ASdeCopas: a syntactic-semantic interface*

Luísa Coheur[1,2], Nuno Mamede[1], and Gabriel G. Bès[2]

[1] L^2F INESC-ID/IST – Spoken Languages Systems Laboratory
{luisa.coheur}{numo.mamede}@l2f.inesc-id.pt
[2] GRIL/Université Blaise-Pascal
Gabriel.Bes@univ-bpclermont.fr

Abstract. ASdeCopas is a syntactic-semantic parser, implemented in Prolog, which uses hierarchically organized order-independent rules. This paper focuses on the formalization of semantic rules, presenting the concepts of well-formed semantic rule, rules hierarchy, and the conditions for rules application. If two rules can apply, only the most specific one does so. Examples are given and some properties of the system are pointed out.

Keywords: syntactic-semantic interface, semantic rules, rules hierarchy, 5P, Minimal Recursion Semantics

1 Introduction

ASdeCopas[3] is a syntactic-semantic parser that takes a graph representing the input sentence and returns a formula, according with a set of semantic rules. The paper focuses on the formalization of semantic rules. Section 2 describes ASdeCopas's input, semantic rules are formalized in section 3, section 4 presents simplified examples and section 5 lists some of the system properties and discusses perspectives on future work.

2 ASdeCopas's Input

Ideas, formalisms and data from the 5P paradigm $[1, 3, 6, 2]$ are followed/used to obtain ASdeCopas' input: a text with an associated graph. A graph is defined as follows (let C be a set of category labels, W a set of words and _ the empty field):

Definition 1. *Graph*
A graph is a pair $G = (\Delta, \Psi)$*, where:*
• Δ *is a set of nodes, each one noted node(w, c, p), where* $w \in W$*,* $c \in C$ *and* $p \in \mathbb{N}$ *(p represents the node's position).*
• Ψ *is a set of arrows, each arrow noted arrow(*p_1*,* p_2*), where* $p_1, p_2 \in \mathbb{N}$ *(*p_1*,* p_2 *being the position of the source and the target, respectively).*

* Paper supported by FCT (Fundação para a Ciência e Tecnologia).
[3] ASdeCopas stands for "Análise Semântica depois de Completada a análise sintáctica".

A well formed graph (wfg) verifies:

• $(\forall\ arrow(p_1,\ p_2) \in \Psi)(\exists\ node(w_1,\ c_2,\ ps_1),\ node(w_2,\ c_2,\ ps_2) \in \Delta)(p_1 = ps_1 \wedge p_2 = ps_2)$ *(that is, each arrow connects existing nodes)*

• $(\forall\ arrow(p_1,\ p_2),\ arrow(p_3,\ p_4) \in \Psi)(p_1 = p_3 \Rightarrow p_2 = p_4)$ *(that is, each node is the source of at most one arrow)*

• $(\forall\ node(w_1,\ c_1,\ p_1),\ node(w_2,\ c_2,\ p_2),\ node(w_3,\ c_3,\ p_3),\ node(w_4,\ c_4,\ p_4) \in \Delta)[(p_1 < p_2 < p_3 < p_4 \wedge (\exists\ arrow(p_1,\ p_3) \vee \exists\ arrow(p_3,\ p_1) \in \Psi)) \Rightarrow \neg(\exists\ arrow(p_2,\ p_4) \vee \exists\ arrow(p_4,\ p_2) \in \Psi)]$ *(that is, no crossing of arrows is allowed)*

Each category is a set of attribute/value pairs[4], *i.e.*, feature structures hierarchically organized (see [1] for details). Notice that no constraint is set on the nature of those pairs: they can have a syntactic or a semantic motivation. Arrows are somehow related to dependencies, but, contrary to mainstream dependency theories, arrows go from dependents to the head [6]. Their motivation is simply to connect two elements, because the established relations are needed to reach the desired semantic representation (see [1, 3] for extra details about this concept). An example of a graph is shown in the next figure, where *A pequena Maria* means *Little Mary*:

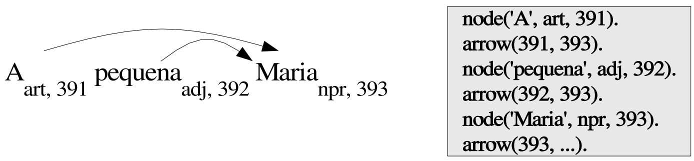

Fig. 1. Graph example.

3 Semantic rules

Definitions 2 to 8 define the syntax of a well-formed semantic rule, while Definitions 9 to 11 define rules hierarchy and Definitions 12 and 13 rules applicability conditions.

Definition 2. *Element (to transform)*
An element has the form $e = elem(w,\ c)$*, where* $w \in \{\_\} \cup W$ *and* $c \in \{\_\} \cup C$.

Definition 3. *Arc*
An arc has the form $a = arc(c_1,\ c_2,\ d)$*, where:*

• $c_1,\ c_2 \in C$ *(*c_1 *and* c_2 *are, respectively, the source and the target of the arc);*

• $d \in \{\_\} \cup \{L,\ R\}$ *(L when the arc goes from right to left, R from left to right).*

[4] For expository reasons we use a unique label to identify those sets.

