

What signs for the semantics-syntax interface?

Sylvain Kahane

Abstract

We propose a compositional analysis of the semantics-syntax interface where the sentence is decomposed in lexical and grammatical signs. The correspondence between a semantic representation (= a predicate-argument structure) and a syntactic representation (= a surface syntactic dependency tree) of a sentence directly results from the combination of the structures of the structural description of the signs composing the sentence.

Our purpose is to address what seems for us to be the first question to answer when we want to develop a grammar for the semantics-syntax interface: what are the units of the description? Syntactic models consider that the units of the description are words. It is justified: words are the units of the sentence when we consider it from the viewpoint of the linear order (= “word order”). But it is well known that there is no simple correspondence between words and semantics units due to morphology and idioms. Moreover prosody and some word order configurations like topicalization are meaningful.

The units of our semantics-syntax interface are the linguistics signs, that is the non-decomposable parts of a compositional analysis. The correspondence between the semantic and syntactic representations is obtained by the combination of the signs, but a sign itself cannot be decomposed and obtained by the combination of two parts. In the sentence *Beans are red*, we have three lexical signs BEAN, BE and RED and five grammatical signs: plural and indefinite combining with the noun and active, indicative and present combining with the verb. Note that BE, active and indicative are only “syntactic” signs without semantic contribution: BE simply allows the predication of the adjective RED to form a sentence with a finite verb and active and indicative are the voice and the mood by default.

We associate to each sign an elementary structure. Elementary structures combine by

unification given the semantics-syntax interface. We will begin our presentation of the formalism with the sign associated to the French verb MANGER ‘eat’. The structure is proposed in XML and in two graphical styles (Fig. 1 and 2).

```
<rule><title> MANGER </title>
<tree>
  <node lex="MANGER" p_lex=+ cat="V"
        p_voice=- p_mood=->
    <sem cont="eat">
      <arg n="1" id="[x]"/>
      <arg n="2" id="[y]"/>
    </sem>
  </node>
  <tree fct="subj">
    <node cat="N" p_lex=->
      <sem id="[x]"/>
    </node>
  </tree>
  <tree fct="dobj" opt=+>
    <node cat="N" p_lex=->
      <sem id="[y]"/>
    </node>
  </tree>
</tree>
</rule>
```

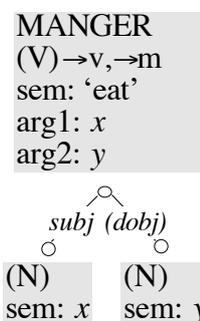


Figure 1 : Rule for MANGER ‘eat’ (first graphical representation)

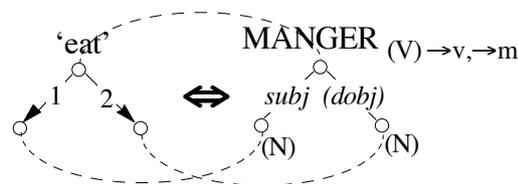


Figure 2 : Rule for MANGER ‘eat’ (second graphical representation)

The structure of MANGER indicates the correspondence between the predicate ‘eat’(x,y) and the lexeme MANGER. The semantic arguments x and y respectively

correspond to the subject (*subj*) and the direct object (*obj*) of MANGER, which must be nouns. The direct object is optional, as indicated by the feature-value $\text{opt}=+$ (and the brackets in the graphical representations). Our conventions favor the syntactic structure, which is encoded by the features $\langle \text{tree} \rangle$ for the dependencies (or, equivalently, for the subtrees) and the features $\langle \text{node} \rangle$ for the nodes. The semantic representation is encoded by the features $\langle \text{sem} \rangle$ for the nodes and $\langle \text{arg} \rangle$ for the predicate-argument relations. The correspondences between the syntactic and semantic nodes are indicated by the embedding of the $\langle \text{sem} \rangle$ into the $\langle \text{node} \rangle$. The features $/\text{p_voice}/$ and $/\text{p_mood}/$ are polar features taking the values $+$ or $-$: a value $-$ indicates a need, while a value $+$ indicates a resource ($+$ and $-$ can combine but not two $+$ or two $-$) (see Bonfante et al. to appear for an efficient use of polar features in parsing). The features $\text{p_voice}=-$ and $\text{p_mood}=-$ thus indicate the verbal lexeme MANGER must be combined with a voice and a mood. The features $/\text{id}/$ allow us to indicate a feature sharing; for instance, here, the first argument of ‘eat’ corresponds to the semantic content of the subject of MANGER.

The structure proposed for MANGER can be represented graphically. Fig. 1 favors the syntactic structure and Fig. 2, using Meaning-Text convention (Mel’cuk 1988), brings to the fore the correspondences between the semantic and syntactic representation. A need $\text{p_feature}=-$ is represented by $\rightarrow \text{feature}$.

