UNDERSTANDING IMPLICIT ENTITIES AND EVENTS WITH *GETARUNS*

Rodolfo Delmonte

Università Ca' Foscari - Department of Language Sciences
E-mail:delmont@unive.it

Abstract

Semantic processing represents the new challenge for all applications that require text understanding, as for instance Q/A. In this paper we will focus on "implicit" or lexically unexpressed linguistic elements that are nonetheless necessary for a complete semantic interpretation of a text. We will address the following types of "implicit" entities and events:

- the grammatical ones, as suggested by a linguistic theories like LFG or similar generative theories;

- the semantic ones suggested in the FrameNet project, i.e. CNI, DNI, INI;

- the pragmatic ones: here we will present a theory and an implementation for the recovery of implicit entities and events of (non-) standard implicatures.

In particular we will show how the use of commonsense knowledge may fruitfully contribute in finding relevant implied meanings. We will also briefly explore the Subject of Point of View which is computed by Semantic Informational Structure and contributes the intended entity from whose point of view is expressed a given subjective statement.

At the end of the paper we propose an evaluation based on section 24 of Penn Treebank as encoded by LFG people in the PARC-700 treebank where lexically unexpressed are adequately classified and diversified.

1. Introduction

In this paper we intend to present a comprehensive method to cope with implicit or lexically unexpressed entities and events which we will call for short IEEs. Our starting point is work on bridging referential links by inferential processes as Hahn et al. 1996 propose for what they call "textual ellipsis"; as well as work on filling in unexpressed semantic (or thematic) roles as contained in the lexical form associated to predicates in computational lexica as proposed by Palmer et al. 1986 in specific domains like maintenance manuals. We intend to propose a more general and linguistically-based principled method which crucially hinges upon semantic representation. Knowledge to be used by the system is provided by computational lexica like WordNet and complex subcategorized lexica like COMLEX or NOMLEX [2], FrameNet [3], or VerbNet [4] and PropBank [5] and other similar tools, where selectional restrictions are encoded at different levels of granularity.

Systems using these lexica will try to distinguish Arguments from Adjuncts and to build appropriate representations for Predicate Argument Structures; this in turn is achieved at clause or propositional level. These latter structures are fed as input to the semantic interpreter that can apply principles of grammatical, lexical and semantic well-formedness to the analysis. It is just in this phase that the presence of IEEs can be detected by the system and an adequate semantic representation can be built. IEEs can be classified according to the following linguistically motivated subdivision:

a. grammatically motivated IEEs
   o as a subtype, IEEs identified by pronominal binding
b. semantically motivated IEEs

o   as a subtype, IEEs identified by anaphoric binding
c. pragmatically motivated IEEs
    o   semantically inferred IEEs
d. discourse motivated IEEs
    o   Centering Main Topic IEEs

We will now give examples of the four types and comment on their status in a theoretical and computational framework. The presentation will use LFG as linguistic theory and FrameNet as semantic lexical theory; semantic representations are inspired by Situational Semantics.

To assess the importance of correctly identifying IEEs, we counted them in the Penn Treebank II (hence PTB) where there are approximately 65000 empty categories, the majority of which is constituted by traces – in a Chomskian sense (our grammatical IEEs) – of moved linguistic material. Empty elements constituted by unexpressed Subjects of untensed clauses are some 27800. Then there is some 580 elliptical empty elements. If we consider that the PTB contains 93539 simple sentences we see that there is almost one such empty element per sentence. Since 38133 are SBAR clauses, they contain each one empty coindexed element. Hence, computing implicit elements is an important component of any semantically viable text analysis component.

1.1 Grammatically motivated IEEs

According to LFG theory as proposed by [1] grammatical IEEs may be classified into three types:
- Lexical Control
- Syntactic Control
- Structural Control

In order to activate control mechanisms, two procedures need to be implemented in the parser: an indexing function that distinguishes phrase structures from one another; a coindexing function and a corresponding chain climbing function to recover the semantic identity or the head of the controller.

1.1.1. Lexical control IEEs are constituted by empty categories that ensue from the presence of a phrase structure which has been computed as argument of the governing predicate and needs at least the presence of a SUBJect. An empty category is created which is then coindexed with the lexical control argument. Relevant examples are (1), (2) and (3) below:

(1) Mary asked John to buy a book
(2) John considers Mary an important ally
(3) Tom is a republican.

