Defining a Lexicalized Context-Free Grammar for a Subdomain of Portuguese Language

Cinthyan Renata Sachs C. de Barbosa
Electronic Engineering and Computation Program of the Aeronautics Technological Institute (ITA) and Computer Sciences
Davidson Cury
Informatic Institute of the Federal University of Espírito Santo (UFES), Brazil.
José Mauro Volkmer de Castilho (*in memorian*)
Informatic Technology Institute of the Federal University of Rio Grande do Sul (UFRGS), Brazil.
Celso de Renna e Souza
Computer Science Division of the Aeronautics Technological Institute (ITA), Brazil.

1. Introduction

According to [1], an emphasis must be given to grammar representations for describing and generating the sentences that make up a given language and an emphasis on processing and computation that demonstrate that the grammar for a language has important implications as to how it can be processed. Some observations by [2] and [3], say that a general-purpose grammars should be linguistically plausive, both to be extensible and to take advantage of work in computational linguistics. Linguistic theories have become much more computationally oriented recently, and some large, general-purpose grammars are now available [4]. But the grammars have to be sufficiently robust to cope gracefully with sentence fragments and ill-formed input. It is also very important in the database context to have a good treatment of proper names, domain-specific identifiers, abbreviations and other items which cannot be part of a general-purpose lexicon [2].

A NLP system's success depends on its knowledge of the application domain, namely the relative completeness of the natural language model it encapsulates and the appropriateness or efficiency of its algorithms [5]. In the specific case of the NLs, the study of context-free languages (CFL) has been of special interest because they permit a simple representation of the syntax, adequate for formal structuring, as for computational analysis [6]. Context-free grammars (CFGs) have been a well accepted framework for computational linguistics for a long time [7], [8]. The recognizer algorithms and generator algorithms that implement CFLs are relatively simple and have a good efficiency. In addition, according to [9], it was the CFGs with some restrictions that had greater progress in the description of NL.

According to [7], lexicalization is important, from a computational perspective, because, other things being equal, lexicalized grammars can often be parsed much more efficiently than non-lexicalized ones. In [10] it is affirmed that this can be done associating each elementary structure in a grammar with a lexical item (terminal symbol in the context of formal grammars). A type of lexicalization is Greibach Normal Form (GNF) for Context-Free Grammar (CFG). In contrast to GNF, that is regarded as a kind of *weak* lexicalization, for not preserving structure of the original, Lexicalized Grammar can be considered as a *stronger* version of GNF, in the sense that the structures are preserved and not just the string sets (weak generative capacity). Lexicalization is of interest from a linguistic perspective, because most current linguistic theories give lexical accounts of a number of phenomena that used to be considered purely syntactic [7]. The information put in the lexicon is thereby increased in both amount and complexity [10]. According to [10], some of the linguistic formalisms illustrating the increased use of lexical information are, lexical rules in Lexical-Functional Grammar (LFG), Generalized Phrase Structure Grammar (GPSG), Head-Driven Phrase Structure Grammar (HPSG), Combinatory Categorial Grammar, Karttunen's version of Categorial Grammar, some versions of Chomsky's Government Binding Theory (GB theory), and Lexicon-Grammars.

Every method for lexicalized CFGs in the strong sense defined has required context-sensitive operations. As a result, every method for lexicalizing CFGs has shared with LTAG the unfortunable feature that lexicalization leads to dramatically decreased rather than increased computational performance [7]. According to [8], although LTAG is considered to be interesting, it is executed at the cost of decreased efficiency, $O(n^6)$-time in the worst case [11], [12], [13]. As a result [7], there are no computational advantages in lexicalizing a CFG using LTAG because of the speed up due to the grammar becoming lexicalized being swamped by the dramatic increase in fundamental worst case cost. So, we will focus on Lexicalized Context-Free Grammars (LCFG), a class of grammars originally introduced in [14]. They are attractive because they combine the

 Proceedings of the Sixth International Workshop on Tree Adjoining Grammar and Related Frameworks (TAG+6), pp. 74-79. Universitá di Venezia.

elegance of LTAG and the efficiency of CFG [14] and, it is also possible through an algorithm [8] to show how the $O(n^3)$-time worst case complexity can be achieved for LCFG (n is the length of sentence). Through this formalism, constructions of CFGs for the definition of the real constructions of an interface, will be dealt with.

