Syntactico-Semantic Learning of Categorial Grammars

Isabelle Tellier
LIFL and Université Charles de Gaulle-lille3 (UFR IDIST)
59 653 Villeneuve d'Ascq Cedex, FRANCE
Tel : 03-20-41-61-78 ; fax : 03-20-41-61-71

tellier@univ-lille3.fr

1. Introduction

Natural language learning seems, from a formal point of view, an enigma. As a matter of fact, every human being, given nearly exclusively positive examples (as psycholinguists have noticed) is able at the age of about five to master his/her mother tongue. Though, no linguistically interesting class of formal languages is learnable with positive data in usual models (Gold's (67) and Valiant's (84)).

To solve this paradox, various solutions have been proposed. Following the chomskian intuitions (Chomsky 65, 68), it can be admitted that natural languages belong to a restricted family and that the human mind includes an *innate knowing* of the structure of this class (Shinohara 90). Another approach consists in putting structural, statistical or complexity constraints on the *examples proposed to the learner*, making his/her inferences easier (Sakakibara 92).

A particular family of research, more concerned with the cognitive relevance of its models, considers that in « natural » situations, examples are always provided with semantic and pragmatic information and tries to make profit of it (Anderson 77; Hamburger & Wexler 75 ; Hill 83 ; Langley 82). This is the family our research belongs to.

But the property of meaningfulness of natural languages is computationally tractable only if we have at our disposal a theory that precisely articulates syntax and semantics. The strongest possible articulation is known as the Fredge's *principle of compositionality*. This principle has acquired an explicit formulation with the works of Richard Montague (Dowty, Wall & Peters 81 ; Montague 74) and his inheritors.

We will first briefly recall an adapted version of this syntactico-semantic framework, based on a type of grammars called « classical categorial grammars » (or CCGs), and we will then show how it can been used in a formal theory of natural language learning.

2. Syntactic analysis with CCGs

A categorial grammar G is a 4-tuple G=<V, C, f, S> with :

- V is the finite *alphabet* (or vocabulary) of G ;
- C is the finite set of *basic categories* of G ;
  From C, we define the set of *all possible categories* of G, noted C', as the closure of C for the operators / and \. C' is the smallest set of categories verifying :
  - $C \subseteq C'$;
  - if $X \in C'$ and $Y \in C'$ then : $X/Y \in C'$ and $Y\backslash X \in C'$;
- f is a *function* : $V \longrightarrow P_f(C')$ where $P_f(C')$ is the set of finite subsets of C', which associates each element v in V with the finite set $f(v) \subseteq C'$ of its categories ;
- $S \in C$ is the *axiomatic category* of G.

In this framework, the set of syntactically correct sentences is the set of finite concatenations of elements of the vocabulary for which there exists an affectation of categories that can be « reduced » to the axiomatic category S. In CCGs, the admitted reduction rules for any categories X and Y in C' are :

- R1 : $X/Y \,.\, Y \longrightarrow X$
- R'1 : $Y \,.\, Y\backslash X \longrightarrow X$

The language L(G) defined by G is then :

$$L(G)=\{w \in V^*;\ \exists n \in N\ \ \forall i \in \{1,..., n\}\, w_i \in V,\ w=w_1...w_n \text{ and } \exists C_i \in f(w_i),\ C_1...C_n \overset{*}{\longrightarrow} S\}.$$

The class of languages defined by CCGs is the class of context-free languages (Bar Hillel, Gaifman & Shamir 60). CCGs are lexically oriented because grammatical information is entirely supported by the categories associated with each word. They are also well adapted to natural languages (Oehrle, Bach & Wheeler 88).

Example :

Let us define a CCG for the analysis of a small subset of natural language, including the vocabulary V={a, every, man, John, Paul, runs, is, ...}. The set of basic categories is C={S, T, CN} where T stands for « terms » and is affected to proper names, CN means « common nouns », intransitive verbs receive the category T\S, transitive ones : (T\S)/T and determiners : (S/(T\S))/CN. Figures 1 and 2 display analysis trees.

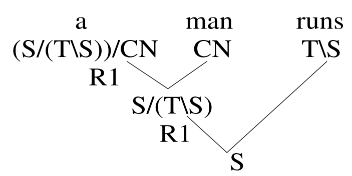

figure 1 : analysis tree n° 1

Isabelle Tellier (1998) **Syntactico-Semantic Learning of Categorical Grammars.** In D.M.W. Powers (ed.) *NeMLaP3/CoNLL98 Workshop on Paradigms and Grounding in Language Learning,* ACL, pp 311-314.