Favoring the syntactic structure gives us a structure very similar to the ones of HPSG: The feature $\langle \text{node} \rangle$ corresponds to the feature HEAD and the features $\langle \text{tree} \rangle$ of a same level, to the subcategorization list (SUBCAT). Nevertheless, our structure also plays the role of the *head-daughter-phrase* schema allowing the combination of a head and its syntactic arguments: In this case, it is possible to interpret the feature $\langle \text{feature HD-DTR (head-daughter)} \rangle$ and the feature $\langle \text{tree} \rangle$ as a feature NHD-DTR (non-head-daughter).

The treatment of raising verbs (such as SEMBLER ‘seem’) and control verbs (such as VOULOIR ‘want’) forces us to enrich

the formalism. We must indicate, in the structure of such a verb, that its subject is also the potential subject of its verbal complement. In order to ensure that, we introduce, in the structure of these verbs, a particular dependency, noted $\langle \text{quasi} \rangle$ for *quasi-dependency* and represented by a dashed arrow. A quasi-dependency unifies with an ordinary dependency (of same function) giving a quasi-dependency.

We give the structure of SEMBLER ‘seem’ in Fig. 3. This verb corresponds to a unary predicate, which does not take its syntactic subject as semantic argument. The subject of SEMBLER is entirely controlled by its verbal complement, which is in the infinitive mood and will lose its subject by unification of its *subj* dependency with the *subj* quasi-dependency of the structure of SEMBLER. On the other hand, the verbal complement of SEMBLER must receive a voice. In the graphical representation, the sequence $\text{p_feature}=+$, $\text{feature}=\text{"value"}$ is abbreviated by $\text{feature}\rightarrow \text{value}$.

As seen before, a verb must combine with a mood and eventually a voice. We consider the following moods for French: indicative, subjunctive, imperative, infinitive, past participle and present participle. Our category of mood mixes two inflectional categories: the finiteness (finite, infinitive, participle) and the mood proper when the verb is finite. Fig. 3 gives a graphical representation of, and present. The mood indicative, as other finite moods, requires the combination with a tense, using a polar feature $/\text{p_tense}/$. The present tense associates the semantic predicate ‘present’ with the grammeme present. This predicate appears as the value of a feature $/\text{sem_tense}/$. By convention, the argument of a $/\text{sem_feature}/$ is the value of the $/\text{sem}/$ feature of the feature structure it belongs.

We can now give an example of a combination of structures for a whole sentence (Fig. 3). The resulting structure is well formed if every polar features have been neutralized. (The $\text{p_lex}=+$ of the syntactic root is neutralized by a $\text{p_lex}=-$ introduced by indicative.)

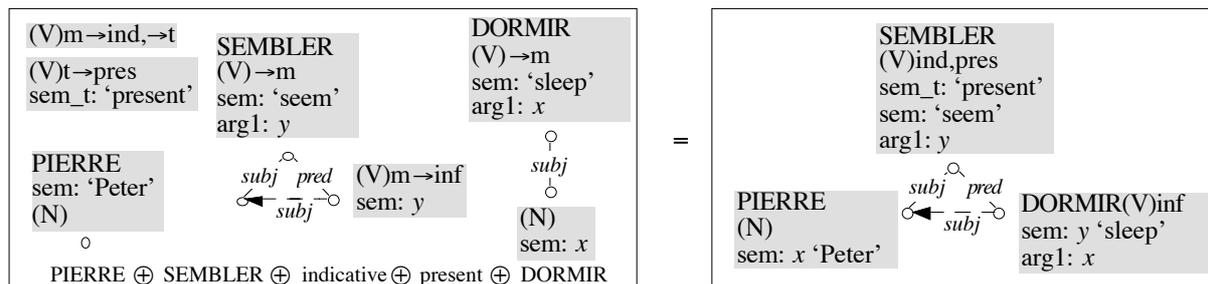


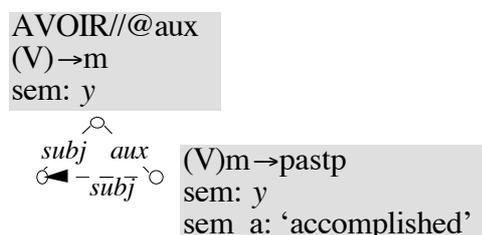
Figure 3: Combination of signs for *Pierre semble dormir* ‘Peter seems to sleep’

The same grammar could be formalized in HPSG provided that we introduce a specific schema for the combination of a lexical sign and a grammatical sign (which will be the head). It results that we will have as many phrasal projections for a verbal construction as grammatical signs that will combine with the verb. The resulting phrase structure will be closer to the syntactic structures of generative grammars or RRG (Van Valin & La Polla 1997) than usual phrase structure in HPSG (Pollard & Sag 1994).