In (1) the predicate BUY is associated to an empty SUBJect which is coindexed with the controller JOHN in force of the existence of a lexical rule that selects a controller between a hierarchy of Grammatical Functions associated to lexically expressed arguments of the governing predicate, in this case ASK. The hierarchy is the Default Rule of Lexical Control and establishes the following order:

(2.1) OBJ2 < OBJ < SUBJ

which simply says that the lexical controller is an OBJ2 if present, an OBJ if present, otherwise a SUBJ. In (2) the predicative NP "an important ally" is lexically controlled by the OBJect Mary in force of the same Default Rule, where the argument receives semantic role also from the lexical controller. Same situation with (3) where however the SUBJect is the controller of the predicative NP "a republican".

1.1.2. Syntactically controlled IEEs refer to what are also called long distance dependencies. These constructions concern two types of clauses: relative clauses and interrogative clauses. These are too well known by computational linguists to require a presentation. We just include two examples to complete the description:

(4) Tom wanted the book that Mary bought.
(5) Which book did Mary buy?

In both examples the predicate BUY needs the existence of an empty category which is then filled by or coindexed with the syntactic controller, the BOOK. This may be achieved by different procedures according to each linguistic theory, but the final result is always the same: a chain between two structures one of which has a control index added by the grammar.

1.1.3. Structurally controlled IEEs are those SUBJects that come into existence whenever a predicative ADJunct appears in a certain structural configuration. We include here below the relevant examples:

(6) John went to the see the movie drunk.
(6.1) John accompanied Mary to the movie naked.
(6.2) Drunk as usual John went to see the movie.
(6.3) Naked as usual John took Mary to the movie.
(7) The company has sold its assets to collect funds.
(7.1) These assets have been sold to collect funds.
**(7.2) These assets sell well to collect funds.
(8) After reading the letter Mary rushed to the school.
(8.1) Mary met John after finishing school.
(8.2) Reading books is important.
(8.2.1) Reading books is important for John.
(9) Ski John loves!
(9.1) At the corner was standing a young girl.

These examples do not exhaust all possible cases of structurally relevant IEEs. We omitted intentionally cases of so-called "parasitic gaps" which we consider

highly rare in real texts to be taken into consideration. We also omitted the case called OBJect intransitivization that will be discussed in the following section.

Example (6) is a case of an adjectival ADJunct which has the SUBJect as controller. This may be due to the unsuitability of MOVIE as controller of DRUNK. If we look at example (6.1), possible controllers are both Mary and John for NAKED. So the OBJect Mary is taken. However, according to the position of the ADJunct control may pass to the SUBJect John. In other words, structural control does not answer only to grammatical criteria, but also to semantic criteria and finally to positional ones.

Example (7) is an interesting case where we see that the controller may also be omitted and in that case it needs to be restored from previous discourse. We deal with such cases below. Notice here the important fact represented by (7.2) where we included an ungrammatical case – a sentence that will not be found in real texts. Whenever the AGENT is not lexically expressed nor can be posited by grammatical principles no control may ensue. Crucially then, in order for a control structure like the RESULTATIVE infinitive to be expressed in a sentence, some controller needs also to be there. Then we have cases represented by examples under (8) which are all gerundives. As it seems, the controller is always the SUBJect disregarding its position. The copulative construction in (8.2) introduces another type of control, the one called ARBITRARY control. As can be noticed, (8.2.1) is no longer a case of Arbitrary control because of the presence of a BENEFICIARY "for John" who becomes the controller. Final cases are those constituted by so-called inverted focus structure – example (9) – and locative inversion in (9.1). These cases do not require the insertion of an empty category but a shallow parser is usually unable to cope with them appropriately [see 4]. Deep processing will impose an appropriate argument structure by means of selectional restrictions, but also compute as SUBJect the inverted NP in the locative construction.

