



**Universidade Federal do Ceará**

**CENTRO DE HUMANIDADES  
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGÜÍSTICA**

**KARINE ALVES DAVID**

**SINTAXE DAS EXPRESSÕES NOMINAIS NO PORTUGUÊS  
DO BRASIL: UMA ABORDAGEM COMPUTACIONAL**

**FORTALEZA-CE  
2007**

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

**KARINE ALVES DAVID**

**SINTAXE DAS EXPRESSÕES NOMINAIS NO PORTUGUÊS  
DO BRASIL: UMA ABORDAGEM COMPUTACIONAL**

Dissertação apresentada ao Programa de Pós-graduação em  
Linguística da Universidade Federal do Ceará –UFC, como pré-  
requisito parcial à obtenção do título de Mestre em Linguística.

**ORIENTADOR: PROF.DR. LEONEL DE ALENCAR ARARIPE**

**FORTALEZA-CE  
2007**

EXAME DE DISSERTAÇÃO

BANCA EXAMINADORA

**DAVID, K. A.** Sintaxe das expressões nominais no português do Brasil: Uma abordagem computacional. **2007. Dissertação de mestrado em Lingüística. Fortaleza, UFC, Programa de Pós-Graduação em Lingüística.**

Orientador: Prof. Dr. Leonel de Alencar Araripe  
Departamento de Letras Estrangeiras, UFC

Profa. Dra. Maria Denilda Moura  
Departamento de Lingüística, UFAL

Prof. Dra. Maria Elias Soares  
Departamento de Letras Vernáculas, UFC

Fortaleza,  
Fevereiro de 2007

Esta pesquisa foi financiada por uma bolsa CNPQ

*Você não sabe o quanto  
Eu caminhei  
Pra chegar até aqui  
Percorri milhas e milhas  
Antes de dormir  
Eu não cochilei*

*Os mais belos montes escalei  
Nas noites escuras de  
Frio chorei, ei , ei  
Ei ei ei..uu..*

*A vida ensina e o tempo  
Traz o tom  
Pra nascer uma canção  
Com a fé do dia-a-dia  
Encontro solução*

*(...)*

*Meu caminho só meu pai pode  
mudar  
Meu caminho só meu pai  
Meu caminho só meu Deus....  
Together...*

*(A estrada, Toni Garrido)*

Aos meus pais pelo apoio, carinho e incentivo e ao meu filho Pedro Hoto, razão principal de todos meus esforços,

Dedico

## AGRADECIMENTOS

A Deus, primeiramente, meu pai e amigo de todas as horas.

Ao meu orientador querido Prof. Dr. Leonel Alencar Araripe, pois, sem ele, esse trabalho não poderia ter nascido.

À minha querida professora e amiga Profa. Dra. Maria Elias Soares, por toda sua dedicação e carinho que sempre me deu e pelas horas em que dela precisei e sempre me atendeu com todo amor.

À Profa. Dra. Denilda Moura, por ter aceitado, com tanta dedicação, participar da minha banca.

À Profa. Dra. Mônica Magalhães Cavalcante, minha amiga que sempre me incentivou, me ajudou em tudo que precisei e compreendeu o meu amor pelo gerativismo.

Aos linguistas Eduardo Raposo, Ana Müller, Gabriel Othero, Arthur Pagani, Ana Castro, Marcelo Sibaldo pelo material bibliográfico a mim concedido e pela atenção que me deram.

A todos os meus colegas da turma de mestrado 2005-2007, pelas aulas tão animadas e construtivas, sobretudo à Josely Teixeira Carlos e Yvanthelmack Valério.

Aos meus irmãos, Renata, Raquel e André Luiz, por terem me apoiado sempre. Sobretudo à Raquel, pelas vezes que me ajudou com meu filho para eu poder ir às aulas do Mestrado e às orientações.

Mais uma vez ao meu pai Luiz, e à minha mãe Zenilda, por todo apoio financeiro, pela paciência e por cuidarem tão bem do meu filho Pedro Hoto.

Ao meu filho querido, agradeço, pela paciência, pelas vezes em que não pude dar-lhe a atenção necessária, por ter que estar escrevendo esta dissertação.

Ao meu amor, Rafael, pelo amor incondicional, pela paciência, pelo carinho e por ter me ajudado, ao final dessa dissertação, com seus conhecimentos gerativistas, traduções inesperadas e com a organização das “arvrinhas”.

Ao meu amigo, Prof. Dr. Clemilton Lopes Pinheiro, que, sem o seu incentivo desde 1997, jamais eu teria chegado até aqui.

Ao meu amigo Beto, pelo carinho de sempre e por ter me ajudado com o *abstract*.

Ao amigo Prof. Ms. Acrísio Senna pela sua contribuição para a minha disposição em fazer esse curso de mestrado.

Ao CNPQ pela bolsa que financiou esta pesquisa.

Às minhas tias Ilza e Icléia por sempre me darem forças e acreditarem na minha capacidade.



