

Graph Interpolation Grammars: a Rule-based Approach to the Incremental Parsing of Natural Languages

John Larchevêque

▶ **To cite this version:**

John Larchevêque. Graph Interpolation Grammars: a Rule-based Approach to the Incremental Parsing of Natural Languages. [Research Report] RR-3390, INRIA. 1998. inria-00073299

HAL Id: inria-00073299
https://inria.hal.science/inria-00073299

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Graph Interpolation Grammars: a rule-based approach to the incremental parsing of natural languages

John Larchevêque

N ° 3390

Mars 1998

THÈME 3



ISSN 0249-6399



Graph Interpolation Grammars: a rule-based approach to the incremental parsing of natural languages

John Larchevêque

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Atoll

Rapport de recherche n ° 3390 — Mars 1998 — 41 pages

Abstract: Graph Interpolation Grammars are a declarative formalism with an operational semantics. Their goal is to emulate salient features of the human parser, and notably incrementality. The parsing process defined by GIGs incrementally builds a syntactic representation of a sentence as each successive lexeme is read. A GIG rule specifies a set of parse configurations that trigger its application and an operation to perform on a matching configuration. Rules are partly context-sensitive; furthermore, they are reversible, meaning that their operations can be undone, which allows the parsing process to be nondeterministic. These two factors confer enough expressive power to the formalism for parsing natural languages.

Key-words: parsing, natural language processing, psycholinguistics, linguistics, grammar, syntactic representation, lexical representation

(Résumé : tsvp)

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : 01 39 63 55 11 - International : +33 1 39 63 55 11
Télécopie : (33) 01 39 63 53 30 - International : +33 1 39 63 53 30

Les Grammaires à Interpolations de Graphes : un formalisme déclaratif pour l'analyse incrémentale du discours

Résumé : Les Grammaires à Interpolations de Graphes sont un formalisme déclaratif muni d'une sémantique opérationnelle. Elles visent à émuler certaines caractéristiques de l'analyseur syntaxique humain, et notamment son caractère incrémental. Le processus d'analyse qu'elles définissent construit une représentation syntaxique incrémentalement, à mesure que chaque lexème est lu. Dans une GIG, une règle comporte une représentation des configurations déclenchantes et une spécification des opérations à appliquer sur ces configurations. Les règles comportent un élément contextuel; de plus, elles sont réversibles, en ce sens qu'il est possible d'en défaire les effets, ce qui permet une analyse non déterministe. Ces deux facteurs contribuent à conférer aux Grammaires à Interpolations de Graphes un pouvoir d'expression suffisant pour l'analyse des langues naturelles.

Mots-clé : syntaxe, langage naturel, psycholinguistique, linguistique, grammaire, représentation syntaxique, représentation lexicale

1 Introduction

1.1 Characteristics and rationale

A graph interpolation grammar is a grammar formalism with an operational semantics. A rule in a graph interpolation grammar specifies not only syntactic relations but also an elementary parsing operation.
Rules are lexicalized in the sense that each rule describes the combinatory properties of a lexical item. Parsing a sentence consists in matching each lexeme in the input string with a rule in the grammar and applying this rule to the current parse representation.
GIG-driven parsing was designed with incrementality and flexibility in mind, so as to emulate some features of the human parsing capability, in particular incremental processing, error tolerance, and handling of complex word orders.
In a complete model of discourse understanding, incremental parsing would be shown to work in tandem with some form of composition between partial semantic representations. It is to be thought that the collaboration between syntax and semantics in natural discourse understanding is fairly close, and that backtracking in the parser is frequently initiated by a clash found between semantic features. This report, however, will not attempt to give even the roughest idea of what semantic representations should look like and will therefore exclusively focus on syntactic phenomena, sometimes at the cost of simplifying the phenomena at hand.

1.2 Plan of the report

The first two sections give a fairly complete presentation of the concepts and processes involved in parsing with a Graph Interpolation Grammar. The syntactic structures generated when parsing with a GIG are described in Section 2, while the grammar rules and the parsing process are described in Section 3.
The next few sections apply the formalism to selected problems. Section 4 compares the handling of a simple expression language using a Context Free Grammar and a Graph Interpolation Grammar, and Section 5 analyzes Dutch Cross-Serial Dependencies.
Finally, the conclusion indicates topics for further research and related works.

2 Syntactic structures

2.1 Phrases

A phrase is made up of a head and its complements. It will be represented by a graph with a root and one labelled edge from the root to each nonroot node.
The root represents the head of the phrase, the other nodes represent its complements; and the edge labels represent grammatical functions.