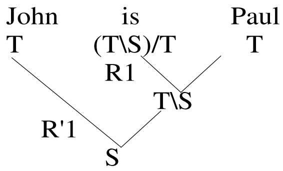

figure 2 : analysis tree n°2

3. From syntax to semantics

The key idea of Montague's work (74) was to define an isomorphism between syntactic trees and semantic ones. This definition is the formal expression of the principle of compositionality. It allows to automatically translate sentences in natural language into formulas of an adapted semantic language that Montague called « intentional logic ».

3.1 The semantic representation

Intentional Logic (or IL) generalizes the first order predicate logic by including typed lambda-calculus and by making a general use of the notion of modality through the concept of intension (Dowty 81). Only a simplified version of this framework (not taking into account intensions) is recalled here.

- IL is a typed language : the set I of all possible types of IL includes
  * elementary types : $e \in I$ (type of « entities ») and $t \in I$ (type of « truth values ») ;
  * for any types $u \in I$ and $v \in I$, $<u,v> \in I$ ($<u,v>$ is the type of functions taking an argument of type u and giving a result of type v).
- semantics of IL : a denotation set D_w is associated with every type $w \in I$ as follows :
  * $D_e = E$ where E is the denumerable set of all entities of the world ;
  * $D_t = \{0,1\}$;
  * $D_{<u,v>} = D_v^{D_u}$: the denotation set of a composed type is a function.

3.2 Translation as an isomorphism

Each analysis tree produced by a CCG can be « translated » into IL :

- translation of the categories into logical types (function $k : C' \longrightarrow I$) :
  * basic categories : in our example, $k(S)=t$, $k(T)=e$, $k(CN)=<e,t>$;
  * derived categories : for any $X \in C'$ and $Y \in C'$: $k(X/Y)=k(Y\backslash X)=<k(Y),k(X)>$.
- translation of the words ($q : V \times C' \longrightarrow IL$) : each couple (v,U) where v is a word in V and $U \in f(v) \subseteq C'$ is (one of) its category(ies) is associated with a logical formula q(v,U) of IL whose type is $k(U) \in I$. The most usual and useful translations are :
  * $q(a,(S/(T\backslash S))/CN)=\lambda P \lambda Q \exists x[P(x) \wedge Q(x)]$
    $q(every,(S/(T\backslash S))/CN)=\lambda P \lambda Q \forall x[P(x) \rightarrow Q(x)]$
    where x and y are variables of type e, P and Q variables of type <e,t>.
  * the verb « to be », as a transitive verb, is translated by :
    $q(is,(T\backslash S)/T)=\lambda x \lambda y[y=x]$
    with x and y variables of type e.
  * Every other word w is translated into a logical constant noted w'.
- translation of the rules of combination :
  Rules R1 and R'1 are translated into oriented functional applications (Moortgat 88) :
  W1 : $f \,.\, x \longrightarrow f(x)$
  W'1 : $x \,.\, f \longrightarrow f(x)$

These definitions preserve the correspondence between categories of the grammar and types of logic. This property assures for example that syntactically correct sentences (of category S) will be translated into logical propositions (of type k(S)=t, i.e. with a truth value).

Example :

The example sentences analyzed in figures 1 and 2 can now be translated into IL, as shown in figures 3 and 4 respectively.

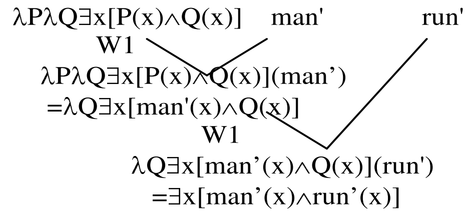

figure 3 : semantic translation of tree n°1

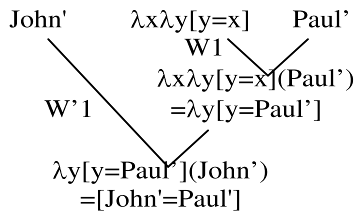

figure 4 : semantic translation of tree n°2

4. The learning model

4.1 Innate knowledge and concepts to learn

When a human being learns a natural language, we suppose that he has at his disposal sentences syntactically correct and semantically relevant. The corresponding situation in our model is an algorithm which takes as inputs a sentence that can be analyzed by a CCG together with its logical translation into IL.