In this work the medical area was chosen, with emphasis on *a portuguese grammar generating a medical language for radiological queries* over a set of radiographs of a Ewing sarcom case. The treated language subset includes active and passive voices, relative and interrogative clauses, their combinations and pronouns. The language generated by this grammar was validated from knowledge obtained in the FLAMA (*Tools and Authoring Language for Modeling Process*). This framework had as objective a set of tools and an authoring language that constitutes the framework of a highly specialized architecture for authoring activities [15], [16]. FLAMA was based on an archetypical reference RUI environment (*Representation for Understanding Images*) for the learning of radiology [17]. The access to knowledge contained in the knowledge base will be done through radiological language which *grammar* is the focus of investigation through this work. We are developing a correlated work dealing with parsing techniques for our grammar.

2. Radiological Grammar

Gs:
1) S' $\rightarrow$ Comp S / Comp S_q / Comp PP_q
2) S_q $\rightarrow$ Pro S
3) S $\rightarrow$ NP (M) VP / VP / ProReto VP
4) S_r $\rightarrow$ Rel S
5) NP $\rightarrow$ (Det) (Mod) N (Mod) (W)
6) W $\rightarrow$ (W) Mod
7) NP $\rightarrow$ (Det) (Mod) Ncom (Mod) (W)
8) NP $\rightarrow$ NP Conec NP (X)
9) X $\rightarrow$ (X) Conec NP
10) NP $\rightarrow$ NP S_r
11) Det $\rightarrow$ (Predet) Detbase (Pósdet)
12) Mod $\rightarrow$ AP / PP
13) N $\rightarrow$ N S_r
14) Ncom $\rightarrow$ Ncom S_r
15) VP $\rightarrow$ (Neg) V_{tr} (Intens) NP (SP_A) (Y) /
(Neg) V_{tr} OblAt (Intens) (SP_A) (Y)
16) Y $\rightarrow$ (Y) PP
17) VP $\rightarrow$ (Neg) V_{tr} (Intens) SP_C (SP_A) (Y)
18) VP $\rightarrow$ (Neg) V_{tr} (proap) (Intens) NP SP_C (SP_A) (Y) /
(Neg) V_{tr} (proap) (Intens) SP_C NP (SP_A) (Y) /
(Neg) V_{tr} OblAt (Intens) NP (SP_A) (Y) /
(Neg) OblAt V_{tr} (Intens) NP (SP_A) (Y) /
(Neg) V_{tr} RefCli (Intens) (SP_A) (Y) /
(Neg) V_{tr} (Intens) NP SP_R (SP_A) (Y) /
(Neg) V_{tr} (Intens) SP_R NP (SP_A) (Y)
19) VP $\rightarrow$ (Neg) V_{tr} (Intens) SP_C SP_C (SP_A) (Y)
20) VP $\rightarrow$ (Neg) V_{intr} (Intens) (SP_A) (Y)
21) SP_C $\rightarrow$ P NP / P OblTon
22) SP_R $\rightarrow$ P RefObl
23) VP $\rightarrow$ (Neg) V_E AP (Y)
24) VP $\rightarrow$ (Neg) V_E NP (Y)
25) VP $\rightarrow$ (Neg) V_E PP (Y)
26) VP $\rightarrow$ (Neg) V_E V_{par} (NP) (por NP) (PP)
27) VP $\rightarrow$ (Neg) V_{estar} V_{ger} (Y)
28) V_{tr} $\rightarrow$ (Neg) V_{tr} S_r
29) Comp $\rightarrow$ $\pm$*wh*
30) PP $\rightarrow$ P NP
31) PP_q $\rightarrow$ P_q Pro
32) PP $\rightarrow$ Adv
33) AP $\rightarrow$ (Intens) (SP_A) Adj (SP_C)
34) AP $\rightarrow$ (Intens) (SP_A) Adjcom (SP_C)
35) AP $\rightarrow$ AP Conec AP (Z)
36) Z $\rightarrow$ (Z) Conec AP
37) Adj $\rightarrow$ Adj S_r
38) Adjcom $\rightarrow$ Adjcom S_r

