Computing implicit entities and events for story understanding

Rodolfo Delmonte, Emanuele Pianta
Università Ca' Foscari and IRST, Fondazione Bruno Kessler, Venezia
delmont@unive.it, pianta@itc.it

1 Introduction

In order to show that a system for text understanding has produced a sound representation of the semantic and pragmatic contents of a story, it should be able to answer questions about the participants and the events occurring in the story. This requires processing linguistic descriptions which are lexically expressed but also unexpressed ones, a task that, in our opinion, can only be accomplished starting from full-fledged semantic representations. The overall task of story understanding requires in addition computing appropriate coreference and cospecification for entities and events in what is usually referred to as a Discourse Model. All these tasks have been implemented in the GETARUNS system, which is subdivided into two main meta-modules or levels: the Low Level System, containing all modules that operate at sentence level; High Level System, containing all the modules that operate at discourse level by updating the Discourse Model. The system 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 phrase-structure rules in a given sequence as they are ordered in the grammar: whenever a syntactic constituent is successfully built, it is checked for semantic consistency, as LFG grammaticality principles require [1].

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

Proceedings of the 8th International Conference on Computational Semantics, pages 277–281,
Tilburg, January 2009.

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. 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 (non factual situations). 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.

2 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 (see [2]), in particular the cooperative principle. We 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 spatio-temporal 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 can be interpreted as bridging referential expression entertaining some semantic relation with previously mentioned entities. If we consider a classical example from [5] (A: *Can you tell me the time?*; B: *Well, the milkman has come*), we see that the request of the current time is bound to a spatio-temporal location. Using the MILKMAN rather than a WATCH to answer the question, is relatable to spatio-temporal 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 implied in the answer, that may trigger the inference.

2.1 The restaurant text

To exemplify some of the issues presented above we present a text by [7]. In this text, entities may be scenario-dependent characters or main characters independent thereof. Whereas the authors use the text for psychological experimental reasons, we will focus on its computability.

(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. (5) He began to read his book.

Sentence (1) introduces both JOHN as the Main Topic in the Topic Hierarchy and RESTAURANT as the Main Location (in the role of LOCATion argument of the governing verb GO and the preposition INTO). Sentence (2) can potentially introduce TABLE as new main Topic. This type of sentences is called *presentational* in the linguistic literature, 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 spatio-temporal component of the system. GETARUNS is equipped with a spatio-temporal inferential module that asserts Main Spatio-Temporal Locations to anchor events and facts expressed by situational infons. This happens whenever an explicit lexical location is present in the text, as in the first sentence (the RESTAURANT). The second sentence contains another explicit 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 triggers the implicature that the TABLE mentioned in sentence (2) is a metonymy for the SITting event. Consequently, the system will not assume that the indefinite expression *a table* has the funciton to present a new entity TABLE, but that an implicit entity is involved with a related event. The entity implied is understood as the Main Topic of the current Topic Hierarchy, i.e. JOHN.

We will now concentrate our attention onto sentence (3). To account for the fact that whenever a waiter takes an order there is always someone that makes the order, GETARUNS computes TAKE ORDER as a compound verb with an optional implicit 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 implicit entity. 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.

The text includes a 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 (5). 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, GETARUNS computes the Point of View and Discourse Domain on the basis of Informational Structure and Focus Topic by means of a Topic Hierarchy algorithm based on [3] and [8].

2.2 Common sense reasoning

GETARUNS is also able to search for unexpressed relations intervening in the current spatio-temporal location. To solve this problem in a principled way we needed commonsense knowledge organized in a computationally tractable way. This is what CONCEPTNET 2.1 ([6]) provides. 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 it includes instances of 19 semantic relations, 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 ([4]).

ConceptNet can be accessed in different ways; we wanted a strongly constrained one. We choose a list of relations from this external resource and combine them with the information available from the processing of the text to derive Implicit Information. In other words, we assume that what is being actually said hides additional information which however is implicitely hinted at. ConceptNet provides the following relations: SubEventOf, First-SubeventOf, DesiresEvent, Do, CapableOf, FunctionOf, UsedFor, EventRequiresObject, LocationOf. Let us see how this information can be exploited to interpret another classical example from the Pragmatics literature: 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 in this text, 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 link *petrol* and *garage* to *gas/gasoline* and *gas station* respectively. Now we can query the ontology 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. Eventually 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.

References

[1] Joan Bresnan. *Lexical-Functional Syntax (Blackwell Textbooks in Linguistics).* Blackwell Publisher, September 2000.

[2] H.P. Grice. Logic and conversation. In P. Cole and J.L. Morgan, editors, *Syntax and Semantics*, volume 3. New York Academic Press, 1975.

[3] B. Grosz. *Focusing and description in natural language dialogues.* Cambridge University Press, 1981.

[4] Douglas B. Lenat. CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38, 1995.

[5] Stephen Levinson. *Pragmatics.* Cambridge University Press, 1983.

[6] Hugo Liu and Push Singh. ConceptNet: A practical commonsense reasoning toolkit. *BT Technology Journal*, 22(211–226), 2004.

[7] A.J. Sanford and S.C. Garrod. Thematic subjecthood and cognitive constraints on discourse structure. *Journal of Pragmatics*, 12(5-6):519–534, 1988.

[8] C. Sidner. Focusing in the comprehension of definite anaphora. In M. Brady and R. Berwick, editors, *Computational models of discourse*, pages 267–330. MIT Press, 1983.