Noun/Verb Inference

Paul Bedaride

INRIA/LORIA
Université Henri Poincaré, Nancy
paul.bedaride@loria.fr

Claire Gardent

CNRS/LORIA
Nancy
claire.gardent@loria.fr

Abstract

We present a system which combines logical inference with a semantic calculus producing "normalised semantic representations" that are robust to surface differences which are irrelevant for entailment detection. We focus on the detection of entailment relations between sentence pairs involving noun/verb alternations and we show that the system correctly predicts a range of interactions between basic noun/verb predications and semantic phenomena such as quantification, negation and non factive contexts.

Keywords: Semantic Role Labelling, Textual Entailment, Noun/Verb Alternations

1. Introduction

As has been repeatedly argued, detecting whether a given sentence S_1 implies some other sentence S_2 is a basic semantic task that natural language understanding systems must be able to perform. Consequently, implication detection has been the focus of intense research in particular since the inception in 2005 of the RTE (Recognising textual entailment) challenge(Dagan et al., 2005).

In this paper, we focus on detecting implications between sentences involving a nominal/verbal alternation such as for instance, the following sentence pairs:

(1) a. Assuming no dramatic *fluctuation* in interest rates , the association expects to achieve near record earnings in 1990 .
→ If interest rates do not *fluctuate* dramatically , near record earnings are expected in 1990 .
→ Unless interest rates *fluctuate* dramatically, near record earnings are expected .

The approach we propose takes a middle path between the logical approach adopted by semanticists and the similarity based reasoning resorted to by many RTE systems (Dagan et al., 2005). As (MacCartney, 2009) has argued, while the first approach produces semantic representations that are too brittle to handle real text (for example, (Bos and Markert, 2006)'s system was able to find a proof for less than 4% of the problems in the RTE1 test set), the second fails to adequately handle commonplace semantic phenomena such as e.g., negation, quantification or non-factive contexts.

To overcome these shortcomings, we combine a logic based approach with a robust calculus of semantic representations in which joint syntactic/semantic structures produced by semantic role labelers (SRL) are rewritten into first order logic (FOL) formulae.

Because our semantic calculus has access to SRL structures, it provides an appropriate level of abstraction from syntactic differences that are irrelevant to entailment detection. For instance, it ensures that the semantic representations for *Rome was destroyed by the Barbarians*, *The Barbarians have destroyed Rome*, *Rome's destruction by the Barbarians* and *Barbarians destruction of Rome* are all identical[1].

[1] Note that although grammar based systems are in principle able to capture most of these equivalences, in practice they often fail to because the representations they produced closely reflects the input string and in particular, contains most of its function words. For instance, the representations associated by Boxer (Curran et al., 2007) with the four "*Rome* phrases" above are all different: the passive differs from the active in that it contains a *by* predication reflecting the use of the agent phrase; the nominal versions differ from the verbal in that that the nominal representation identifies a topic whilst the verbal one does not; and the two nominal versions differ in that they each contain either a *by* or an *of* predication depending on the function words they contain.

Furthermore, the use of a general rewrite system allows for abstractions and generalisations that are difficult to capture in a strictly compositional system à la Montague. For instance, it permits capturing both the restriction and the scope of a quantifier with a single rule applying to a non local fragment of the dependency tree namely, the fragment containing the determiner, its associated nominal and the verb phrase over which the quantifier scopes. More generally, the use of a general rewrite system on joint syntactic/semantic structures permits abstracting over surface details that are irrelevant to the detection of sentential implication.

The paper is structured as follows. We first describe the semantic role labeler we developed for verbs and nouns (Section 2). We then show how logic based representations are derived from the joint syntactic/semantic structures produced by this labeler and how they are used to recognize (non) sentential entailments (Section 3). Finally in section 4, we report on a first evaluation showing that the resulting system permits recognising some of the expected inferences.

2. Semantic Role Labelling

In order to recognise that *David Hiddleston was killed* can be inferred from *The avalanche killed David Hiddleston on the spot* , it must first be recognised that *X kill Y* entails *Y was killed*. Consequently, many of the the recent entries in the annual Recognizing Textual Entailment (RTE) competition have used rewriting in a variety of ways, though often without distinguishing it as a separate subproblem.