FIGURE 2.1 - Surface Grammar

A set of rules that identify references to the various kinds of diagnosis and symptoms are needed. Through Portuguese Transformational Grammar (GT) [3], [18], [19] it was possible to work the proposed domain for the following kinds of constructions: **grammatical groups** as sentence, noun phrase (e.g. "*the exam*"), verb phrase (e.g. "*has some differential diagnosis*"), adjective phrase (e.g. "*the osteomyelitis <u>is chronic</u>*"), prepositional phrase (e.g. "*the osseous texture is <u>of diffuse form</u>*"), adverb phrase, restrictive clause (e.g. "*the parameters <u>that were analyzed led to osteomyelitis</u>*"), **yes/no sentences** (e.g. "*The femur is reduced?*"), ***wh*-sentences** (e.g. "*What is the patient age?*"), **alternative sentences** in the usual form (not cleaved) (e.g. "*appear lithic or blastic lesion?*"), **sentences of solicitation of explanation** (e.g. "*Why* S?"), **existential S** (e.g. "*Some region was examined?*"), **S in the active voice** (e.g. "*The lesion compromise some region?*"), **S in the passive voice** (e.g. "*Some region is compromised by lesion?*"), **cleaved S** (e.g. "*Is the exam that confirmed the diagnosis?*"). A grammar was proposed (fig.2.1) which foresees the treatment of expressions of *specific vocabulary* (words, expressions and medical jargon [20]) utilized in radiology. A meticulous study of TG and of Transformation Rules for portuguese [21] was necessary for constructing a *surface grammar* (in TG the rewriting rules generate the *deep structure* of the sentence) that treats several syntactical aspects of radiological language. Next the relationship between the surface grammar and LCFG, will be shown.

3 Lexicalization of LCFG

In this section, we propose to extend the domain of locality of Radiological CFG in order to make lexical items appear local to the production rules. The domain of locality of a CFG is extended by using a tree rewriting system that uses substitution and a restricted form of adjunction. So, the syntactical relations described in the FCG presented in [21] are needed also in LCFGs. However, these relations are expressed by trees (*initial* and *auxiliary* trees) and operations above. In [22], it is affirmed that, due to formal properties of the adjunction, the formalism utilized becomes more powerful than FCGs. Not every grammar is in a lexicalized form. Given a grammar *G* stated in a formalism, we will try to find another grammar G_{lex} (not necessarily stated in the same formalism) that generates the same language and also the same tree set as *G* and for which the lexicalized property holds. We refer to this process as *lexicalization* of a grammar. We say that a formalism *F* can be lexicalized by another formalism *F'*, if for any finitely ambiguous grammar *G* in *F* there is a grammar *G'* in *F'* such that *G'* is a lexicalized grammar and such that *G* and *G'* generate the same tree set (and a fortiori the same language) [10]. Taking our *surface grammar* [21], [23] it is possible to map LCFG, permitting in the last representation an improvement in the computational performance. According to [7], the presence of lexical item as its leftmost non-empty constituent in LCFG facilitates efficient left to right parsing. The main reason motivate the construction of an CFG (fig.2.1) is due to number of initial and auxiliary trees in LCFG can be, in the worst case, much greater than the number of production rules in a FCG (in this case the surface grammar). The number of elementary trees in G_{lex} is related to the number of acyclic and minimal cycle paths in LG (*lexicalization graph*). More details in [7]. By using the *theorem* exhibited in [8], it is possible to transform the surface grammar into a LCFG, by means of a special artifice in the treatment of terminals, obtaining consequently, as in the surface grammar, valid constructions for the radiological language.

In linguistic context, according to [24], the nodes on the frontier in LCFG are *preterminal* lexical category symbols such as N (Noun), V (Verb), etc. For denoting these categories, the artifice utilized in this work will be the insertion of *pseudoterminals* in the surface grammar so that the theorem can be applied. The pseudoterminals, denoted by detbase, n, v, etc. (all denoted in small letter) will be in the productions of the preterminals as in N□ n, V □ v, etc. After applying the theorem the previous preterminals and these pseudoterminals will be marked with □, as seen in fig.3.1, indicating that its pseudoterminal child node, will have to activate a search in a dictionary for verifying if it is possible to use, in this point, the token from the input string. If possible, the pseudoterminal will be substituted by the terminal at issue. The lexical items, utilized in the radiological language, are in a dictionary, which is exhibited in [21].