## RESUMO

Neste trabalho desenvolvemos um programa computacional na linguagem Prolog que foi capaz de analisar expressões nominais em língua portuguesa, atribuindo a elas sua estrutura de constituintes e sua representação através de colchetes rotulados e árvores. A análise feita nesta pesquisa se baseia em dados coletados no *corpus* NILC, a partir de 36 expressões nominais possivelmente geradas por falantes de língua portuguesa, dentre elas, expressões que trazem em sua estrutura interna pronomes pessoais assumindo o papel de determinantes que tomam sintagmas nominais enquanto complemento. Um outro objetivo dessa pesquisa foi testar hipóteses sobre o sintagma determinante (DP), especialmente a respeito do estatuto dos pronomes pessoais. Os resultados encontrados por nós comprovaram que os pronomes pessoais *eu*, *você*, *nós* e *vocês* são núcleos de um sintagma determinante que selecionam sintagmas nominais como complemento. Esta pesquisa, outrossim, contribui não só para preencher lacunas no âmbito da lingüística descritiva, mas também, para o desenvolvimento tecnológico do país, por ainda existir uma grande carência de pesquisas nessa área.

Palavras-chave: lingüística computacional, processamento computacional da linguagem natural, Prolog, gramática de unificação, DCG, LFG, teoria X-barras, DP, pronomes pessoais.

## ABSTRACT

In this work we developed a computational program in Prolog language able to analyze nouns expressions in Portuguese, attributing to them their constitutive structure and their representation through labeled clasps and trees. The analyzes done in this research is based in collected data from NILC *corpus*, from 36 possible nouns expressions created by Portuguese language speakers, in the midst of, expressions that brings in its internal structure personal pronouns assuming the role of determiners that took nouns phrase as complement. Another objective of this research was test hypotheses about determiner phrase (DP), especially concerning of personal pronouns statute. The results finding for us prove that the personal pronouns *eu*, *você*, *nós* e *vocês* are nucleus of determiner phrase that select noun phrases as complement. This research, moreover, add not only for fill blanks in descriptive linguistics ambit, but also, for the technological development of the country, because exist a large necessity for researches in this area.

Key-words: computacional linguistics, computacional process of natural language, Prolog, grammar unification, DCG, LFG, X-bar Theory, DP, personal pronouns.

# SUMÁRIO

<b>INTRODUÇÃO</b> .....	11
<b>CAPÍTULO 1: PRESSUPOSTOS TEÓRICOS</b> .....	15
1.1. Os três tipos de análise das expressões nominais na gramática gerativa.....	17
1.1.1. Análise gerativa tradicional .....	17
1.1.2. A teoria X-barra “clássica” .....	19
1.1.3. A teoria X-barra contemporânea .....	22
1.2. A teoria da checagem no Programa Minimalista: A concordância dentro do DP	28
1.3. Modelos computacionais do funcionamento da linguagem: DCG e LFG .....	33
1.4. Uma análise das expressões nominais do português segundo uma abordagem computacional.....	37
1.4.1. Construção de fragmentos computacionais em língua portuguesa no formalismo DCG em Prolog.....	37
<b>CAPÍTULO 2: HIPÓTESE E OPÇÕES METODOLÓGICAS</b> .....	42
2.1. Hipóteses .....	42
2.2. Opções Metodológicas .....	43
2.3. Descrição da análise .....	47
<b>CAPÍTULO 3: DESCRIÇÃO SINTÁTICA DA ESTRUTURA INTERNA DAS EXPRESSÕES NOMINAIS DO PORTUGUÊ DO BRASIL</b> .....	49
3.1. Os pronomes pessoais em língua portuguesa .....	49
3.1.1. A crítica de Monteiro (1994) à análise tradicional dos pronomes pessoais.....	50
3.1.2. Os pronomes pessoais no Programa Minimalista: Um exemplo de análise da expressão nominal em língua inglesa .....	51
3.1.3. Um estudo dos pronomes pessoais em língua portuguesa enquanto núcleo de um DP: Uma breve análise.....	52
3.2. As expressões nominais analisadas segundo a teoria X-barra contemporânea ....	55
3.3. Fragmento de gramática livre de contexto CFG.....	79

<b>CAPÍTULO 4: IMPLEMENTAÇÃO COMPUTACIONAL DE EXPRESSÕES NOMINAIS DO PORTUGUÊS DO BRASIL EM PROLOG.....</b>	<b>83</b>
4.1. Reconhecer gramatical do português no formato DCG sem traços.....	83
4.2. Analisador gramatical do português no formato DCG sem traços.....	89
4.3. Analisador gramatical do português no formato DCG com traços .....	99
4.4. Fragmento de analisador gramatical do português no formato DCG com traços	107
 <b>CONCLUSÃO.....</b>	 <b>112</b>
 <b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	 <b>114</b>

# INTRODUÇÃO

---

Creemos que, ao menos, duas afirmações acerca da Gramática Gerativa são consensuais para os lingüistas: a de que os enunciados efetivamente produzidos em uma língua são, ao menos parcialmente, determinados por estados mentais; e a de que a natureza desses estados mentais pode ser captada por sistemas computacionais, os quais podem ser entendidos como hipóteses explicativas para o conhecimento lingüístico dos falantes/ouvintes de uma dada língua. Conseqüentemente, o que buscam, em última instância, as pesquisas gerativistas é a construção dos mecanismos que simulem o conhecimento lingüístico de um falante de uma língua natural.

Esta pesquisa se insere nesse contexto. Aqui vamos discutir alguns dos dados teóricos da Gramática Gerativa para descrever/explicar um fenômeno da língua portuguesa: a estrutura interna das expressões nominais. Temos, pois, dois principais objetivos: a) discutir e rever alguns dos mecanismos envolvidos na estrutura das expressões nominais em português, e b) elaborar um fragmento de gramática computacional capaz de gerar, reconhecer e analisar um subconjunto das expressões nominais presentes nessa língua. Teoricamente, a discussão parte da teoria X-barra, postulada por Fukui (1986), Abney (1987) e Raposo (1992). Mas, recorre, também, a abordagens no âmbito da Gramática de Unificação, como os formalismos sintáticos da DCG (*Definite Cause Grammar*), conforme Blackburn *et al* (2001), e no âmbito da LFG (*Lexical-Functional Grammar*), de acordo com Bresnan (2001) e Falk (2001).