2. Semantically motivated IEEs

In this section we discuss cases of lexical-semantic IEEs which are also discussed in the FrameNet project, which is theory of lexical representation and is based on its underlying linguistic theory, i.e. Constructional Grammar. Differently from our approach which is mainly computational, this project is descriptive and wrongly conflates cases of lexical semantics IEEs with cases of grammatical IEEs in the same typology. Starting from CNI, this class of implicit entities concerns structurally omitted constituents as can be

gathered from the definition given in their Manual (ibid. p. 54)

Under the term CNI we find three types of IEEs conflated under the same definition, some of which have already been discussed above. Computationally speaking, these types require totally different tools and procedures to be activated. They may be redefined as follows:

2.1. pronominal binding cases

(Big-)PRO cases as found in independent or Adjunct infinitives, participials and gerundives, i.e. in clauses with an untensed verb.

These cases have already been presented above. The PRO SUBJect inherits lexical properties associated to the subcategorization frame and may thus be pronominally bound to a structural controller, if any exist. Otherwise, the PRO is computed as generic or arbitrary: as a result, PRO cannot possibly be computed as external pronouns that can corefer in the discourse.

Finally, for these CNI to be computed, their presence is posited by the Interpretation Component of the system, which recovers Predicate Argument Structures or PASs by applying grammatical completeness and other principles to the output of the parser – in our case to c-structure. In our case, the output of the Interpretation Component are f-structures, i.e. semantically complete PASs. Big-PROs will then be bound by the Pronominal Binding component of the system, which only works at sentence level, using structural information and principles of the grammar.

2.2. anaphoric binding cases

Little-PRO cases for those languages – Romance but not only - that allow a SUBJect to be left lexically unexpressed in clauses with a tensed verb. The pronoun is added by the Interpretation Component as above and may be bound at sentence level. In addition, and differently from big-PRO it may become an external pronoun, which is then bound at discourse level. In this case, discourse level processing components like Topic Hierarchy and Centering - that will be presented below - will contribute to find the appropriate antecedent. This case includes IMPERATIVE mood sentences which require the SUBJect to be left unexpressed. It also includes COORDINATE structures omitted understood SUBJect as in the example, "John went out and *pro* met Mary" which we comment below, where the omitted SUBJect must be copied from the previous clause by the grammar embodied by the parsing module, rather than by interpretation or by lexically related principles. There is always the need to have a SUBJect expressed with all verb predicates.

2.3 Ellipsis: on the edge between Syntax and Semantics

We take all cases of ellipsis to be collapsed in a single computational action: copy of the elliptical material in the place where it is missing. This can only be done once the complete utterance that precedes the elliptical one is fully parsed. Also it is important to remind the fact that this process can be spotted only in case there is no ambiguity, as shown in the examples below:

(10) John went out and [*John*] met Mary
- SUBJECT null in coordination
(11) John often kisses Mary, and Bill does/will [*kiss Mary*], too.
- VP ellipsis
(12) John carefully counted the money, and Bill did/will [*carefully count the money*], too.
- VP ellipsis and Adverb ellipsis
(13) Harry lives in Boston and Mike [*lives*] in New York
- Gapping
(14) Susie wants to buy a car and my brother [*wants [to buy*]] a bike.
- LD Gapping
(15) Venice is the city where I live and [*where*] I work
- Forward Conjunction Reduction
(16) Some have served mussels to Sue while others have [*served*] swordfish
- PseudoGapping
(17) He takes and never gives back.
- Indefinite Null Instantiation (INI)
(18) What? Who?
- Sluices

As can be easily noticed, the only case in which ambiguity may constitute a problem is PseudoGapping (see example 16), which is a case of Auxiliary ellipsis, and English auxiliaries are ambiguous between lexical and non lexical usage. Before continuing examining other IEEs, we need to briefly describe the system GETARUNS – which may be freely downloaded from the www.sigsem.org Main page following the link of the SharedTask held in Venice and then cliking on GETARUNS. The working of the system is described in [6].

3. The system GETARUNS

Here we can only point to the fact that the system is organized as the usual pipeline of modules, divided up into a lower and an upper level, where the lower level computes sentence level interpretation and records the output in DAGs (Direct Acyclic Graphs). At this level the system also computes pronominal binding and quantifier raising.

3.1 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 a 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 (direct acyclic graphs) 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. Inferences can be drawn on the facts repository as will be discussed below.