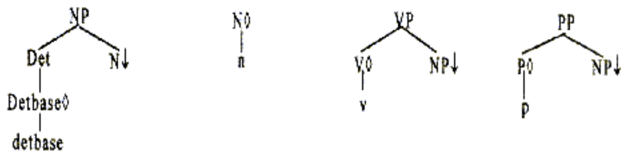

FIGURE 3.1 - Examples of trees before lexicalization

The principal unit of syntactic information associated with a LCFG entry is a tree structure in which the tree nodes are labeled with syntactic categories and feature information and there is at least one leaf node labeled with a *lexical* category (such lexical leaf nodes are known as *anchors*). Thus, the *lexicon* consists of a finite set of structure each associated with an anchor. The structures defined by the lexicon are called *elementary structures*. Structures built by combination of others are called *derived structures* [10]. With respect to a LCFG lexicon, this can be defined as the LTAG lexicon, for example, where this consists of a set of trees each one associated with one or more lexical items [14]. These elementary trees can be viewed as elementary clauses (including their transformational variants) in which the lexical items participate. The trees are combined by *substitution* and a restricted form of *adjunction* that is context-free in nature.

Fig.3.2 exhibits some of the constructions LCFG utilized in radiological domain. In grammars as LCFG, and others of the TAG family, the linguistic unit is this elementary tree, which corresponds to a minimal predicative structure. According to [22], these structures are syntactic and semantic units at the same time. They are gathered in tree families which encode the different lexical and syntactic rules which may apply to them [25]. When trees are combined by substitution or adjunction, the corresponding semantic representations are combined. In [24] focus that the elementary trees are the appropriate domains for characterizing certain dependencies (e.g., subcategorization and filler-gap dependencies). That is, an elementary tree localizes agreement dependencies, filler-gap dependencies and predicate-argument dependencies and also serves as a complex description of the anchor.

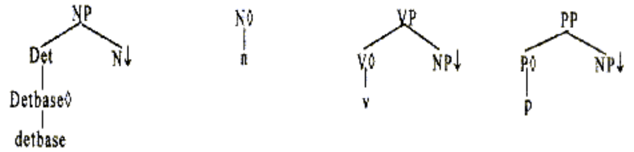

FIGURE 3.1 - Examples of trees before lexicalization

The principal unit of syntactic information associated with a LCFG entry is a tree structure in which the tree nodes are labeled with syntactic categories and feature information and there is at least one leaf node labeled with a *lexical* category (such lexical leaf nodes are known as *anchors*). Thus, the *lexicon* consists of a finite set of structure each associated with an anchor. The structures defined by the lexicon are called *elementary structures*. Structures built by combination of others are called *derived structures* [10]. With respect to a LCFG lexicon, this can be defined as the LTAG lexicon, for example, where this consists of a set of trees each one associated with one or more lexical items [14]. These elementary trees can be viewed as elementary clauses (including their transformational variants) in which the lexical items participate. The trees are combined by *substitution* and a restricted form of *adjunction* that is context-free in nature.

Fig.3.2 exhibits some of the constructions LCFG utilized in radiological domain. In grammars as LCFG, and others of the TAG family, the linguistic unit is this elementary tree, which corresponds to a minimal predicative structure. According to [22], these structures are syntactic and semantic units at the same time. They are gathered in tree families which encode the different lexical and syntactic rules which may apply to them [25]. When trees are combined by substitution or adjunction, the corresponding semantic representations are combined. In [24] focus that the elementary trees are the appropriate domains for characterizing certain dependencies (e.g., subcategorization and filler-gap dependencies). That is, an elementary tree localizes agreement dependencies, filler-gap dependencies and predicate-argument dependencies and also serves as a complex description of the anchor.