Partimos da hipótese DP, ou seja, do pressuposto de que a categoria D é um elemento funcional que seleciona um NP enquanto complemento, do mesmo modo que as outras categorias funcionais, como I (flexão) e C (complementador)<sup>1</sup>, de acordo com os princípios da teoria X-barra, a qual denominamos nesta pesquisa por “contemporânea”. Conforme essa hipótese, as expressões nominais são DPs (sintagmas de determinante), projeções da categoria D e não da categoria N. Em teorias X-barra anteriores, como a de Chomsky (1970), também adotada por outros estudiosos, como Mateus *et al.* (1989), Galves (1993), Lobato (1994), Devlin (2005) dentre outros, as expressões nominais são consideradas projeções da categoria N, constituindo, portanto, um NP (sintagma nominal).

A hipótese DP tem sido assumida por vários autores. Cerqueira (1993), por exemplo, enfoca o contraste entre os pronomes possessivos *seu* e *dele*, buscando determinar condições

---

<sup>1</sup> As abreviaturas I e C correspondem, respectivamente, a *Inflexion* e *Comp*, do Inglês.

de licenciamento dessas formas no NP. Para isso, busca uma relação entre os traços de concordância (AGR) no interior do DP e a presença de formas possessivas. No trabalho, esse autor considera uma versão mais articulada da estrutura DP proposta por Abney (1987). Já Castro (1997) utiliza-se da hipótese DP, em sua pesquisa, para trabalhar a categoria vazia no interior do sintagma determinante. Muller (1997) também aborda os pronomes possessivos no Português do Brasil (doravante PB), porém, deixa exposto, em sua tese, que não pretende encontrar a solução para a representação do DP ao estudar esses pronomes na Teoria Gerativa; pois, segundo ela, os sintagmas funcionais ainda são objeto de intenso debate. Magalhães (2004) argumenta, em sua pesquisa, a favor da valoração de traços de concordância dentro do DP em termos da operação *Agree*.

Dentre outros autores gerativistas, em nível internacional, que aderem também à hipótese DP, podemos citar Valois (1991), Ticio (2006), Coene e D'hulst (2003) e Herdan (2005).

Todos esses trabalhos nos foram importantes, na medida em que podemos observar como, a cada momento, os lingüistas gerativistas vêm aderindo, em suas pesquisas, a estudos relacionados à hipótese DP seguindo os pressupostos, principalmente, de Abney (1987).

Ainda dentro da discussão da hipótese DP, consideramos o que postula Radford (1997) no que diz respeito à classificação dos pronomes pessoais como núcleos de uma expressão nominal, os quais selecionam NPs enquanto complementos. Tradicionalmente, como pode ser visto em Cunha e Cintra (1985), os pronomes são tratados apenas como meros substitutos do nome. Mateus *et al.* (1989), por outro lado, que seguem a linha gerativista, propõem que os pronomes pessoais constituem NPs.

Entre as pesquisas que versam sobre essa análise, atualmente, para a língua portuguesa, já foram apresentadas, em nível de comunicação, de nossa autoria (DAVID, 2006a e 2006b), em que discutimos o pronome pessoal enquanto núcleo de um DP que toma NP como complemento. Alencar (2006b) também sugere que os pronomes pessoais são núcleo de sintagmas determinantes em Língua Portuguesa.

Em pesquisas anteriores, Radford (1997) investigou os pronomes pessoais em inglês e constatou que eles podem exercer o papel de núcleo de um DP. Do mesmo modo, fizemos, nesta pesquisa, uma revisão do conceito tradicional de pronome em língua portuguesa, analisando suas propriedades morfossintáticas, e verificamos que essa categoria, assim como no inglês, possui propriedades de seleção de NP, do mesmo modo como os artigos.

Realizamos um trabalho de investigação empírica da estrutura interna de 36 tipos de expressões nominais, explicadas pela hipótese DP, em dados do *corpus* NILC. E assim

podemos assumir, com os autores já citados, a hipótese de que os pronomes pessoais *eu, você, vocês e nós* em língua portuguesa são núcleos de um DP, que trazem um NP enquanto complemento.

Com base na teoria da Checagem de Chomsky (1995,1999), mas recorrendo também aos princípios de formalismos gramaticais de unificação, como a LFG, elaboramos regras para gerar, reconhecer e analisar estruturas de DP da língua portuguesa, acrescentando aos itens lexicais traços morfossintáticos que permitem a unificação de elementos dentro de uma determinada expressão. Alguns autores, como Raposo (1992), entendem que, para gerar um fragmento de gramática de estrutura sintagmática do português, faz-se necessário que se determinem regras capazes de gerar e reconhecer essas expressões nominais como bem formadas ou gramaticais. Assim é que Raposo (1992) apresenta regras de reescrita categoriais, as quais não conseguem assegurar que os traços gramaticais, que diferentes palavras trazem em sua estrutura, possam ser compatíveis com aqueles de outras palavras da mesma expressão. Portanto, em nosso trabalho, acreditamos que conseguimos preencher, pelo menos em partes, essa lacuna apresentada por essas regras de reescrita, que é a de hipergerar estruturas agramaticais. Assim, tornou-se indispensável, em nossa pesquisa, construirmos entradas lexicais que apresentassem, em sua estrutura interna, traços morfossintáticos, os quais representam as propriedades de cada categoria, não deixando de alterar, também, as regras sintáticas, de modo a torná-las sensíveis a tais propriedades.

