Procesamiento del Lenguaje Natural, núm. 35 (2005), pp. 449-450 recibido 18-05-2005; aceptado 01-06-2005

VENSES – a Linguistically-Based System for Semantic Evaluation

Rodolfo Delmonte
Ca' Garzoni-Moro, San Marco 3417
Università "Ca' Foscari"
30124 - VENEZIA
Tel. 39-041-2349464/52/19 –
E-mail: delmont@unive.it
website: project.cgm.unive.it

Resumen: The system for semantic evaluation VENSES (Venice Semantic Evaluation System) is organized as a pipeline of two subsystems: the first is a reduced version of GETARUN, our system for Text Understanding. The output of the system is a flat list of head-dependent structures (HDS) with Grammatical Relations (GRs) and Semantic Roles (SRs) labels. The evaluation system is made up of two main modules: the first is a sequence of linguistic rule-based subcalls; the second is a quantitatively based measurement of input structures. VENSES measures semantic similarity which may range from identical linguistic items, to synonymous or just morphologically derivable. Both modules go through General Consistency checks which are targeted to high level semantic attributes like presence of modality, negation, and opacity operators, temporal and spatial location checks. Results in cws, accuracy and precision are homogenoues for both training and test corpus and fare higher than 60%.

Palabras clave: Parsing, Dependency Structures, Logical Form, Semantic Evaluation

Time required for demonstration: 20 minutes

1 Introduction

We present our system for semantic evaluation which has obtained one of the best results in the RTE Challenge (Delmonte et al., 2005). The system for semantic evaluation VENSES (Venice Semantic Evaluation System) is organized as a pipeline of two subsystems: the first is a reduced version of GETARUN, our system for Text Understanding; the second is the semantic evaluator which was previously created for Summary and Question evaluation and has now been thoroughly revised for the new more comprehensive RTE task.

The reduced GETARUN is composed of the usual sequence of submodules common in Information Extraction systems, i.e. a tokenizer, a multiword and NE recognition module, a PoS tagger based on finite state automata; then a multilayered cascaded RTN-based parser which is equipped with an interpretation module that uses subcategorization information and semantic roles processing. Eventually, the system is equipped with a pronominal binding module that works for lexical personal, possessive and reflexive pronouns, which are substituted by the heads of their antecedents - if available. The output of the system is a flat list of head-dependent structures (HDS) with Grammatical Relations (GRs) and Semantic Roles (SRs) labels. Notable additions to the usual formalism is the presence of a distinguished Negation relation; we also mark modals and progressive mood. All other non semantic elements like auxiliaries and determiners are erased.

The evaluation system uses a strategy of rewards/penalties for T/H pairs where text entailment is interpreted in terms of semantic similarity: the closest the T/H pairs are in semantic terms, the more probable is their entailment. Rewards in terms of scores are assigned for each "similar" semantic element; penalties on the contrary can be expressed in terms of scores or they can determine a local failure and a consequent FALSE decision.

The evaluation system accesses the output of GETARUN which sits on files and is totally independent of it. It is made up of two main Modules: the first is a sequence of linguistic rule-based subcalls; the second is a quantitatively based measurement of input structures. The latter is basically a count of heads, dependents, GRs and SRs, scoring only similar elements in the H/T pair. Similarity may range from identical linguistic items, to synonymous or just morphologically derivable. As to GRs and SRs they are scored higher according to whether they belong to the subset of core relations and roles, i.e. obligatory arguments, or not, that is adjuncts. Both Modules go through General Consistency checks

Linguistic rule-based subcalls are organized into a sequence of calls going from rules containing axiomatic-like paraphrase HDSs which are ranked higher, to rules stating conditions for similarity according to the scale of argumentality which are ranked lower. All rules address HDSs, GRs and SRs. Both Modules strive for True assessments: however, Calls 1 are then followed by Calls 2 which can output True or False according to general consistency or scoring. Modifying the scoring function may thus vary the final result dramatically: it may contribute more True decisions if relaxed, so it needs fine tuning. More experimentation is ccurrently being carried out on a bigger data set the MRS made available by Microsoft, to achieve a more general definition of this function.

2 An A-As Hybrid Parser

Our parser has been presented in detail lately in a number of papers and has achieved 90% recall on Greval Corpus (see Delmonte 2004) and 89% recall on the XEROX-700 corpus, limited only this latter test to SUBJ/OBJ GRs. As in most robust parsers, we use a sequence or cascade of transducers: however, in our approach, since we intend to recover sentence level structure, the process goes from partial parses to full sentence parses. Sentence and then clause level parsing are crucially responsible for the right assignment of Arguments and Adjuncts (hence A-As) to a governing predicate head. This is paramount in our scheme which aims at recovering predicate-argument structures, besides performing a compositional semantic translation of each semantically headed constituent.

3 The Semantic Evaluator (SE)

As said above, the SE is organized into two main modules: a quantitatively based module, and a sequence of rule-based subcalls where scoring is also taken into account when needed, to increase confidence in the decision process. The two modules must then undergo general consistency checks which have the task to ascertain the presence of possible mismatches at semantic level. In particular, these checks take care of the following semantic items:

- presence of spatiotemporal locations relatively to the same governing predicate;
- presence of opacity operators like discourse markers for conditionality having scope over the governing predicate under analysis;
- presence of quantifiers and other referentiality related determiners attached to the same nominal head in the T/H pair under analysis;
- presence of antonyms in the T/H pair at the level of governing predicates;
- presence of predicates belonging to the class of "doubt" expressing verbs, governing the relevant predicate shared by the T/H pair.

Test-set Results	Training-set Results
cws: 0.6257 **accuracy: 0.5925** **precision: 0.6242** **recall: 0.4650** **f: 0.5330**	**cws: 0.6396** **accuracy: 0.6032** **precision: 0.6261** **recall: 0.5088** **f: 0.5614**

Tab.1 Results for training and test-set

4 References

Delmonte, R. (2004), Evaluating GETARUNS Parser with GREVAL Test Suite, Proc. ROMAND - 20th International Conference on Computational Linguistics - COLING, University of Geneva, 32-41.

Delmonte R., Sara Tonelli, Marco Aldo Piccolino Boniforti, Antonella Bristot, Emanuele Pianta (2005), VENSES – a Linguistically-Based System for Semantic Evaluation, RTE Challenge Workshop, Southampton, PASCAL - European Network of Excellence, pp. 49-52.