The sentences of radiological language will be obtained utilizing these structures and the lexical items contained in dictionary presented in [21]. Some of these sentences are exhibited in fig.3.3 which utilize elementary trees a13, a50, a26 e a38, and lexical items a, tíbia, tem, um, contorno, totalmente, regular, which are necessary to preterminals Detbase, N, Vtr, Detbase, N, Intens, Adj respectively. The trees that describe portuguese syntactic structures are grouped in families, specially according to criteria of verbal regency and transitivity [26]. Hierarchical representations of portuguese LCFG have been proposed, e.g., defining tree families [21]. A tree family contains the different possible trees for a given canonical subcategorization (or predicate-argument structure).

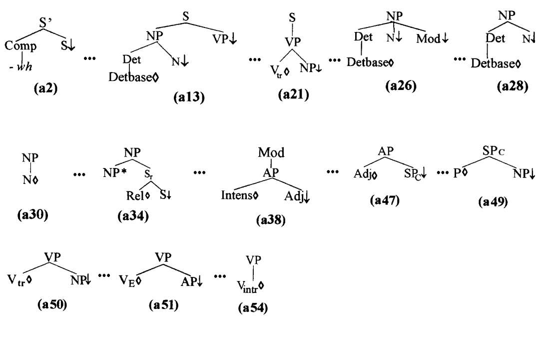

FIGURE 3.2 - Elementary structures utilized in radiological domain

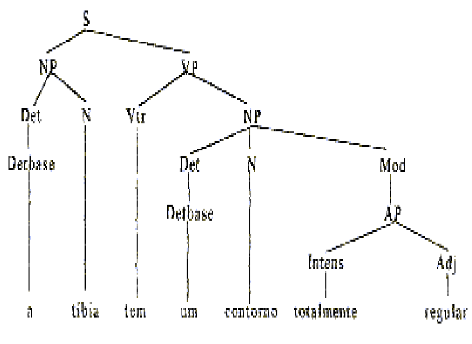

FIGURE 3.3 - Substitution in *a tíbia tem um contorno totalmente regular*

4. Conclusions

In this paper, we have presented some novel applications of LCFG. We have illustrated a radiological grammar, based on the LCFG formalism. A new approach to grammar in natural language to description of portuguese utilized in Natural Language Interfaces to Database (NLIDBs) was applied. The formalism TAG family was used, which is a tree-generating system rather than a string generating system. The set of trees derived from this family constitutes the object language. To describe these structures of portuguese, the conventional CFGs gave support in the definition of the syntactic structures of portuguese for LCFG. A meticulous study about context-free rules of phrase structure of TG was done. These rules of TG that foresee various linguistic problems of the NLIDBs [27] were defined for the deep structures. Transformation rules defined for the portuguese were described in [21] for attainment of surface structures. These rules were mapped for a surface grammar, which was taken as entry in the utilization of theorem presented in [8] that has as output an LCFG. LCFG is a formalism integrating lexicon and grammar. It has both linguistic advantages (e.g. elegant handling of unbounded dependencies and idioms [28]) and computational advantages, particularly due to lexicalization ([29]). The lexicalization in the LCFG showed to be interesting not only from the linguistic perspective, as such formal interest, because it is systematically associated with a lexical anchor. These structures specify extended domains of locality over which constraints can be stated. In [30] emphasizes the importance of key-concept of *extended domain of locality*, which is capable of allowing the dependency information and phrase structure information to be represented in one structure, that is, an elementary tree. This feature, according to [30], allows the phrase structure parse tree to be represented in terms of dependency information. The advantage of the extended domain of locality, according to [31], is that many relationships that must be mediated via grammar rules in other formalisms can be stated directly in the lexical entries.

The linguistic and mathematical advantages of lexicalized formalisms are useful in practical applications. Currently, most large scale NLP systems adopt CFG-like grammars, that have to face the problem of syntactic ambiguity, because the constraints expressed on the syntactic categories are too general to limit the huge hypothesis space generated by wide-coverage grammars [32]. So, LCFG favored the syntactical analysis of NL for the portuguese, because a broader domain of locality than usual phrase structure rules. This allows us to state, for example, subcategorization imposed on an element by another that does not directly dominate it, e.g., between the verb and the determiner of its first complement. Other linguistic advantage is that the adjunction permits the recursion in the composition of trees. The lexicalization notion in the LCFG is also linguistically very significant by preserving not only the string sets (weak generative capacity) but also the structures, i.e., *strong generative capacity*. The *syntactic criteria* defined by LCFG restrict the number of possible structures for a sentence, simplifying the semantic analysis. With relation to *concordance criteria*, these can be taken outside grammar, as proposed by [33] for not causing exponential explosion of the rules. In this work the concordance criteria are similar to *top* and *bottom* features [34] that are unified in LTAG.