The innate knowing supposed is reduced to the inference rules R1 and R'1 and the corresponding translation rules W1 and W'1. As opposed to usual semantic-based methods of learning, *no word meaning is supposed to be initially known*.

Finally, what does the learner has to learn ? In our linguistic framework, syntactic and semantic information are attached to the members of the vocabulary by functions f and q. These functions are the target outputs of the algorithm. More precisely, the syntactic and semantic knowledge to be learned can be represented as a finite list of triplets of the form : (v,U,w) where $v \in V$, $U \in f(v) \subseteq C'$ and $w=q(v,U) \in IL$.

Example :

Learning the example grammar previously used means learning the following set :

H={(John, T, John'), (Paul, T, Paul'),
(is, (T\S)/T, $\lambda x \lambda y[y=x]$), (runs, T\S, run'),
(a, (S/(T\S))/CN, $\lambda P \lambda Q \exists x[P(x) \wedge Q(x)]$)...}.

4.2 The learning algorithm

The proposed leaning strategy, given in figure 5, consists in building a hypothesis set, updated after each new input, to approach the target set.

For every couple <s,τ(s)> where s is a sentence and τ(s) its logical translation in IL, do :
- if there is one, affect to the words in s their category in the current hypothesis set ;
  else, make hypotheses on the category associated by f with the unknown words of s ;
- For every possible analysis tree :
  * translate the tree into IL ;
  * compare the final translation with τ(s) and infer possible values for the unknown semantic translation of words to update the current hypothesis set.

Figure 5 : the learning strategy

4.3 A detailed example

At the beginning, the current hypothesis set is the empty set. Let us suppose that the first given example is <John runs, run'(John')>.

- the syntactic hypotheses : the only categories allowing to build an analysis tree are
  * first possibility : f(John)=A and f(runs)=A\S ;
  * second one : f(John)=S/B and f(runs)=B.
  where A and B can be any category in C', basic or not.
- the semantic translation :
  * first possibility : see fig. 6 (the input data are put into rectangles).

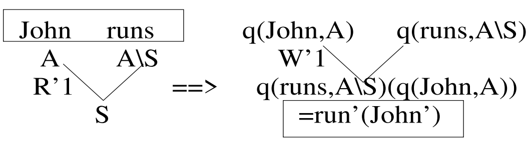

figure 6 : hypothesis H1

If we compare q(runs,A\S)(q(John,A)) with τ(s)=run'(John'), it leads to :

q(run,A\S)=run' and q(John,A)=John'.
So a possible hypothesis set is :
H1={(John,A,John'), (runs,A\S,run')}.
Similarly, the second possibility leads to another possible hypothesis set :
H2={(John,S/B,run'), (runs,B,John')}.

At this stage, we have no reason to prefer one hypothesis to the other (the learner does not know that John is linked with John', neither about runs and run'). The current hypothesis is then : H1 OR H2. But suppose now that a second given example is <Paul runs, run'(Paul')>. The same process applies to this example, except that « runs » now belongs twice to the current hypothesis set.

- the syntactic hypotheses : the new sentence treated with H1 forces to affect the category A to « Paul », while H2 forces to affect the category S/B.
- the semantic translation :
  * in the first possibility, H1 becomes
    H1'={(John,A,John'),(runs,A\S,run'), (Paul,A,Paul')}
  * it is impossible to provide a value to q(Paul,S/B) following the tree built with hypothesis H2.

So H2 is abandoned and only H1' remains. It can be noticed that a similar conclusion would have followed if the second example had been :
<John sleeps, sleeps'(John')>.

Any other example sentence including one of the words concerned by the current hypothesis is enough to discredit hypothesis H2.

5. Evaluation and conclusion

The choices made in this model have theoretical backgrounds and consequences.

First, CCG seem to be particularly adapted to the learning process. Recent researches have found conditions under which the syntax of these grammars is learnable (Buszkowski & Penn 90, Kanazawa 96). But, in these frameworks, tree structures are provided as inputs to the learning algorithm : in our model, the semantic translation plays a close role but in a weaker and more cognitively relevant fashion. Adriaans (92) also proposed a learning algorithm for categorial grammars, using both syntactic and semantic inputs, but he treated them separately : the semantic learning could only start when the syntactic learning was achieved, instead of helping it as we propose.