3.1.1. Discourse Model cases. Omitted Agent of passive sentences already discussed above, which we treat as we do with cases of OBJect intransitivization, i.e. by adding a dummy existential quantifier. The Agent of passive sentences will then be identified by the semantic processing module which will look for a similar governing predicate in the context, or previous stretch of discourse. When the predicate is found the argument will be identified and the current existential bound to it in the Discourse Model.
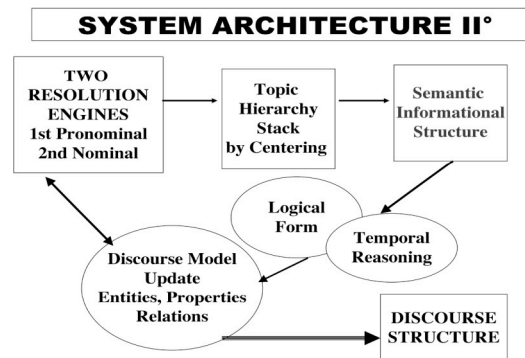

Figure 1. High Level Modules of GETARUNS

3.1.2. Semantic coreference cases. There is no need to specify a dummy (big)-PRO, little pro or existential quantifier in these cases because the missing element is an Adjunct and not an Argument as was the case with the examples discussed above. So the only way to recover the identity of the lexical entities coreferred by the optional adjuncts "evoked" by these structures is to search in the context, or in the previous stretch of discourse – the Discourse Model – for a similar semantic relation. When an identical predicate is found, the arguments are recovered and their semantic identifiers used to complete the extended PAS for the current predicate.

3.1.3. Other Discourse Model cases. We postulate a semantic treatment of empty deleted OBJect for those transitive verbs that allow it that is lexical. The solution of the problem lies in the lexical nature of the phenomenon of OBJect intransitivation, which must be marked for Intransitivization, i.e. these are transitive verbs that may become Intransitives. Seen that transitive verbs constitute the great majority of all verbs in any language, and the ones allowing intransitivization is a small subset, they shall have to be marked so. The empty OBJect can then be added to the extended PAS by the semantic component. Similar cases are constituted by the deletion of OBJ2 in ditransitive verbs, as also shown by example (17) above. So, we prefer to consider the OBJ2 as an existential that needs to be recovered, and leave the OBJ unexpressed.

4. Implicit Entities and Implicatures

Conversational implicatures and implications in general, are based on an assumption by the addressee that the speaker is obeying the conversational maxims [8], in particular the cooperative principle. The well-known example from Levinson [8, p. 107],

Text 1.
A: Can you tell me the time?
B: Well, the milkman has come.

requires that both interlocutors share the same spatiotemporal location, besides the same conventions and habits. Not everywhere you can find milkmen go around delivering milk.
Now consider the following example always from Levinson [8, p.104],

Text 2.
A: I've just run out of petrol.
B: Oh; there's a garage just around the corner.

Here we see that spatiotemporal locations are even more important: if speaker A needs fuel then the addressee indicates a spatial location, the garage, which in addition has to be open – hence a temporal location. More on this example below.

So, we would like to regard the mechanism that recovers standard implicatures and conversational implications in general, as a reasoning process that uses the knowledge contained in the semantic relations actually expressed in the utterance to recover hidden or implied relations or events as we call them. This reasoning process can be partially regarded as a subproduct of an inferential process that takes spatiotemporal locations as the main component and is triggered by the need to search for coreferent or cospecifiers to a current definite or indefinite NP head. This could be interpreted as bridging referential expression entertaining some semantic relation with previously mentioned entities. In Text (2) the initial inference would be triggered by the metonymy relation intervening between "petrol" and CAR. At the same time CAR would be the trigger of the GARAGE reference, always metonymic. If we consider now Text (1), we see that the request of the current time is itself bound to a spatiotemporal location. Using the MILKMAN rather than a WATCH to answer the question, is relatable to spatiotemporal triggers. In fact, in order to infer the right approximate time, we need to situate the COMING event of the milkman in time, given a certain spatial location. Thus, it is just the "pragmatic restriction" associated to SPACE and TIME that is implied in the answer, that may trigger the inference. More on this topic below.