An important fact, undoubtedly, is that the LCFGs do not require more computational recourses than CFGs. The greatest computational advantage [7] is that the parsing of an LCFG (obtained by CFG through lexicalization) is significantly faster than the one of the CFG. Although the string sets generated by LCFG are the same as those generated by CFG, LCFG is capable of generating more complex sets of trees than CFGs [7]. The fact that LCFG lexicalizes CFG is significant, because every other method for lexicalizing CFG without changing the trees derived require context-sensitive operations [10] and therefore dramatically increases worst case processing time. Context-sensitive operations (e.g., mildly context-sensitive formalisms [11]) entail much larger computational costs for parsing and recognition than CFGs. In particular, the fastest known LTAG parser require $O(n^6)$ - time in the worst case [12] in contrast to $O(n^3)$ for CFG. Since LCFG is a restricted case of TAG, standard $O(n^6)$-time TAG parsers [11], [12], [13] can be used for parsing LCFG. Although they require $O(n^6)$-time for TAG parsing, they can be made very easily to require at most $O(n^4)$-time for LCFG. This bound is still too high since we know that LCFGs generate only context-free languages. Then, recognizer and left to right parsing algorithm which requires $O(n^3)$-time in the worst case for LCFG has been proposed [14], [7], [21] to process a sentence of length *n*. Since the attractive aspects of LTAGs come at some computational cost, LCFG provides an efficient alternative which does not sacrifice the elegance of the LTAG analyses and which may be useful in different areas of computational linguistics.

Very restricted natural language does not work particularly well as a straight substitute for a traditional formal query language. A broader system covering a larger part of the Portuguese grammar and medical vocabulary is currently under development. The grammar has to be sufficiently robust to cope gracefully with sentence fragments and ill-formed input, words, expressions and medical jargon utilized in medicine. The grammar described in this paper has the possibility of to be inserted in broader studies which generate the portuguese language and not only the radiological language ones. Portuguese language interfaces for another applications as Operational Systems [35], Expert Systems [36], [37], [38], Intelligent Tutor Systems, etc. [15], [16], [21], [39] can use these grammar. Stochastic extensions [40] can also be incorporated to grammar for making it possible to capture both distributional and hierarchical information about portuguese words.

References

[1] KRULEE, G. K. *Computer Processing of Natural Language*. Englewood Cliffs, NJ: Prentice-Hall, 1991. 456p.

[2] COPESTAKE, A.; JONES, K. S. Natural Language Interfaces to Databases. *The Knowledge Engineering Review*, [S.l.], v.5, n.4, p.225-249, 1990.

[3] BARBOSA, C. R. S. C. de. Interfaces em Linguagem Natural para Banco de Dados. Porto Alegre: CPGCC da UFRGS, 1997. 165p. Trabalho Individual n.640.

[4] GROVER, C. et. al. The Alvey Natural Language Tools Grammar (second release). Cambridge: Computer Laboratory, University of Cambridge, 1989. (Technical Report 162).

[5] MANARIS, B. Z.; SLATOR, B. M. Interactive Natural Language Processing: Building on Success. IEEE Computer, New York, v.29, n.7, p.28-32, July 1996.

[6] MENEZES, P. F. B. Linguagens Formais e Autômatos. Porto Alegre: Instituto de Informática da UFRGS: Sagra Luzzatto Editores, 1998. 165p.

[7] SCHABES, Y.; WATERS, R. C. Lexicalized Context-Free Grammar: A Cubic-Time Parsable, Lexicalized Normal Form For Context-Free Grammar That Preserves Tree Structure. Broadway, Cambridge: Mitsubishi Electric Research Laboratories, June 1993. 30p. (Technical Report 93-04).