Definition 4. *Semantic Rule*
A semantic rule is a triple $R_i = (\Sigma, \Theta, \Gamma)$ *(also notated* $[R_i]\ \Sigma : \Theta \looparrowright \Gamma$*), where:*
- Σ *is a (non empty) set of elements;*
- Θ *is a (possibly empty) set of arcs;*
- Γ *is a set of translating functions.*[5]

For the following definitions, let $R_i = (\Sigma, \Theta, \Gamma)$ be a semantic rule.

Definition 5. *Connections between elements*
Let $e_1 = elem(w_1, c_1)$, $e_2 = elem(w_2, c_2) \in \Sigma$.
e_1 *and* e_2 *are said to be directly connected* $(e_1 \bowtie_e e_2)$ *if and only if (iff) there is an arc* $a = arc(c_3, c_4, d) \in \Theta$*, such that*
- $(c_1 = c_3 \wedge c_2 = c_4) \vee (c_1 = c_4 \wedge c_2 = c_3)$

e_1 *and* e_2 *are said to be connected* $(e_1 \bowtie_e^* e_2)$ *iff*
- $(e_1 \bowtie_e e_2) \vee (\exists\ e_3 \in \Sigma)(e_1 \bowtie_e e_3 \wedge e_3 \bowtie_e^* e_2)$

Definition 6. *Connections between arcs*
Let $a_1 = arc(c_1, c_2, d_1)$, $a_2 = arc(c_3, c_4, d_2) \in \Theta$ *and* $a_1 \neq a_2$.
a_1 *and* a_2 *are said to be directly connected* $(a_1 \bowtie_a a_2)$ *iff*
- $c_1 = c_4 \vee c_2 = c_3 \vee c_2 = c_4$

a_1 *and* a_2 *are said to be connected* $(a_1 \bowtie_a^* a_2)$ *iff*
- $(a_1 \bowtie_s a_2) \vee (\exists\ a_3 \in \Theta)(a_1 \bowtie_a a_3 \wedge a_3 \bowtie_a^* a_2)$

Definition 7. *Relation arc/element*
Let $a_1 = arc(c_1, c_2, d_1) \in \Theta$ *and* $e = elem(w, c) \in \Sigma$.
a_1 *and* e *are said to be directly connected* $(a_1 \twoheadrightarrow e)$ *iff*
- $c = c_1 \vee c = c_2$

a_1 *and* e *are said to be connected* $(a_1 \twoheadrightarrow^* e)$ *iff*
- $(a_1 \twoheadrightarrow e) \vee (\exists\ a_2 \in \Theta)(a_1 \bowtie_a^* a_2 \wedge a_2 \twoheadrightarrow e)$

Definition 8. *Well formed semantic rule*
$R_i = (\Sigma, \Theta, \Gamma)$ *is a well formed semantic rule (wfsr) iff:*
- $(\forall\ e_1, e_2 \in \Sigma)(e_1 \bowtie_e^* e_2)$*(all the elements are connected)*
- $(\forall\ a_1 \in \Theta)(\exists\ e \in \Sigma)(a_1 \twoheadrightarrow^* e)$*(all the arcs are related with an element)*

Example 1. Well formed semantic rule
The following rule, whose target is Minimal Recursion Semantics (MRS) [5], is a wfsr (`n` is the label of the category associated with nouns):

$[R_1]$ {`elem(_,n,_)`} : $\emptyset \looparrowright$ {`handle(n) : sem(n)(var(n))`}

For the following definitions let $\sqsubseteq$ be the subsumption relation between two sets.

Definition 9. *Element subsumption*
$e_1 = elem(w_1, c_1)$ *subsumes* $e_2 = elem(w_2, c_2)(e_1 \sqsubseteq_e e_2)$ *iff*
$(c_1 \sqsubseteq c_2) \wedge (w_1 \neq \_ \Rightarrow w_2 = w_1)$

[5] For expository reasons we will not detail translating functions. However, consider defined sem(c), var(c) and handle(c), returning, respectively, the semantics, a variable and an handle [5] associated with the element identified by category c.

Definition 10. *Arc subsumption*
$a_1 = arc(c_1, c_2, d_1)$ subsumes $a_2 = arc(c_3, c_4, d_2)$ ($a_1 \sqsubseteq_a a_2$) iff
$(c_1 \sqsubseteq c_3 \wedge c_2 \sqsubseteq c_4) \wedge (d_1 \neq \_ \Rightarrow d_1 = d_2)$

Definition 11. *Rule subsumption*
Let $R_1 = (\Sigma_1, \Theta_1, \Gamma_1)$, $R_2 = (\Sigma_2, \Theta_2, \Gamma_2)$ be wfsr.
R_1 subsumes R_2 ($R_1 \sqsubseteq_r R_2$) iff
$(\forall\, e_1 \in \Sigma_1)(\exists\, e_2 \in \Sigma_2)\ (e_1 \sqsubseteq_e e_2) \wedge (\forall\, a_1 \in \Theta_1)(\exists\, a_2 \in \Theta_2)(a_1 \sqsubseteq_a a_2)$

