbine, [ind:id61], 1, id60, id20),
  fact(infon258, inst_of, [ind:id61, class:thing], 1, univ, univ),
  fact(infon259, isa, [ind:id61, class:[wind, turbine]], 1, id60, id20),
  fact(infon264, of, [arg:id63, specif:id61], 1, univ, univ),
  fact(infon267, isa, [arg:id61, arg:wind_turbine], 1, id60, id20),
  fact(infon268, isa, [arg:id61, arg:power], 1, id60, id20),
  fact(infon269, typical, [arg:id61], 1, id60, id20),
  fact(infon271, power, [nil:id61, arg:id61], 1, id60, id20)])).
```

```
entity(ind,id14,2,facts([
  fact(infon52, inst_of, [ind:id14, class:abstract_state], 1, univ, univ),
  fact(infon53, inst_of, [ind:id14, class:energy], 1, univ, univ),
  fact(infon54, isa, [ind:id14, class:energy], 1, univ, univ),
  fact(infon55, isa, [ind:id14, class:wind_energy], 1, univ, univ),
  fact(infon58, of, [arg:id15, specif:id14], 1, univ, univ)])).

entity(ind,id22,1,facts([   in(infon90, id22, id15),
  fact(infon88, inst_of, [ind:id22, class:thing], 1, univ, univ),
  fact(infon89, isa, [ind:id22, class:wind], 1, id19, id20),
  fact(*, isa, [ind:id22, class:wind], 1, univ, univ),
  fact(*, of, [arg:id22, specif:id14], 1, univ, univ)])).
```

We know that WIND and ENERGY are related, and also that there is one such technology, but is semantically set apart, due to orthography.

```
entity(class,id11,1,facts([
  fact(infon39, 'wind-energy', [ind:id11], 1, id1, univ),
  fact(infon44, of, [arg:id11, specif:id12], 1, univ, univ),
  fact(infon45, inst_of, [ind:id11, class:abstract_state], 1, univ, univ),
  fact(infon46, isa, [ind:id11, class:technology], 1, id1, univ)])).

entity(class,id12,0,facts([
  fact(infon41, inst_of, [ind:id12, class:astratto], 1, univ, univ),
  fact(infon42, isa, [ind:id12, class:energy], 1, univ, univ),
  fact(infon44, of, [arg:id11, specif:id12], 1, univ, univ),
  fact(infon103, has, [arg:id26, tema:id12], 1, id19, id20),
  fact(infon109, of, [arg:id26, specif:id12], 1, univ, univ)])).
```

I assume that scientific language requires a different setup of semantic rules of inference, which can only be appropriately specified in a domain ontology.

References

Bos, J. (2008). Introduction to the Shared Task on Comparing Semantic Representations. In J. Bos and R. Delmonte (Eds.), *Semantics in Text Processing. STEP 2008 Conference Proceedings*, Volume 1 of *Research in Computational Semantics*, pp. 257–261. College Publications.

Bresnan, J. (2001). *Lexical-Functional Syntax*. Oxford: Blackwell.

Buitelaar, P. (1998). *CoreLex: Systematic Polysemy and Underspecification*. Ph. D. thesis, Brandeis University.

Delmonte, R. (2007). *Computational Linguistic Text Processing: Logical Form, Semantic Interpretation, Discourse Relations and Question Answering*. New York: Nova Science Publishers.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Cambridge (MA): MIT Press.