Aqui, chegamos ao foco principal de nossa pesquisa, que, talvez, seja o que a diferencia de outras pesquisas de cunho gerativista sobre o fenômeno em questão. Tanto essas entradas lexicais quanto as regras sintáticas que combinam os elementos lexicais em sintagmas foram formuladas numa linguagem computacional. Defendemos que essa linguagem é mais eficiente e tem uma aplicação prática: fornecer subsídios para a construção de aplicativos para o processamento automático da língua portuguesa. A utilização de uma linguagem computacional para descrever uma língua natural não se circunscreve à ciência da computação. Também Schwarze (2001) defende um método de representação formal para a lingüística, baseado na linguagem Prolog<sup>2</sup>. A LFG constitui exemplo de teoria lingüística gerativa formulada numa linguagem computacional, de acordo com Bresnan (2001) e Falk (2001).

No que se refere aos trabalhos que assumem, no Brasil, essa abordagem computacional, podemos citar Alencar (2004, 2006a, b), Vieira e Lima (2001), Pagani (2004,

---

<sup>2</sup> Para um conhecimento aprofundado desse assunto ver Alencar (2006a).

2005), Othero e Menuzzi (2005) e Othero (2004, 2006). Esses autores fazem uso de programas computacionais para descrever vários tipos de estruturas da língua portuguesa. Nos seus trabalhos, encontramos a construção de minigramáticas do português do Brasil, elaboradas a partir de programas computacionais implementados principalmente na linguagem Prolog.

Acreditamos, desse modo, que a nossa pesquisa é bastante relevante para o desenvolvimento da ciência lingüística no Brasil, onde os estudos sobre processamento computacional da linguagem natural e engenharia da linguagem (*language engineering*) constituem área ainda incipiente. Graças a esse forte vínculo com a ciência da computação, nossa pesquisa se reveste, igualmente, de importância para o desenvolvimento tecnológico, na medida em que pretende fornecer subsídios para a elaboração de softwares lingüísticos de correção gramatical, de tradução automática e de ensino computacionalmente assistido de Língua Portuguesa.

Tudo isso se torna bastante importante, pois, apesar dos grandes avanços na área computacional, ainda encontramos sérios problemas de difícil solução nessa área, como, por exemplo, a baixa qualidade fornecida por corretores textuais e a dificuldade de tradução automática.

Todo esse trabalho teórico-analítico com as expressões nominais em português, explicadas pela hipótese DP, está textualizado, nesta dissertação, da seguinte forma: Inicialmente, apresentamos um capítulo de natureza mais teórica. Realizamos, numa primeira parte, uma retrospectiva da teoria gerativa no que diz respeito especificamente aos tipos de expressões nominais, e uma breve exposição sobre a teoria da checagem, no Programa Minimalista. Numa segunda parte, fazemos uma também breve exposição sobre os modelos computacionais com os quais trabalhamos, e sobre a construção de fragmentos computacionais em Língua Portuguesa. O segundo capítulo apresenta os aspectos metodológicos. No terceiro capítulo, realizamos a revisão e a discussão dos mecanismos envolvidos na estrutura das expressões nominais em Português, a partir dos 36 tipos de expressões encontradas no *corpus* NILC. Finalmente, no quarto e último capítulo, mostramos, passo a passo, a construção de um fragmento computacional capaz de gerar, reconhecer e analisar alguns dos tipos de expressões nominais em língua portuguesa, a partir da hipótese DP.



# CAPÍTULO I

## PRESSUPOSTOS TEÓRICOS

---

De acordo com Alencar (2006b, p.3), o termo *gramática* ou *sintaxe* gerativa, em sentido estrito, designa a abordagem de Chomsky, lançada na década de 50 e que passou por sucessivos aperfeiçoamentos, sendo os mais recentes a Teoria de Princípios e Parâmetros (ou Teoria da Regência e da Ligação – TRL), surgida nos anos 80, e o atual Programa Minimalista, lançado em 1995. A principal característica da TRL é a postulação de uma Gramática Universal (*Universal Grammar – UG*), constituída de princípios, comuns a todas as línguas, geneticamente determinados. Um desses princípios universais é o da dependência estrutural, que afirma que todas as operações gramaticais são caracterizadas estruturalmente (RADFORD, 1997, p.12).

Seguindo na esteira de Chomsky, muitos lingüistas computacionais adotam a concepção de gramática como dispositivo para a geração e análise de um número infinito de estruturas e assumem a hipótese do inatismo, mas rejeitam a postulação de transformações. Esse é o caso de muitos dos principais representantes da LFG (Gramática Léxico-Funcional) como Bresnan (2001) e Falk (2001). Uma definição precisa para o termo transformação é dada por Alencar (2006, p. 6) de que “transformações são operações que derivam uma estrutura sintagmática a partir de uma outra estrutura sintagmática já formada. Tipicamente a transformação de uma estrutura em outra resulta do movimento de constituintes”.

Em nossa pesquisa, nos deteremos a trabalhar com alguns aspectos de teorias não-transformacionais, também, conhecidas por teorias não-derivacionais ou teorias lexicalistas baseadas na unificação (*unification-based*) ou baseadas em restrições (*constraint-based*) (ALENCAR, 2006b, p.7).