Example 2. Rule subsumption
As $\mathtt{n} \sqsubseteq \mathtt{np}$, rule R_1 subsumes the following rule:
$[R_2]$ {`elem(_,np,_)`} : $\emptyset \looparrowright$ {`handle(np) : NAME(var(np), sem(np))`}

Definition 12. *Conditions for the application of a semantic rule*
Let $R_j = (\Sigma, \Theta, \Gamma)$ be a wfsr and $G = (\Delta, \Psi)$ a wfg. R_j can apply to G iff:
- *$(\forall\ elem(w_i, c_i) \in \Sigma)(\exists\ node(w_j, c_j, p_j) \in \Delta)[(w_i \neq \_ \Rightarrow w_i = w_j) \wedge c_i \sqsubseteq c_j]$*
- *$(\forall\ arc(c_n, c_m, d) \in \Theta)(\exists\ arrow(p_k, p_l) \in \Psi, \exists\ node(w_k, c_k, p_k), node(w_l, c_l, p_l) \in \Delta)(c_n \sqsubseteq c_k \wedge c_m \sqsubseteq c_l \wedge (d = R \Rightarrow p_k < p_l) \wedge (d = L \Rightarrow p_k > p_l))$*

Definition 13. *Application of a semantic rule*
Being given a wfg, let R_1 and R_2 be wfsr, verifying the conditions to be applied to it. If $R_1 \sqsubseteq_r R_2$, then R_1 is not applied.

Example 3. Semantic rules applicability
Both rules R_1 and R_2 are in conditions to be applied to the graph from Fig. 1. As $R_1 \sqsubseteq_r R_2$ only R_2 is triggered.

4 Example

If rule R_1 is applied to the graph from Fig. 1, the following formula is obtained: $\mathtt{h}_{393}$ `: Maria(`$\mathtt{x}_{393}$`)` (notice that variable generation is not carried out randomly: variable indexes are given by the position of the associated element or of the element it arrows). However, as only R_2 is triggered, we obtain: $\mathtt{h}_{393}$ `: NAME(`$\mathtt{x}_{393}$`, Maria)`.
Next we present a rule for intersective and subsective adjectives [4] (we will use the notation from [7]).[6]
$[R_3]$ {`elem(_,adj,_)`} : {`arc(adj, n)`} $\looparrowright$ {`handle(n) : AM(var(n), sem(adj))`}
After adding this rule to the system, the following formula is generated is addition to the previous: $\mathtt{h}_{393}$`: AM(`$\mathtt{x}_{393}$`, pequena)`.

5 Brief discussion and future work

Briefly, ASdeCopas has the following properties:

[6] We should point out that reification of variables [8] over adjectives will be needed to allow modification. However, we will ignore this problem for expository reasons.

– propagation of ambiguous values can be precluded through the use of (syntactic) information presented in the rules;
– rules can be applied in any order, as they are intrinsically independent (that is, their output does not depend on the output of other rule);
– information can be modularly added by profiting from the hierarchical organization of both categories and semantic rules;
– partial results can be produced;
– different semantic processes (such as role extraction, anaphora resolution, ...) can run at different times and their final results merged, due to the controlled generation of variables.[7]

The main problem with this system is that the immediate production of a structured formula is not easy, even though equivalent "flatter" structures can be easily produced. Moreover, the production of partial results should be taken carefully, as inconsistent representations can be produced.
ASdeCopas is implemented in Prolog and is being tested in question interpretation and in a more formal framework where the output is Minimal Recursion Semantics.

References

1. Gabriel G. Bès. La phrase verbal noyau en français. In *Recherches sur le français parlé, 15*, pages 273–358. Université de Provence, France, 1999.
2. Gabriel G. Bès. Empiricité en linguistique et grammaire de montague: la sémantique en 5P et la compositionnalité (avec en annexe grammaire de montague et ambiguïté). Technical report, GRIL, Université Blaise-Pascal, Clermont-Ferrand, France, Avril, 2001.
3. Gabriel G. Bès and Caroline Hagège. Properties in 5P (soon in the GRIL web page). Technical report, GRIL, Clermont-Ferrand, France, November, 2001.
4. Gennaro Chierchia and Sally McConnell-Ginet. *Meaning and Grammar - an Introduction to Semantics (second edition).* The MIT Press, 2000.
5. Ann Copestake, Dan Flickinger, and Ivan A. Sag. *Minimal Recursion Semantics. An introduction. CSLI, Stanford University.* 1997.
6. Caroline Hagège. *Analyse Syntatic Automatique du Portugais.* PhD thesis, Université Blaise Pascal, Clermont-Ferrand, France, 2000.
7. Daniel Jurafsky and James Martin. *Speech and Language Processing.* Prentice Hall, 2000.
8. Diego Mollá. Ontologically promiscuous flat logical forms for NLP. In *IWCS-4, Tilburg, The Netherlands*, 2000.

[7] We are not assuming DRT to solve anaphora.