

A Fully Lexicalized Grammar for French Based on Meaning-Text Theory

Sylvain Kahane

CNRS & Lattice-Talana, Université Paris 7, UFRIL, case 7003,
2, place Jussieu, 75251 Paris Cedex 05, France
sk@ccr.jussieu.fr
<http://www.linguist.jussieu.fr/~skahane>

Abstract. The paper presents a formal lexicalized dependency grammar based on Meaning-Text theory. This grammar associates semantic graphs with sentences. We propose a fragment of a grammar for French, including the description of extractions. The main particularity of our grammar is it that it builds bubble trees as syntactic representations, that is, trees whose nodes can be filled by bubbles, which can contain others nodes. Our grammar needs more complex operations of combination of elementary structures than other lexicalized grammars, such as TAG or CG, but avoids the multiplication of elementary structures and provides linguistically well-motivated treatments.¹

1 Introduction

Meaning-Text theory (MTT) has been developed since more than thirty years, but no complete formalization of the model has been achieved. Our main goal in this paper is to propose a formal grammar based on MTT. We insist on the fact that our grammar in any case is a ‘kosher’ implementation of MTT.

Following the MTT postulates ([16]: 53), we consider that 1) a grammar² is a formal system which ensures the bidirectional correspondence between texts and meanings (= semantic representations) and that 2) intermediate levels of representation—a morphological level and a syntactic level—must be considered and that a grammar consists of several modules which establish correspondence between representations of adjacent levels. Our grammar is composed of three modules: the morphological module ensures the correspondence between sentences and morphological representations, the syntactic module ensures the correspondence between morphological and syntactic representations and the semantic module ensures the correspondence between syntactic and semantic representations.

In section 2, we present the different levels of representation, in section 3, the syntactic module and in section 4, the semantic module. Our formalism will be exemplified by a fragment of French grammar. Extractions, which need extensions of the formalism, will be treated separately in section 5.

¹ I want to thank Jasmina Miličević, as well as two anonymous referees, for many valuable comments and corrections.

² The term grammar is used in its Chomskian sense of ‘linguistic model’.

A. Gelbukh (Ed.): CICLing 2001, LNCS 2004, pp. 18-31, 2001.

© Springer-Verlag Berlin Heidelberg 2001

2 Different Levels of Representation of a Sentence

A *morphological representation* of a sentence is the sequence of the morphological representations of the words of the sentence; the morphological representation of a word is *surface lexical unit* accompanied with a list of *surface grammemes*. Consider the sentence:

- (1) Zoé a parlé à un type étrange.
 Zoé has talked to a guy strange
 ‘Zoé has talked to a strange guy’

The morphological representation of (1) is:

- (2) ZOÉ_{sg} AVOIR_{ind,présent,3,sg} PARLER_{part passé} À UN_{masc,sg} TYPE_{sg} ÉTRANGE_{masc,sg}

The *syntactic representation* of a sentence is a non ordered *dependency tree* similar to the surface syntactic trees of MTT ([16]) or the stemmas of Tesnière ([23]). The nodes of the structure are labeled with surface lexical units, each being accompanied with a list of *surface grammemes*, and the dependencies are labeled with (surface) *syntactic relations*.

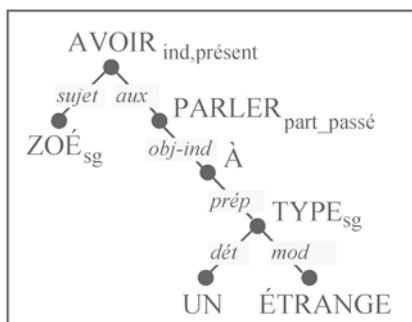


Fig. 1. Syntactic representation of (1)

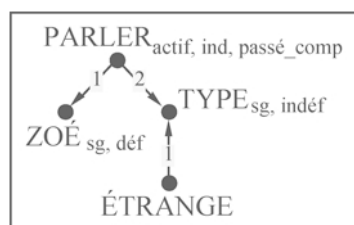


Fig. 2. The semantic representation of (1)

The *semantic representation* of a sentence is a directed graph whose nodes are labeled by deep lexical units, each being accompanied by a list of deep grammemes. A *deep lexical unit* corresponds to a surface lexical unit or a group of surface lexical units making an idiom. A *deep grammeme* is a grammeme with a meaning (including voice grammemes, which do not exactly express a meaning, but depend on semantic communicative choices). A deep grammeme can correspond to a surface grammeme or a complex expression including surface lexical units: that is the case of the French *passé composé* (AVOIR_{ind,présent} + V_{art,passé}) or the French passive voice (ÊTRE + V_{part,passé}). A deep lexical unit acts like a predicate and is linked to its arguments by arrows pointing on them. The different arrows emerging from a deep lexical unit are numbered from 1 to n following the increasing syntactic salience of the arguments. Such an arrow, representing a predicate-argument relation, is called a *semantic dependency*; the predicate is *the semantic governor* and its argument, the *semantic dependent*. Our semantic representation is a compromise between the semantic and deep syntactic representations of MTT. On the one hand, the nodes of our semantic representation are labeled with deep lexical units, rather than by semantemes (= the mean-