[8] SCHABES, Y.; WATERS, R. C. Lexicalized Context-Free Grammars. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 31., 1993, Ohio, USA. Proceedings... Ohio: [s.n.], 1993. p.121-129.

[9] RICH, E.; KNIGHT, K. Inteligência Artificial. São Paulo: Makron Books, 1993. 722p.

[10] JOSHI, A. K.; SCHABES, Y. Tree-Adjoining Grammars and Lexicalized Grammars. In: EUROPEAN SUMMER SCHOOL IN LOGIC, LANGUAGE AND INFORMATION, 4., 1992, Colchester, U.K. Proceedings... Colchester: University of Essex, 1992. p.1-23.

[11] VIJAY-SHANKER, K. A Study of Tree Adjoining Grammars. Philadelphia, USA: Department of Computer Science, University of Pennsylvania, 1987. PhD Thesis.

[12] LANG, B. The Systematic Constructions of Earley Parsers: Application to the Production of O(n6) Earley Parsers for Tree Adjoining Grammars. In: INTERNATIONAL WORKSHOP ON TREE ADJOINING GRAMMARS, 1., 1990, Dagstuhl Castle, FRG. Proceedings... Castle: [s.n.], 1990.

[13] SCHABES, Y. The Valid Prefix Property and Left to Right Parsing of Tree-adjoining Grammar. In: INTERNATIONAL WORKSHOP ON PARSING TECHNOLOGIES, 2., 1991, Cancun, Mexico. Proceedings... Cancun: [s.n.], 1991.

[14] SCHABES, Y. Lexicalized Context-Free Grammars. Broadway, Cambridge: Mitsubishi Electric Research Laboratories, Jan. 1993. 16p. (Technical Report TR 93-01).

[15] CURY, D. FLAMA: Ferramentas e Linguagem de Autoria para a Modelagem da Aprendizagem. São José dos Campos: CPG da Engenharia Eletrônica e Computação do Instituto Tecnológico da Aeronáutica, 1996. 151p. PhD Thesis.

[16] CURY, D.; OMAR, N.; DIRENE, A. I. Modelos Baseados em Estereótipos e Oráculos para a Aprendizagem de Conceitos Visuais. Revista Brasileira de Informática na Educação, n.02, p.43-53, Apr. 1998.

[17] DIRENE, A. I. Methodology and Tools for Designing Concept Tutoring Systems. Sussex: School of Cognitive and Computing Sciences, University of Sussex, 1993. PhD Thesis.

[18] CHOMSKY, N. Aspects of the Theory of Syntax. Cambridge, MA: MIT Press, 1965.

[19] ZORZO, A. F. Gramática Transformacional com Atributos. Porto Alegre: CPGCC da UFRGS, 1994. 99p. Master Thesis.

[20] BAUD, R. H.; RASSINOUX, A. M.; SCHERRER, J. R. Natural Language Processing and Knowledge Representation of Medical Texts. Methods of Information in Medicine, [S.l.], v.31, p.117-125, 1992.

[21] BARBOSA, C. R. S. C. de. Gramática para Consultas Radiológicas em Língua Portuguesa. Porto Alegre: CPGCC da UFRGS, 1998. 143p. Master Thesis.

[22] ABEILLÉ, A. et al. Non Compositional Discontinous Constituents in Tree Adjoining Grammar. In: EUROPEAN SUMMER SCHOOL IN LOGIC, LANGUAGE AND INFORMATION, 4., 1992, Colchester, U.K. Proceedings... Colchester: University of Essex, 1992. p.1-20.

[23] BARBOSA, C. R. S. C. de.; CASTILHO, J. M. V. de. Gramática Livre de Contexto Lexicalizada para a Análise Sintática da Língua Portuguesa - Uma Experiência na Geração de Consultas de uma Interface em Linguagem Natural para Banco de Dados. In: ENCONTRO PARA O PROCESSAMENTO COMPUTACIONAL DA LÍNGUA PORTUGUESA ESCRITA E FALADA, 5., 2000, Atibaia, SP. Anais... Atibaia: ICMC/USP, 2000. 193p. p.155-164.

