Expert model in Compounds, an ITS for second language learning

Pascale Sébillot, Frédéric Danna
Irisa, Campus de Beaulieu, F-35042 Rennes Cédex, France
Paul Boucher
Cerlico, Université de Nantes, F-44000 Nantes, France

Keywords: knowledge representation in ITS, expert model, English compounds

1 Introduction

In this paper, we present a part of *Compounds*, an Intelligent Tutoring System (ITS) which addresses the problem of learning English as a foreign language and deals specifically with the formation and interpretation of compounds. Its aim is to bring French students to understand, produce and translate English compounds as fluently as a native Englishman. The reasons that led us to develop this ITS are of two types: on the one hand, the compounding process is an essential part of English and an extremely productive process in this language. Therefore, a French student who wants to learn English needs to know how to produce and understand existing and new compounds. On the other hand, French students have a lot of difficulties to deal with English compounding process; these problems are due to a large extend to the fact that *English is a word composition language while French is a stem composition language* [Blo33, Ben74]. Moreover, recent linguistic litterature devotes relatively little space to compounding processes in English, and, among the works in this domain, very few concern language pedagogy.

An ITS is usually composed of four components [NV88]: *the expert model* which contains the knowledge of the domain, *the student model* which contains information on the student and on his knowledge, *the teaching module* which manages the teaching plans and *the interface* which permits the communication between the student and the ITS. Here, we focus on the first developed part of Compounds: the expert (English) knowledge and its representation. *The linguistic competence* of the fluent speaker is idealised as the *ability to generate the appropriate combination* of words corresponding to a given *semantic relationship*, and conversely, as the *ability to analyse a given surface configuration* in terms of its *underlying semantic relationship.* Among the different possible patterns for English compounds, we have decided to focus firstly on the most productive and easily interpreted ones : N-N (dog-fish), the primary compounds V-N (cut-throat) and N-V (spoon-feed), and the synthetic compounds N-Ving (data-processing), Ving-N (swimming-pool), N-Ver (window-washer) and Ver-N (killer-shark) with right heads.

We present the linguistic theories that we have studied in order to determine the rules that permit to find a definition for a given compound and, conversely, to find a compound corresponding to a definition. We have extended these theories to treat all the designated compounds. We explain how these strong linguistic bases enable us to build and represent the expert knowledge in Compounds and we briefly describe the implementation and its results.

2 The linguistic framework

We have studied the works of different linguists to determine the expert knowledge for our ITS. *Downing* [Dow77] is the only one who really treats the problem of N-N compounds. She gives definitions for such compounds based on the semantic class (human, animal, ...) of their heads. *Lieber* [Lie83] deals with all types of compounds though she gives no precisions for N-N compounds. She defines two principles (the Feature Percolation Conventions and the Argument Linking Principle) to determine the correctness and the semantic of the compounds. We have discussed some of her results, modified the FPC and limited the contribution of her works to ours for primary compounds. *Selkirk* [Sel82] proposes a convention (the First Order Projection Condition) to deal with endocentric (headed) adjective or noun compounds whose head is derived from a verb and whose nonhead constituent is in a thematic relation (agent, theme, ...) with the head adjective or noun. We have extended her condition in order to treat all the synthetic compounds.

3 The knowledge representation

The previous works enable us to represent the expert knowledge, i.e. to determine the composition patterns used by a native Englishman to give a definition for a compound and, conversely, to find the compound corresponding to a given definition. The different patterns are represented in the next two tables.

Conversion of a compound into a definition		
example	compound	definition
bird-dog	object-subject	a subject which is in some relation with a object
swimming-pool	predicate-ing -subject	a subject which is used for predicate-ing
worker-bee	predicate-er -subject	a subject who predicate-s
data-processing	object- predicate-ing	the act of predicate-ing object-s
window-washer	object- predicate-er	someone who predicate-s object-s

The first definition can be improved thanks to Downing's method.

The first table is interpreted as follows: thanks to the previous linguistic theories, if a compound, like for example swimming-pool, is parsed as a predicate (swim) plus a suffix (ing) plus the subject of the predicate (pool), then the correct corresponding definition (or semantic relationship) is *a subject which is used for predicate-ing*, i.e. a pool which is used for swimming.

Conversion of a definition into a compound		
definition	compound	example
a subject {which,who,...} is used for predicate-ing	predicate-ing -subject	swimming-pool
a subject {which,who,..} predicate-s	predicate-er -subject	worker-bee
a subject {which,who,...} predicate-s object-s	object -subject	bird-dog
someone {which,who,...} predicate-s object-s	object -predicate-er	window-washer
the act of predicate-ing object-s	object -predicate-ing	data-processing

This table is interpreted as follows: thanks to the previous linguistic theories, if a definition is, for example, parsed as a subject which is used for predicate-ing (a pool which is used for swimming), then the correct corresponding compound is *predicate-ing - subject*, i.e. swimming-pool.

This type of representation also enables us to represent the correct and incorrect knowledge of a student, i.e. the correct and incorrect patterns that he has in mind.

4 Results and future works

We have developed an implementation of the expert model of our ITS on a SUN-4 station in Prolog-II/Mali. We have built a logical grammar which permits to transform a compound into a logical representation in order to get its definition, and, conversely, to transform a definition into a logical representation in order to get the corresponding compound. We have added a control of morphosyntactic features to verify the correctness of the proposed compounds and definitions. The times to get the compound or the definition are included between 0.2 and 0.6 second.

Our works are based on a strong linguistic model which is used, together with lexical information, to represent the expert knowledge in terms of a collection of composition patterns which depend on the type of the compound. The same sort of patterns is used to represent the correct and incorrect knowledge of the students.

The current limits of our works are of two types: on the one hand, there exist some linguistic limitations. For example, there is no strict correspondence between the world of compounds and the world of definitions. The pattern *someone who predicate-s object-s* may correspond to the compound *an object- predicate-er* (window-washer), but, in some rare cases, it may correspond to *a predicate-object* (cut-throat). We have not find anything to distinguish between both of them and we produce throat-cutter for the second case. On the other hand, we have not currently extended our system to all sorts of compounds (left-headed, exocentric ones, ...). Moreover, up to now, we have only proposed a generic student model for Compounds, i.e. all the possible right and wrong composition patterns that students may have in mind. Future works will consist in developing a real student model and in detecting the learning and forgetting possibilities of the considered student.

References

[Ben74] E. Benveniste. Problèmes de linguistique générale. *Vol. 2, Gallimard, Paris*, 1974.

[Blo33] L. Bloomfield. *Language.* G. Allen and Unwin Ltd. London, 1933.

[Dow77] P. Downing. On the creation and use of English compound nouns. *Language, Vol. 53, No 4*, pages 810–842, 1977.

[Lie83] R. Lieber. Argument linking and compounds in English. *Linguistic Inquiry, Vol. 14, No 2*, pages 251–285, 1983.

[NV88] J.F. Nicaud and M. Vivet. Les tuteurs intelligents : réalisations et tendances de recherches. *Technique et Science Informatiques, Vol. 7, No. 1*, 1988.

[Sel82] E. Selkirk. *The syntax of words.* Cambridge Mass., MIT Press, 1982.