Based on this observation, we undertake to develop an entailment detection system in which rewriting is modelled using a standard rewriting system called GrGen (Geiß et al., 2006) which is at once efficient, notationally expressive and used in multiple domains (e.g., formal calculus, combinatoric algebra, operational semantics).

Our semantic role labeller uses rewrite rules to rewrite the dependency structures output by the Stanford statistical parser into joint syntactic/semantic structures where verbs and nouns predications are assigned identical thematic representations. In particular, the thematic structure of the four versions of *Rome was destroyed by the Barbarians* given above will be identical.

In previous work (Bedaride and Gardent, 2009), we presented the derivation process of the normalising rewrite rules for verbs and the resulting labeller was shown to achieve 72.6% F-mesure on the PropBank data, 86.3% precision on positive entailment detection on a benchmark of 4976 constructed examples and 99.28% precision for non entailment cases.

Here, we focus on SRL for nouns and describe how the syntax-to-semantic rewrite rules required to normalise the representation of nouns are derived from existing resources namely NomLex, NomBank and Comlex. The rules are automatically derived from these resources in three main steps. First, a subcategorisation lexicon for nouns is extracted from NomLexPlus by unfolding its factorised entries and filtering out linguistically illicit combinations. Second, the mapping between syntactic and semantic arguments is derived from ComLex and integrated in the subcategorisation lexicon derived in step 1. Third, lexicalised rewrite rules are derived from the resulting lexicon.

We now explain each of these steps in more detail.

Unfolding NomLexPlus. NomLexPlus is a subcategorisation lexicon for nouns built as an extension of NomLex and derived semi automatically from NomBanK. It contains 7 000 lexical entries, each entry encoding the subcategorisation frames of a nominal predicate. When the noun is a deverbal, the verb to which it is related is also indicated.

Figure 1 shows the NomLexPlus entry for the noun *articulation*. The `VERB-SUBC` field indicates the possible frames of that noun namely `NOM-NP` , `NOM-NP-PP` and `NOM-INTRANS`[2]. As can be seen, NomLexPlus entries are factorised in that each frame is associated with several possible realisations for each argument in that frame. For instance, the frame `NOM-NP` indicates that the `SUBJECT` argument can be realised as a `N-N-MOD`, a `DET-POSS` or a `PP :PVAL "by"` and similarly, that the `OBJECT` argument can be realised as a `N-N-MOD`, a `DET-POSS` or a `PP :PVAL "of"`.

To identify the possible syntactic configurations associated by NomLexPlus with each noun, we unfold each entry and associate with each noun the set of unfolded frames that can be thus derived. The unfolding consists in taking the product of each list of possible realisation for each argument and filtering out the result to eliminate illicit combinations such as for instance, frames with several arguments having the same realisation.

[2] The conventions used for naming frames are taken over from ComLex. Briefly, the frame name indicates the major category of each argument whereby the subject is systematically ommitted. Each argument is futhermore associated with a function and each function with a set of possible realisations.

```
(NOM :ORTH "articulation"
     :PLURAL *NONE*
     :VERB "articulate"
     :NOM-TYPE ((VERB-NOM))
     :VERB-SUBJ ((PP :PVAL ("by")))
     :VERB-SUBC ((NOM-NP :SUBJECT ((N-N-MOD)
                                   (DET-POSS)
                                   (PP :PVAL ("by")))
                         :OBJECT ((DET-POSS)
                                  (N-N-MOD)
                                  (PP :PVAL ("of")))
                         :REQUIRED ((OBJECT)))
                 (NOM-NP-PP :SUBJECT ((N-N-MOD)
                                      (DET-POSS)
                                      (PP :PVAL ("by")))
                            :OBJECT ((DET-POSS)
                                     (N-N-MOD)
                                     (PP :PVAL ("of")))
                            :PVAL ("with"))
                 (NOM-INTRANS :SUBJECT ((N-N-MOD)
                                        (DET-POSS)
                                        (PP :PVAL ("of" "by")))
                              :REQUIRED ((SUBJECT))))
```

Figure 1: NomLexPlus entry

Deriving the syntax to semantics mapping. ComLex specifies for each verb frame the mapping between semantic and syntactic arguments.We use this information to directly associate the nominal syntactic frame of our lexicon with thematic roles (rather than syntactic functions).

Specifying rewrite rules. Based on the subcategorisation lexicon extracted from NomLexPlus and the syntax to semantics mapping derived from ComLex, we specify a set of rewrite rules that maps the dependency graphs produced by the Stanford parser to dependency graphs enriched with thematic information.