Previous models built in the syntactico-semantic spirit (Anderson 77, Hamburger & Wexler 75, Hill 83, Langley 82,) used more traditional syntax and semantic representations very close to syntactic structures (Pinker 79) : they failed to represent complex logical relations like quantification or Boolean operators. Logical languages like IL are more powerful and *a priori* independent from linguistic structures. In fact, our approach assumes that logic is the natural « language of the mind » in that situations perceived by our learner are supposed to be automatically translated into logical formulas *before* being compared with linguistic expressions.

Fundamentally, what makes natural languages learnable in our model is the presupposition that *there exists an isomorphism between the syntax of sentences and their semantics*. This strong principle of compositionality is contested by some linguists but remains an interesting approximation. The « graph deformation condition » used in (Anderson 77) was a weaker version of it. Under this condition, the inputs provided to the learner are the leaves and root respectively of two isomorphic trees and what is to be reconstituted is the body of these trees, as displayed in figure 6. But, as opposed to (Anderson 77), there is an asymmetry : the formalism chosen is adapted to language analysis but not to language generation.

The efficiency of the algorithm seems to crucially rely on *the complexity of the input relatively to the current hypothesis*. This complexity can be measured by the number of new words appearing in a sentence example. If few new words are introduced in each new example, the number of hypotheses to explore will remain reasonable. Else, the learning may be too complicated. Of course, this valuable intuition still needs to be formulated and proved in a more formal way.

It is not possible to develop here how to treat the cases when a word needs more than one category, but it remains possible to learn in this context. The learning is incremental.

The framework is still incomplete because we haven't chosen any learning model and we haven't proved the learnability of any language in it with our strategy. An extended and more general version of the algorithm in figure 5, using Lambek grammars (Lambek 58), is being implemented and tested. But the approach seems original and interesting enough to be developed further.

6. Bibliography

Adriaans, P. W. (1992). *Language Learning from a Categorial Perspective*, Doctoral dissertation, University of Amsterdam.

Anderson, J., R. (1977). Induction of Augmented Transition Networks. *Cognitive Science*, 1, 125-157.

Bar Hillel, Y. (1953). A quasi-arithmetical notation for syntactic description. *Language* 29, 47-58.

Bar Hillel, Y., Gaifman, C. & Shamir, E. (1960). On Categorial and Phrase Structure Grammars, *Bulletin of the Research Council of Israel*. 9F, 1-16.

Buszkowski, W., Penn, G. (1990). Categorial Grammars Determined from Linguistic Data by Unification, *Studia Logica*. 49, 431-454.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*, Cambridge, MIT Press.

Chomsky, N. (1968). *Language and Mind*. Brace & World.

Dowty, D. R., Wall, R. E., Peters, S. (1981). *Introduction to Montague Semantics*. Reidel, Dordrecht.

Gold, E. M. (1967). Language Identification in the Limit. *Information and Control*, 10, 447-474.

Hamburger, H., Wexler, K. (1975). A Mathematical Theory of Learning Transformational Grammar. *Journal of Mathematical Psychology*, 12, 137-177.

Hill, J. A. C. (1983). A computational model of language acquisition in the two-year-old. *Cognition and Brain Theory*, 6(3), 287-317.

Kanazawa, M. (1996). Identification in the Limit of Categorial Grammars. *Journal of Logic, Language & Information*, 5(2), 115-155.

Lambek, J. (1958). The Mathematics of Sentence Structures., *American Mathematical Monthly*, 65, 154-170.

Langley, P. (1982). Language acquisition through error discovery. *Cognition and Brain Theory*, 5, 211-255.

Montague, R. (1974). *Formal Philosophy; Selected papers of Richard Montague*. Yale University Press, New Haven.

Moortgat, M. (1988). *Categorial investigations, logical and linguistic aspects of the Lambek Calculus*. Foris, Dordrecht.

Oehrle, R. T., Bach, E., & Wheeler, D. (Eds.) (1988). *Categorial Grammars and Natural Language Structure*. Reidel, Dordrecht.

Pinker, S. (1979). Formal models of language learning. *Cognition*, 7, 217-283.

Shinohara, , T. (1990). Inductive inference of monotonic formal systems from positive data, in Arikara, S., Goto, S., Oshuga, S. & Yokomori, T. (Eds) *Algorithmic Learning Theory*, 339-351, Ohmsha and New York and Berlin, Springer.

Sakakibara, Y. (1992). Efficient learning of context-free grammars from positive structural examples. *Information & Computation*, 97, 23-60.

Valiant, L. G. (1984). A theory of the learnable. *Communication of the ACM*, 1134-1142.