[24] KROCK, A. Analysing Extraposition in a Tree Adjoining Grammar. In: EUROPEAN SUMMER SCHOOL IN LOGIC, LANGUAGE AND INFORMATION, 4., 1992, Colchester, U.K. Proceedings... Colchester: University of Essex, 1992. p.107-149.

[25] ABEILLÉ, A. et al. A Lexicalized Tree Adjoining Grammar for English. Philadelphia, USA: Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, 1990. (Technical Report MS-CIS-90-24).

[26] JOSHI, A.; VIJAY-SHANKER, K.; WEIR, D. The Convergence of Mildly Context-Sensitive Grammatical Formalisms. In: SELLS P.; SHIEBER, S.; WASOW, T. (Eds.). Foundational Issues in Natural Language Processing. Cambridge, MA: MIT Press, 1991. p.31-81.

[27] ANDROUTSOPOULOS, I.; RITCHIE, G. D.; THANISCH, P. Natural Language Interfaces to Databases - An Introduction. Journal of Natural Language Engineering, Cambridge, p.1-50, 1994.

[28] CANDITO, M.H. A Principle-based Hierarchial Representation of LTAGs. In: COLING'96., 1996, Copenhagen, Proceedings... Copenhagen, [s.n.], 1996. p.194-199.

[29] SCHABES, Y.; ABEILLÉ, A.; JOSHI, A. K. Parsing Strategies with 'Lexicalized' Grammars: Application to Tree Adjoining Grammars. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, 12., 1988, Budapeste, Hungria. Proceedings... Budapeste: ACL, 1988. 843p. p.578-583.

[30] JOSHI, A. K.; SRINIVAS, B. Using Parsed Corpora for Circumventing Parsing. In: WERMTER, S.; RILOFF, E.; SCHELER, G. (Eds.). Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing. Berlin: Springer-Verlag, 1996. p.413-424.

[31] POPOWICH, F. Lexical Characterization of Local Dependencies with Tree Unification Grammar. Burnaby, Canada: School of Computing Science, Simon Fraser University, Sep. 1993. 36p. (Technical Report CMPT TR 93-13).

[32] BARBERO, C.; LOMBARDO, V. Wide-coverage Lexicalized Grammars. [S.l.:s.n.], 1996.

[33] MONTEIRO, S. L. Um Subconjunto Regular do Português. Rio de Janeiro: Laboratório Nacional de Computação Científica, 1988. 43p. (Relatórios de Pesquisa e Desenvolvimento ISSN 0101 6113).

[34] XTAG RESEARCH GROUP. A Lexicalized Tree Adjoining Grammar for English. Pennsylvania: Institute for Research in Cognitive Science, University of Pennsylvania, 1995. 155p. (Technical Report IRCS 95-03).

[35] MANARIS, B. Z.; DOMINICK, W.D. NALIGE: a User Interface Management System for the Development of Natural Language Interfaces. International Journal of Man-Machine Studies, [S.l.], v.38, n.6, p.891-892, June 1993.

[36] BARBOSA, C. R. S. C. de. Sistema Especialista Pedagógico. Rio Claro: DEMAC-UNESP. 1992. 120p. Trabalho de Conclusão.

[37] OLIVEIRA, C. A. de. Interface em Linguagem Natural para Sistemas Especialistas. In: 2o SIMPÓSIO BRASILEIRO DE INTELIGÊNCIA ARTIFICIAL. 20-22 nov. 1985, São José dos Campos. Anais... São José dos Campos: 1985. p.173-177.

[38] DREW, C. F. A Natural Language Interface for Expert System Help Desks. In: CONFERENCE ON MANAGING EXPERT SYSTEM PROGRAMS AND PROJECTS, 1990, Bethesdas, MD. Proceedings... Bethesdas: 1990. p.209-215.

[39] BARBOSA, C. R. S. C. de.; OMAR, N. Detecção de Erros Durante um Diálogo Tutorial. In: WORKSHOP DE COMPUTAÇÃO, 3., 2000, São José dos Campos, SP. Anais... São José dos Campos: CTA/ITA, 2000. 164p. p.1-8.

[40] SCHABES, Y.; WATERS, R. C. Stochastic Lexicalized Context-Free Grammars. Broadway, Cambridge: Mitsubishi Electric Research Laboratories, July 1993. 12p. (Technical Report 93-12).