First, we manually identify the mapping between the possible argument realisations (`DET-POSS, N-N-MOD` and `PP :PVAL`) and the Stanford parser dependency structures. The mapping is straigthfoward and involves roughly 20 rules.

Next, we define the rewrite rules by mapping the Stanford dependency graph corresponding to a possible frame to the thematic structure resulting from mapping each of the argument present in that frame to the thematic role defined for this argument in that frame by the syntax-to-semantic mapping defined in the lexicon produced in the previous step.

In this way, we derive a set of 453 general syntax to semantics rewrite rules. These general rules are furthermore associated with conditions which restrict their application depending on the specific noun and prepositions found in the input text. In other words, the rules are lexicalised but lexical conditions are only

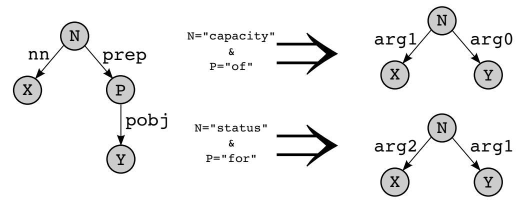

Figure 2: Rewrite rule example

checked once a general, non lexical rule has been found which matches the input. We found that roughly 75% of the rules have a narrow coverage in that they apply to less than 100 nouns whilst the remaining 25% are fairly general rules applying to many nouns.

3. Logic based reasoning

To capture the interaction between basic, nominal or verbal predications and standard semantic phenomena such as quantification, negation and non factive contexts, we rewrite the joint structures produced by the labeller presented in the previous section to a first-order logic (FOL) formula. We then use theorem proving on these formulae to detect (non) entailment relations between sentences.

From labelled dependency structures to FOL formulae. The translation of a LSD into a FOL formula is again performed using rewriting. Different rewriting rules will apply depending on the particular structures present in the input. Briefly, the set of rewrite rules used to handle the examples in the benchmark used (cf. section 4) can be sketched as follows.

A first set of rules deals with quantification and ensures both the proper translation of the determiner (e.g., that *every* rewrites to a universal) and the appropriate binding of both scope and restrictions (e.g., *every man runs* will be assigned the semantics $\forall x(Mx \rightarrow Rx)$).

A second set of rules ensures that each node in a labelled dependency structure (LSD, that is, the thematic grid associated with a noun or a verb) is associated with an existentially quantified variable and a predication over that variable where the predicate used is the word labelling the node. Furthermore, each edge translates to a binary relation between the source and the target node variables. The overall formula associated with an LSD is then the bracketed conjunction of the predications introduced by each node. For instance, the formula for *John loves Mary* will be $\exists e, y, z : (love(e) \wedge john(y) \wedge mary(z) \wedge arg0(e, y) \wedge arg1(e, z))$.

Finally, a third set of rules caters for the translation of negation and sentence connectives. These rules search the syntactic structure for connectors and negation words and rewrite these by combining the semantics of these words with that of their arguments. If necessary, the semantics of the arguments can be modified to account for the interaction with the context. For instance, an existential will be rewritten to a universal when it occurs in a universally quantifying context (e.g., *If a man owns a donkey, he feeds it*).

Checking entailment. Given the above translation into FOL, textual entailment between sentences can be tested by checking for logical entailment between the associated FOL formulae (Blackburn et al., 1999). In practice, we get formulas for the 5 first syntactic analyses and select the analysis with the highest semantic score where the scoring system favors longer predications (i.e., predications with a higher number of dependents) over shorter ones. We then check logical entailment between the two representations associated with the most highly scored analysis of each of the two sentences to be compared.

4. Evaluation

We evaluate the proposed approach in two ways. First, we examine its capacity as a SRL by computing recall and precision against the ConLL 2000 data. This gives a mesure of how good the system is at normalising verbal and nominal dependency structures. Second, we test the approach on entailment related sentence pairs where the (non) entailment is mediated through syntax based equivalence and in particular, nominalisations.

SRL evaluation. The evaluation against the ConLL data yields the results given in the following table:

	precision	recall	f-score
Unlabeled	83.56%	65.36%	73.35
Labeled	72.66%	56.83%	63.78

The unlabeled score gives the proportion of correct predicate/argument dependency found (for verbs and nouns) while the labeled score additionally checks the specific relation holding between predicate and argument. The overall score situate our labeller in the middle range of ConLL 2009 joint labellers (F1 ranging from 36.05 to 85.44) with a reasonably good precision but a low recall due partly to the fact that the Stanford parser often fails to return the correct analysis[3].