Não achamos conveniente fazer, aqui, todo um percurso histórico da teoria lingüística, desde a gramática tradicional até à teoria a qual nos propomos a assumir para este trabalho. Assim, a seguir, iremos apresentar um quadro resumido proposto por Klenk (2003), o qual demonstra detalhadamente um breve histórico que parte da gramática tradicional, passa pelos estruturalismos europeu e norte-americano, e, em seguida pela gramática gerativa a qual se divide em gramáticas transformacionais e não-transformacionais, conforme o exposto acima.

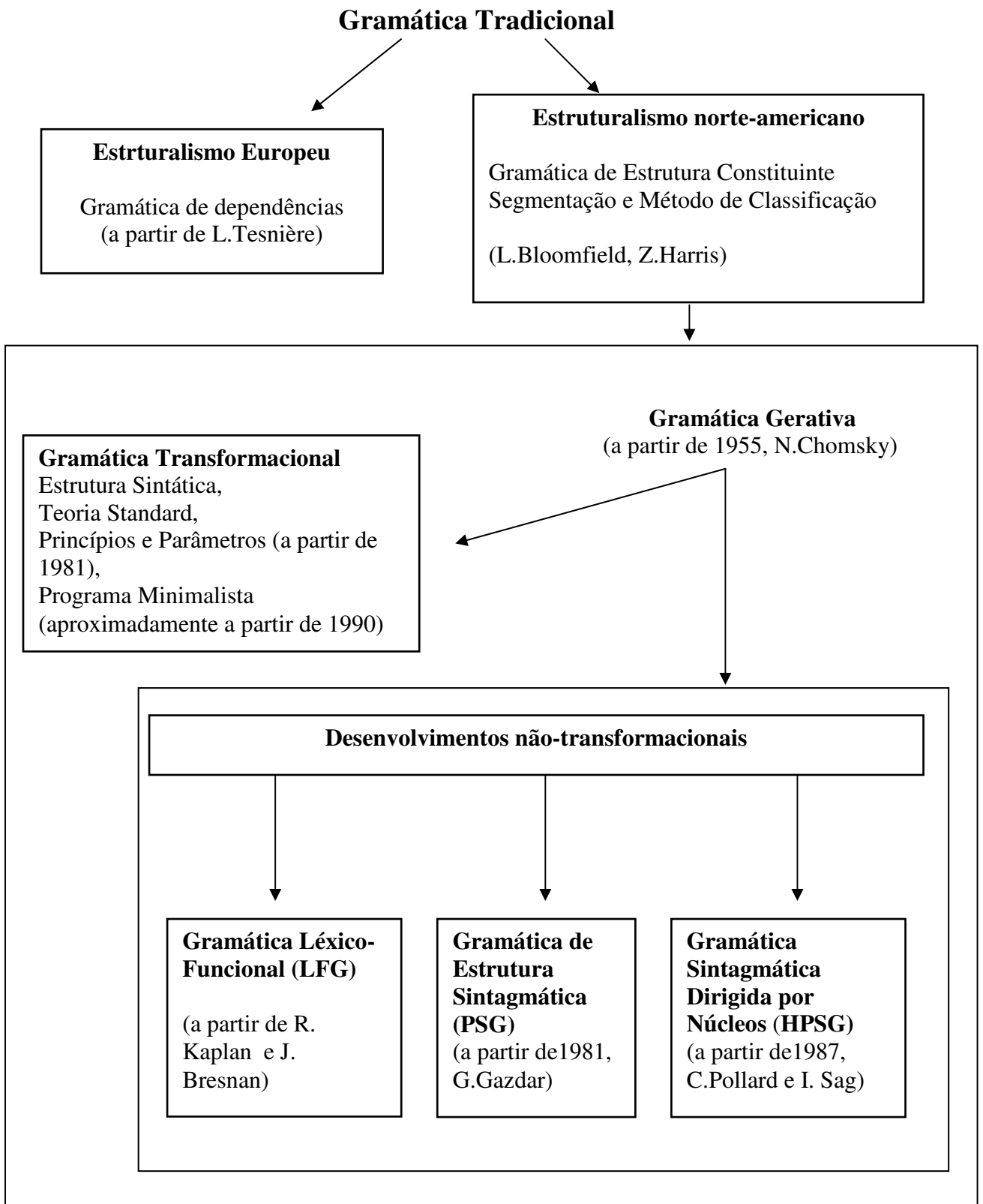


Figura 1. Breve Histórico da Gramática Gerativa

## 1.1. Os três tipos de análise das expressões nominais na gramática gerativa

### 1.1.1. Análise gerativa tradicional

As categorias gramaticais de um língua são divididas em dois tipos: categorias lexicais e categorias sintagmáticas. As categorias lexicais são consideradas o centro das categorias sintagmáticas.

Segundo Raposo (1992), já que se estabelece que as frases estão organizadas a partir de estruturas constituintes, para isso é necessário que se determinem quais regras permitem gerar estruturas e à qual natureza pertencem.

Em modelos anteriores à TRL, caracteriza-se a estrutura de constituintes por um conjunto de regras que podem ser chamadas de algumas maneiras, como: regras de reescrita categorial, ou somente regras categoriais e, ainda, por regras sintagmáticas. Em (2), apresentamos o exemplo de uma minigramática do português do Brasil capaz de gerar e reconhecer a sentença em (1):

(1) Um jogador de xadrez venceu uma partida em Berlim.