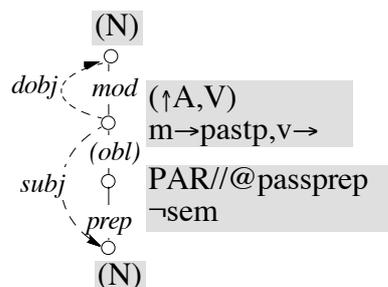
For the modeling of verbal compound forms such as *il semble avoir dormi* ‘he seems to have slept’, we introduce a grammatical sign we call accomplished, expressed by the auxiliary AVOIR (or the auxiliary BE if the feature /aux/ is instantiated). The structure of this sign ensures that the potential subject of the verb will be realized on the auxiliary. The auxiliary does not have its own content and receives the semantic content of the participle. The content of accomplish is a unary predicate ‘accomplished’ given as the value of a sem_feature we call /sem_a/.¹ Finally, as the auxiliary requires a mood but not a voice, the accomplish will take place between the mood and the voice (the voice is the most closer morpheme from the verb). The accomplished is compatible with all the moods. It tends to forms “idioms” with the simple tenses of the indicative, like the *passé composé*, syntactically equivalent

¹ Another possibility will be to introduce a real semantic vertex for the accomplish that would be associated the syntactic vertex of the auxiliary. Nevertheless, we need to ensure that the syntactic governor of the auxiliary takes the content of the verb as semantic argument and not the accomplish meaning.

to a present perfect but semantically equivalent to English simple past.



The structure for the active voice is trivial: It simply provides a feature-value p_voice=+. The treatment of the passive voice is more complex. We defend a compositional analysis of the passive where both past participle and copula are signs. We consider that in *Le poulet est mangé par Pierre* ‘the chicken is eaten by Peter’, the verb ÊTRE is the copula as in *La question est facile* ‘the question is easy’ and that the past participle mangé ‘eaten’ is the same element as in *le poulet mangé par Pierre* ‘the chicken eaten by Peter’ or *Le poulet semble mangé par Pierre* ‘the chicken seems (to be) eaten by Peter’ (cf. Abeillé & Godard 2002 for a more detailed argumentation). Consequently, we introduce a structure directly for the passive past participle; this structure can combine with the copula to form the passive.



In the passive past participle structure, the direct object becomes the nominal governor and the subject becomes an optional oblique

complement introduced by the preposition PAR ‘by’ (unless the feature /passprep/ specifies another value, that is, DE ‘of’ as in *Pierre est aimé de Marie* ‘Peter is loved by Mary’).

We can note that the syntactic roles on the dependencies play an important role in our grammar. For instance the *doj* relation validates the combination of a verb with the passive voice. It will also validate the combination with the accusative clitic pronoun or the agreement with past participle.

Conclusion

Our formalism can be compared with categorial grammars or TAG, but rather than considering words as units, it associates elementary structures to both lexical and grammatical signs. Some constructions like clefting, topicalization or dislocation (*Pierre, je le connais* Peter, I know it) will also be associated to signs (cf. for instance Kahane 2003 for a description of *tough*-movement in this formalism).

The formalism we propose for the semantics-syntax interface has several advantages: 1) it is powerful enough for allowing us to use a same structure for the very different uses of a lexeme (while a formalism such as requires a different structure for each use of a lexeme); 2) it avoids us to recourse to phrasal descriptions or schemata and is thus lighter than a formalism such as HPSG; moreover, this makes the formalism nearly associative, in the sense that the signs can be combined in whatever order (except for the grammatical signs) and the combination is not constrained by the need to saturate the phrases before a combination as in HPSG.

One of the qualities of the formalism comes from the use of quasi-dependencies, which allows the derivation structure not to be a tree (the derivation structure is the structure that describes which sign combines with which other; see Vijay-Shanker 1987 for TAG and Kahane 2003 for this formalism). For instance, in *Peter seems sick*, the three lexical signs PETER, SEEM and SICK combine pairwise, and the derivation structure contains a cycle.

References

- Abeillé Anne, Godard Danièle, 2002, “Course on the Syntax of Romance Languages”, *ESSLLI*, Bucarest.
- Bonfante Guillaume, Guillaume Bruno, Perrier Guy, to appear, “Analyse syntaxique électrostatique”, *TAL*.
- Kahane Sylvain, 2003, “On the status of the deep syntactic structure”, *MTT 2003*, Paris.
- Mel'cuk Igor, 1988, *Dependency Syntax: Theory and Practice*, State Univ. of New York Press, Albany.
- Pollard Carl, Sag Ivan, 1994, *Head-Driven Phrase Structure Grammar*, Stanford CSLI.
- Tesnière Lucien, 1959, *Eléments de syntaxe structurale*, Klincksieck, Paris.
- Vijay-Shanker K., 1987, *A Study of Tree Adjoining Grammars*, PhD thesis, University of Pennsylvania.