Entailment detection. To evaluate our approach on entailment detection, we manually built a benchmark of 20 sentence pairs involving a N/V variation. The benchmark encompasses 4 main types of entailment patterns dubbed respectively, *simple, light-verb, sem* and *adj/adv*. A *simple* pattern is one such as 2 where the entailment depends only on a nominalisation.

(2) Legislation to lift the debt ceiling is ensnarled in the fight over cutting capital-gains taxes.
$\rightarrow$ Capital-gains are taxed.

A *light-verb* pattern involves a light verb construction such as for example:

(3) a. An acceleration of investments gives Japanese companies control of large, highly visible U.S. corporations, such as Columbia Pictures Entertainment Inc.
$\rightarrow$ Japanese companies control U.S. corporations.
$\not\rightarrow$ An acceleration of investments controls U.S. corporations.

b. It is operating under Chapter 11 of the federal Bankruptcy Code, giving the company court protection from creditors'lawsuits.
$\rightarrow$ Chapter 11 of the federal Bankruptcy Code protects the company.
$\not\rightarrow$ the company court protectsfrom creditors'lawsuits

The *adj/adv* type illustrates the interaction between predication and modifiers:

(4) Countries with inadequate protections for intellectual-property rights could be hurting themselves.
$\rightarrow$ Countries which inadequately protect intellectual-property rights could be hurting themselves.

Finally, a *sem* pattern illustrates the interaction between basic predications and semantic phenomena

[3](Klein and Manning, 2003) report a label F-mesure of 86.3% on section 23 of the Penn Treebank.

such as quantification, negation and non factive contexts. For instance, (1) illustrates the interaction of *if* (verbalised by *assuming, if* or *unless*) negation (verbalised by *unless, no*) and N/V relations (*fluctuation/fluctuate*).

For each of the sentence pairs contained in the benchmark, the system correctly predicts the (non) entailment relation.

5. Conclusion

Although it remains limited in scope, the system presented here lays the basis for an approach to entailment detection that combines a robust semantic calculus with logical based reasoning. It thereby departs from (Bos and Markert, 2006) in that the semantic representations are less britle and from (MacCartney, 2009) in that it integrates both the role labelling abstraction of SRLs and logical rather than natural logic reasoning. We have illustrated the potential of the approach by showing how it could handle a limited range of interaction between nominal predication, verbal predication and logical connnectives. Current work concentrates on extending the system coverage and on evaluating it on a full size benchmark designed to illustrate a wider range of interaction between basic predications and the various semantic phenomena potentially present in their sentential context.

References

Bedaride, P. and Gardent, C. (2009), Semantic normalisation : a framework and an experiment., *in* 'IWCS'09: 8th International Conference on Computational Semantics', Tilburg, Netherland.

Blackburn, P., Bos, J., Kohlhase, M. and de Nivelle, H. (1999), Inference and computational semantics, *in* H. .Bunt and E.Thijsse, eds, 'IWCS-3', pp. 5–19.

Bos, J. and Markert, K. (2006), When logical inference helps determining textual entailment (and when it doesn't), *in* 'Pascal, Proceedings of the Second Challenge Workshop, Recognizing Textual Entailment'.

Curran, J. R., Clark, S. and Bos, J. (2007), Linguistically motivated large-scale nlp with c&c and boxer, *in* 'ACL-07 demo', pp. 29–32.

Dagan, I., Glickman, O. and Magnini, B. (2005), The PASCAL Recognising Textual Entailment Challenge., *in* 'Machine Learning Challenges Workshop', pp. 177–190.

Geiß, R., Batz, G., Grund, D., Hack, S. and Szalkowski, A. (2006), Grgen: A fast spo-based graph rewriting tool, *in* 'ICGT 2006', Heidelberg, p. 383–397.

Klein, D. and Manning, C. D. (2003), Accurate unlexicalized parsing, *in* 'ACL', Sapporo, Japan, pp. 423–430.
URL: *http://www.aclweb.org/anthology/P03-1054*

MacCartney, B. (2009), Natural language inference, PhD thesis, Stanford University.