4.1 The Restaurant text

To exemplify some of the issues presented above we present a text by Sanford and Garrod [13;14] called the Restaurant text. In this text, entities may be "scenario-dependent" [15] or main characters that are independent thereof. While the authors use the text for psychological experimental reasons, we will focus on its computability. So first of all the sentences making up the text, here below,

Text 3[1].
0. At the restaurant.
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.

---

[1] The text has also been used in the challenge of the Shared Task associated to STEP2008, and its full analysis is available at the link with the same name under SIGSEM main page.

5. He began to read his book.

Here below we will only comment on implicatures and implicit arguments. The text is also defined as a "psychological statement" text, i.e. it includes sentence (4) that represents a psychological statement, that is it expresses the feelings and is viewed from the point of view of one of the characters 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". However, in order to capture such information, a system has to compute Point of View and Discourse Domain on the basis of Informational Structure and Focus Topic by means of a Topic Hierarchy algorithm based on [10], which has been lately evaluated in [6].

Before going into more detailed explanations, a brief comment on Schank's approach [16] – but see also Mueller [18] – is in order. Schank introduces *scripts* and a theory of conceptual dependencies which are based on primitive actions which are very close to FrameNet's frames. In Schank's perspective, *plans* are the means for satisfying goals and they are composed of *scripts*. To understand a story one needs *scripts* and a *plan*. However, differently from what we do here, *scripts* are organized with metadata that contain for instance preconditions, instrumental relations etc. and other precompiled schemes of knowledge. So the great difference in our approach is represented by the fact that we certainly make no use of preorganized conceptual structures or schemes. We do not even make use of Frames, in the literal sense, in order to produce lexical inferences. All we accomplish is based on the Discourse Model which is a sequence of facts built by the actual linguistic structures instantiated in a given text as it is analyzed and represented in situational semantics. A lot of additional information might be gathered from the commonsense reasoning repository we shall comment on in the following section, and introduced in the DM. However, we don't want to clutter our dynamic knowledge base with irrelevant and redundant information. So only the information strictly required will be searched for as explained below.

4.2 Commonsense reasoning and IEEs

We will concentrate our attention to sentence (3) at first, which is an example of INI. To account for the fact that whenever a waiter takes an order there is always someone that makes the order, we compute TAKE_ORDER as a compound verb with an optional GOAL argument that is the person ORDERing something. The system then looks for the current Main

Topic of discourse or the Focus as computed by the Topic Hierarchy Algorithm, and associates the semantic identifier to the IEE. This latter procedure is triggered by the "existential" dummy quantifier associated to the implicit optional argument. However, another important process has been activated automatically by the presence of a singular definite NP, "the WAITER", which is searched at first in the Discourse Model of entities and properties asserted for the previous stretch of text. Failure in equality matching activates the bridging mechanism for inferences which succeeds in identifying the WAITER as a Social Role in a Restaurant, the current Main Location.

Consider now sentence (2) which introduces a TABLE as main Topic. This type of sentences is called "presentational" and has the pragmatic role of "presenting" an entity on the scene of the narration in an abrupt manner, or as Centering would definite it with a SHIFT move. However, the TABLE does not constitute a suitable entity to be presented on the scene and the underlying import is triggering the inference that "someone is SITting at a TABLE". This inference is guided by the spatiotemporal component of the system. GETARUNS is equipped with a spatiotemporal inferential module that asserts Main SpatioTemporal Locations to anchor events and facts expressed by situational infons. This happens whenever an explicit lexical location is present in the text. In our case, the location expressed is the Restaurant. This can either be part of the title or just be derived from the first sentence of the text, where it has the role of LOCATion argument of the governing verb GO and the preposition INTO. The second sentence contains an expressed location: the CORNER. Now, the inferential system will try to establish whether the new location is either a deictic version of the Main Location; or it is semantically included in the Main Location, or else it is a new unconnected location that substitutes the previous one. The "corner" is in a meronymic semantic relation with "restaurant" and thus it is understood as being a part_of it. This inference is the trigger of the IMPLICATURE that the TABLE is a metonymy for the SITting event. Consequently, when the system tries to corefer, cospecify or assert new semantic individuals, it will find an Indefinite expression "a table" which will not just constitute literally that the text presents a new entity TABLE, but that the IE is involved with a related event. The Entity implied is again understood as the Main Topic of current Topic Hierarchy, i.e. JOHN.