(2) Fragmento de gramática de estrutura sintagmática do português:

$S \rightarrow NP \ VP$

$NP \rightarrow (D) \quad N \quad (PP)$

$VP \rightarrow NP \quad (PP)$

$PP \rightarrow P \quad NP$

$N \rightarrow$  jogador, xadrez, partida, Berlim ...

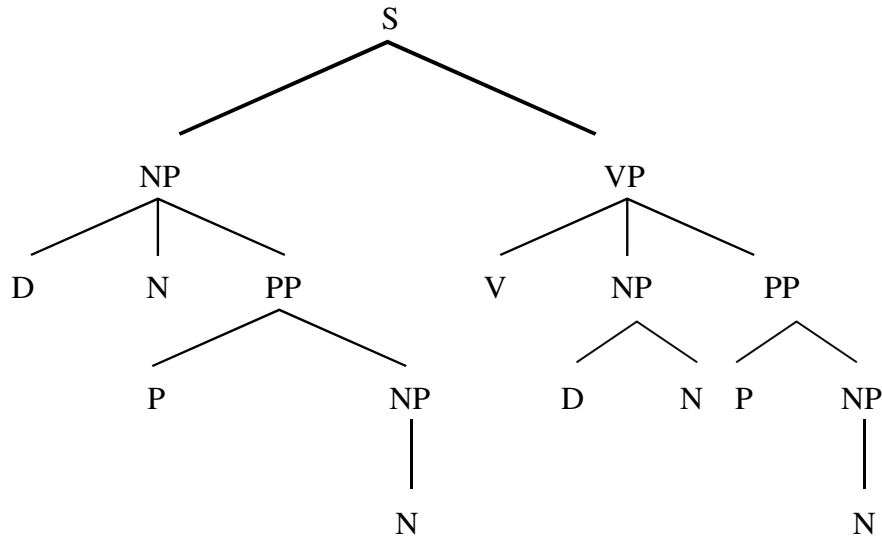
$V \rightarrow$  venceu

$P \rightarrow$  em, de, ...

$D \rightarrow$  um, uma, o, ...

Vejamos, a seguir, a representação arbórea em (3) desta sentença:

(3)



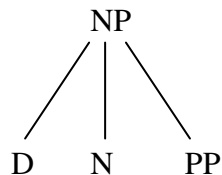
As regras de (2) podem gerar algumas estruturas abstratas que representam estruturas sintáticas, através da introdução de nódulos categoriais que introduzem, em seguida, os itens lexicais.

Conforme Raposo (1992), há uma estreita relação entre as categorias lexicais principais, que são N (substantivo ou nome), P (preposição), A (adjetivo), Adv (advérbio) e V (verbo), e as categorias sintagmáticas NP, PP, AP, AdvP e VP, uma vez que as primeiras são consideradas elementos centrais das últimas. Desse modo, em uma determinada estrutura XP qualquer, onde X é uma variável representando uma das categorias lexicais, o elemento principal ou nuclear será X.

Em (4a) apresentamos o exemplo de como regras categoriais vão permitir gerar configurações sintáticas como (4b):

(4) a.  $NP \rightarrow D N PP$

b.



Raposo (1992) chama esta teoria de “teoria da estrutura de constituintes”. De acordo com ela, e pela representação ilustrada em (4a), constatamos que a ramificação não é necessariamente binária, mas pode ser, também, ternária, e que há pouca estruturação interna, ou seja, os três filhos do nódulo NP encontram-se no mesmo nível hierárquico, e existem

apenas dois níveis hierárquicos: o da categoria lexical com os elementos que a modificam e o da categoria sintagmática.

### 1.1.2. A teoria X-barra “clássica”

A teoria X-barra, objeto da presente seção, foi postulada pela primeira vez “no âmbito da gramática generativa em Chomsky (1970), mas com antecedentes em vários trabalhos do estruturalismo norte-americano” (RAPOSO, 1992, p.159). Trata-se de um dos componentes do modelo gerativo conhecido como Teoria da Regência e da Ligação ou modelo de Princípios e Parâmetros.

Um dos pontos mais discutidos dentro dessa teoria é o que conhecemos por princípio da endocentricidade. De acordo com esse princípio, uma determinada categoria sintagmática XP, por exemplo, deve ter obrigatoriamente um núcleo, no caso X, o qual pertence a uma categoria lexical principal. Assim, o núcleo de uma categoria como NP, não pode ser V, o núcleo de PP, não pode ser N, e assim sucessivamente. Segundo Raposo (1992, p.166), o núcleo de um XP é “a categoria lexical X imediatamente dominada por XP, i.e. [X, XP].” Por essa notação, o nó X é filho de XP.

Por meio do princípio da endocentricidade, que é considerado por Raposo como uma condição de boa-formação (*well-formedness condition*) sobre as regras de um componente gramatical, formulado em (5), são excluídas automaticamente regras do tipo (6):

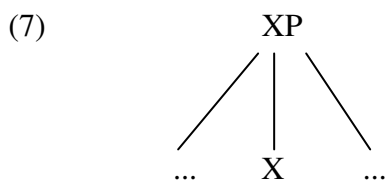
(5)  $XP \rightarrow \dots X \dots$

(6) (i)  $NP \rightarrow \dots V \dots$

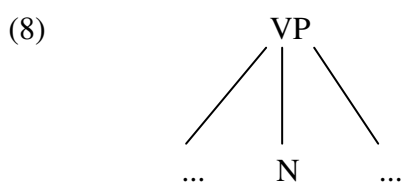
(ii)  $PP \rightarrow \dots N \dots$

(iii)  $AP \rightarrow \dots P \dots$

A regra de (5) gera a configuração abstrata de (7):