The procedure invoked by the system to produce such an implicature makes a call to WordNet that checks all possible inclusion relations: hyponymy, hyperonymy, meronymy, etc. Then we look for current main location specified for spatial "place" locations and recover the Predicate the "restaurant". Then the semantic index – Idy - of the Main Topic is searched and passed down to

the predicate that will compute the implicature. The following procedures produce the other semantic index associated to TABLE and assert its semantic properties; it asserts an inclusion for the location Table into the Main location.

Then there is the final call which has the task to search for unexpressed relations intervening in the current spatiotemporal location. To solve this problem in a principled matter we needed commonsense knowledge organized in a computationally tractable way. This is what CONCEPTNET 2.1 [12] actually constitutes. ConceptNet - available at www.conceptnet.org - is the largest freely available, machine-useable commonsense resource. Organized as a network of semi-structured natural language fragments, ConceptNet consists of over 250,000 elements of commonsense knowledge. At present there are 19 semantic relations used in ConceptNet, representing categories of, inter alia, temporal, spatial, causal, and functional knowledge. The representation chosen is semi-structured natural language using lemmata rather than inflected words. The way in which concepts are related reminds "scripts", where events may be decomposed in Preconditions, Subevents and so on, and has been inspired by Cyc [11].

ConceptNet can be accessed in different ways, we wanted a strongly constrained one. We chose a list of functions that encode pieces of knowledge and use those functions together with the information available at a certain point of the computation to derive Implicit Information. In other words, we assume that what is being actually said hides additional information which however is only implicitly hinted at. What we need is a predicate constrained by a conceptual function and other predicates. So first of all the list of functions,

allsceneryevents(['SubEventOf','FirstSubeventOf', 'DesiresEvent','Do',CapableOf','FunctionOf','UsedFor','Ev entRequiresObject',LocationOf']

Then the call that searches ConceptNet for implicit information,

create_infer_rel(NoFr, MainLoc, AgentId, CurrLocatId, Temp, Loc):-
allsceneryevents(CondEvents),
member(Type, CondEvents),
MatchScenery=..[Type, [go, Prep, MainLoc], [Event, Preps, CurrLocat])
 …
 Infon=..[fact, EvId, Event, [actor:AgentId, locat: CurrLocatId], 1, tes(Tr3), Loc],
      assert(Infon),

If the call has success, we end up by recovering a predicate SIT in the slot Event, and use this predicate to assert an additional property associated to the Topic of

discourse. So eventually, the system checks for implicatures because it is triggered by the unsuitability of the current entity – the TABLE – as topic of discourse.

With a similar strategy can be resolved the non-standard implicature involved in Text 2, that we repeat here below,

Text 2.
A: I've just run out of petrol.
B: Oh; there's a garage just around the corner.

There are a number of missing conceptual links that need to be inferred, as follows:
Inf1: the CAR has run out of petrol
Inf2: the CAR NEEDS petrol
Inf3: garages SELL PETROL for cars
In addition, in order to use ConceptNet we need to translate "petrol" and "garage" into "gas/gasoline" and "gas station" respectively. This passage is not just a mere translation but requires contextual information to tell apart the two meanings associated to the word "garage" – that is the place where to keep your car, and the place where to get gas. Now we can query the ontology as we did previously and will recover the following facts. The whole process starts from the first utterance and uses RUN OUT OF GAS,

(Do "car" "run out of gas")

Then we can use GAS STATION and CAR to build another query and get,

(Do "car" "get fuel at gas station")

where FUEL and GASoline are in IsA relation. We may still get additional information on the reason why this has to be done,

(Do "person" "don't want to run out of gas")
(SubeventOf "drive car" "you run out of gas")
(Do "car" "need gas petrol in order to function")
(Do "gas station" "sell fuel for automobile")

These may all constitute additional commonsense knowledge that may be used to further explain and clarify the implicature.

5. Experimental Results

We used Parc-700 annotated treebank because it contains precise information about a subset of our IEEs, i.e. unexpressed subjects of coordinate structures, and unexpressed subjects of infinitivals, gerundives and participials which we treat separately as untensed clauses. Data related to the linguistic items to analyse are as follows:

i. 512 IEEs of Untensed Clauses
ii. 47 IEEs of Tensed Clauses

Precision and Recall for i. is 75% and 85.54% with an F-score of 79.92%. As to the other category of IEEs we got respectively 83.6% and 87.4% with an F-score of 85.46%.