Do mesmo modo, a configuração em (8) jamais seria permitida por esta teoria:



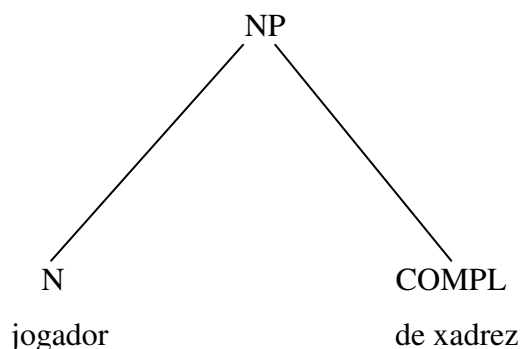
A seguir, apresentaremos algumas das propriedades centrais da versão da teoria X-barra “clássica” desenvolvida por Chomsky (1970).

Para Chomsky (1970) a tese central da teoria X-barra é a de que, em vez de apenas uma projeção da categoria lexical principal, agora seriam duas projeções sucessivas  $X'$  e  $X''$ <sup>3</sup>, em que o número de apóstrofes representa o número de barras. Além disso, em vez de dois níveis hierárquicos seriam três níveis hierárquicos:  $X^0$ ,  $X'$  e  $X''$ . Há, assim, uma forte hierarquização entre os constituintes. Finalmente, a ramificação é estritamente binária nessa abordagem, diferentemente do modelo gerativo tradicional, apresentado na seção anterior 1.1.1.

A proposta de Chomsky é de que os elementos que modificam as categorias lexicais passem a pertencer a duas classes diferentes: a dos especificadores e a dos complementos.

Para ele a projeção  $X'$  é a projeção sintática de nível 1, que é obtida por meio da composição da categoria lexical com os seus complementos. Observemos os exemplos (9) e (10).

(9) jogador de xadrez



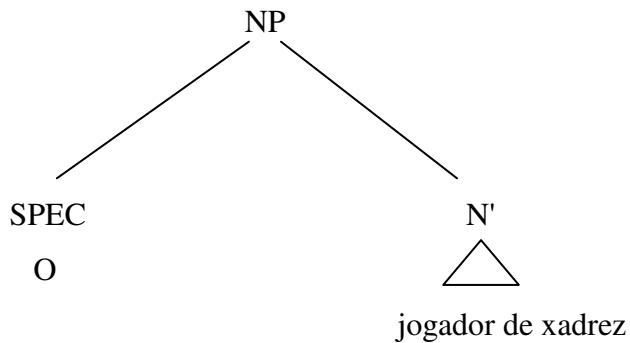
(10)  $N' \rightarrow N$  (COMPL)

Já a projeção  $X''$  é definida como a projeção sintática de nível 2, que é obtida através da composição da categoria  $X'$  (que resulta da primeira projeção) com o especificador da categoria lexical. Esta representação está descrita em (11) e (12):

---

<sup>3</sup> Lê-se "X barra" e "X duas barras", respectivamente

(11) o jogador de xadrez

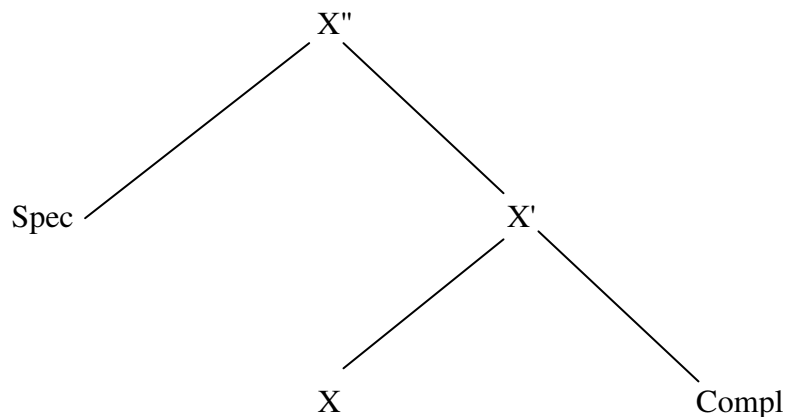


(12)  $N'' \rightarrow (\text{SPEC}) N'$

Chomsky (1970) define *complemento* (Compl) como um constituinte que é subcategorizado por itens que fazem parte de uma das categorias lexicais  $X^0$ , como, por exemplo, N, P, A e V. Já o *especificador* (Spec), segundo esse autor, é um modificador não subcategorizado das categorias lexicais e ocorre em português à esquerda do núcleo central.

Em (13), apresentaremos a configuração sintática de *o jogador de xadrez*, conforme analisado segundo a teoria X-barrado de Chomsky:

(13) O jogador de xadrez



Raposo (1992) segue a convenção da teoria gerativa tradicional de representar a projeção máxima de uma categoria lexical (i.e. de um X qualquer) como XP. Assim, temos  $NP = N''$ ,  $VP = V''$ ,  $AP = A''$  e  $PP = P''$ .

Quanto ao *núcleo*, Chomsky apresenta a possibilidade de que, em uma teoria X-barrado com duas projeções, é possível definir duas noções de núcleo: A primeira é que o núcleo de  $X'$  é a categoria lexical  $X^0$  que é imediatamente dominada por  $X'$ ; a segunda é que o núcleo de

X" é a categoria X' imediatamente dominada por X". O autor considera, igualmente, que a categoria lexical X° é, também, o núcleo da projeção máxima X".