6. Conclusions

We have presented a thorough treatment of implicit entities and events that encompasses most if not all semantically relevant lexically unexpressed elements. This has been implemented in a system called GETARUNS which deals with all these phenomena in a principled way by means a theoretically validated division of labour between the different modules that make up the whole pipeline. We subdivided IEEs into different categories according to both theoretical and computational criteria. In this way grammatical IEEs are taken care before lexically semantically motivated ones. In turn these latter come before the need to carry out pronominal binding and anaphora resolution. Finally, when the semantic components are completing their mapping and search the preceding Discourse Model for coreferring/cospecifying entities, procedures that look for implicatures are activated and inferences are fired. This can only be done in presence of a full-fledged semantic interpretation of the current utterance, because it is just by means of its PAS that the appropriate implicit events may be recovered. World knowledge is represented by two repositories: a generic semantic network like WordNet and the commonsense ontology ConceptNet. Again, other similar repositories may be used, but the mechanisms to access them should be the same: no implicature may be recovered without a full semantic interpretation of the triggering utterances. Dialogues and texts are full of IEEs either as elliptic material or as implicated events and we are currently experimenting with Multiparty Meetings Dialogues from ICSI in order to verify what impact may they have on the overall interpretation process.

6. References

[1] Bresnan, J., 2000. *Lexical-Functional Syntax*, Blackwell.

[2] ComLex:- http://nlp.cs.nyu.edu/comlex.

[3] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In Proceedings of COLING-ACL-98, Montreal, Canada.

[4] Palmer M, Kingsbury P, Gildea D (2005). "The Proposition Bank: An Annotated Corpus of Semantic Roles". Computational Linguistics 31 (1): 71–106.

[5] Hoa Trang Dang, Karin Kipper, Martha Palmer. Integrating compositional semantics into a verb lexicon. COLING-2000 Eighteenth International Conference on Computational Linguistics, 8 pg. Saarbrucken GERMANY, July 31- August 4, 2000.

[6] Delmonte R., (2007), *Computational Linguistic Text Processing*, Nova Science Publishers, New York.

[7] Fellbaum, Christiane, (ed.)(1998), *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge MA.

[8] Levinson, Stephen C. 1983. *Pragmatics*. Cambridge, Cambridge University.

[9] Grice, H. P., 1975. Logic and Conversation. in P. Cole & J. L. Morgan, Syntax and Semantics, Vol. 3: Speech Acts. New York : Academic Press, 41-58.

[10] Grosz, B. 1981 Focusing and description in natural language dialogues. In Joshi, A., B. Webber & I. Sag (eds) *Elements of discourse understanding*. Cambridge UP.

[11] Lenat, D.B. (1995). CYC: A large-scale investment in knowledge infrastructure. Communications of the ACM, 38(11).

[12] Liu, H., Singh, P. (2004). ConceptNet: A Practical Commonsense Reasoning Toolkit. At http://web.media.mit.edu/~push/ConceptNet.pdf.

[13] Sanford, A.J., & Garrod, S. (1998). The role of scenario mapping in text comprehension, Discourse Processes, 26, 159 - 190.

[14] Anderson, A., S. C. Garrod, A. J. Sanford, The accessibility of pronominal antecedents as a function of episode shifts in narrative text, 1983.

[15] Sanford, A.J. and Garrod, S.C., Thematic subjecthood and cognitive constraints on discourse structure, 1988. Journal of Pragmatics, Volume 12, Issues 5-6, December 1988, pp. 519-534.

[16] Schank R C & Abelson R P. 1977. *Scripts, plans, goals and understanding: an inquiry into human knowledge structures*. Hilsdale, NJ: Erlbaum.

[17] Sidner, C. 1983. Focusing in the comprehension of definite anaphora. In M. Brady & R. Berwick (eds) *Computational models of discourse*. MIT Press, Cambridge, MA: 267-330.

[18] Erik T. Mueller, 2007. Modelling Space and Time in Narratives about Restaurants, *Literary and Linguistic Computing*, Vol. 22, No. 1, 67-84.