Um outro princípio que Chomsky adota para essa teoria é o princípio da *uniformidade*, através do qual nos diz que as únicas categorias que podem projetar, ou seja, que obedecem ao princípio da endocentricidade, são as categorias lexicais principais: V, N, A e P e que as categorias menores D, Q etc., por outro lado, funcionam apenas como meros especificadores das categorias lexicais principais. Na teoria a qual chamaremos de “contemporânea”, mais adiante, veremos que Fukui (1986, p.49) vai de encontro a este princípio de Chomsky, no que diz respeito a alguns pontos, pois este autor considera determinadas categorias “menores” como elementos que projetam em uma determinada estrutura, como é o caso dos determinantes (D), que projetam também, conforme veremos na seção 1.1.3.

Outro aspecto importante da teoria X-barras de Chomsky (1970) é que o número máximo permitido de barras é o mesmo para todas as categorias que projetam. É através deste aspecto que uma teoria X-barras pode ser chamada de *uniforme*; mas a característica de uniformidade não é, como o princípio da endocentricidade, uma característica necessária para todas as teorias X-barras.

Outra restrição colocada por Chomsky (1970) é referente ao número máximo de barras que pode ser projetado. Para ele, apenas duas barras podem ser licenciadas, ou seja, todas as categorias projetam o nível X' (associado aos seus complementos) e o nível X" (associado a X' com um especificador). Entretanto há quem não aceite esta restrição, como Jackendoff (1977, *apud* RAPOSO, 1992)<sup>4</sup>, que propõe uma teoria de projeção de três barras. Fukui (1986) concorda com a restrição proposta por Chomsky, no que diz respeito à quantidade do número de barras permitido, mas não segue o princípio da uniformidade, proposto por ele.

### **1.1.3. A teoria X-barras contemporânea**

Em nossa pesquisa, partiremos do pressuposto de Abney (1987, p.38), que atribui ao determinante um tratamento de categoria funcional, assim como ao C e I, que, também, têm sua própria projeção. Assim, o NP passa a ser projetado como parte dessa estrutura maior. De acordo com esta configuração de Abney, o D seleciona projeção num processo semelhante à seleção realizada pelas categorias funcionais C e I.

---

<sup>4</sup> JACKENDOFF, R.S, *X' Syntax: A Study of Phrase Structure*. Cambridge: The MIT Press, 1977.



“Fukui (1986, p.50) propõe, também, uma teoria X-barra em que somente as categorias funcionais projetam o nível X" através da combinação da projeção X' com um especificador. Já as categorias lexicais, pelo contrário, não possuem especificadores e, desse modo, somente projetam o nível X'.

Para Fukui (1986, p.50), as categorias funcionais apresentam um conjunto de traços que as vão diferenciar das categorias lexicais. Vejamos estas propriedades: (i) possuem uma, e somente uma, posição de especificador; (ii) formam classes fechadas, e restritas, de elementos; (iii) não possuem o valor semântico normalmente associado às categorias lexicais, (iv) subcategorizam um, e somente um, complemento. Por sua vez, as categorias lexicais não possuem especificador, formam classes abertas, e têm (na maioria das vezes) valor semântico.

Um dos postulados de Fukui (1986, p.51), feito também, simultaneamente por Abney (1987, p.40) é o de que a categoria D projeta, do mesmo modo que as outras categorias, de acordo com os princípios da teoria X-barra (esta idéia é conhecida como a hipótese DP). De acordo com esta hipótese, as expressões nominais de uma língua, segundo Raposo (1992, p.210):

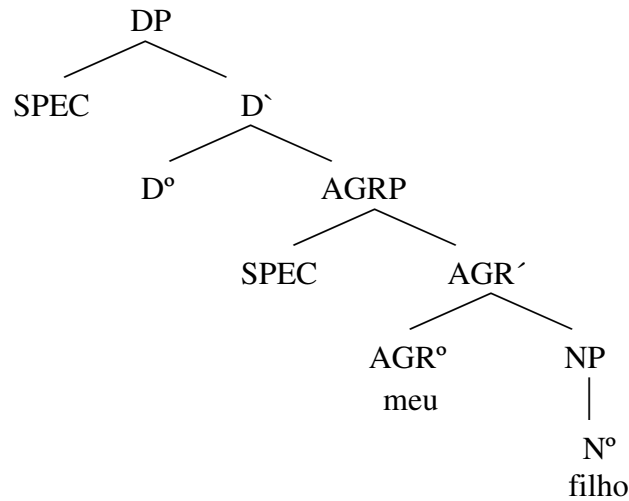
são, na realidade, Grupos de Determinantes (DPs, no inglês *Determiner Phrase*) projeções da categoria D e não da categoria N. O papel reservado à projeção de N (ou seja, NP) é o de complemento de D, da mesma forma que VP é complemento de I e IP é o complemento de C.

A respeito da classe dos determinantes, Mateus *et al* (1989) consideram-na como uma classe limitada que precede o nome, e tem como base determinar este semanticamente. Segundo estas autoras, os determinantes podem ser os artigos e adjetivos determinativos (demonstrativos, possessivos e indefinidos).

Cerqueira (1996) postula que, quando há co-ocorrência de um determinante e um possessivo em língua portuguesa, há a possibilidade de haver duas posições distintas: o elemento D° e AGR nominal, onde estará o possessivo.

Desse modo, uma estrutura como *o meu filho*, de acordo com Cerqueira (1996, p.144), pode ser representada do seguinte modo:

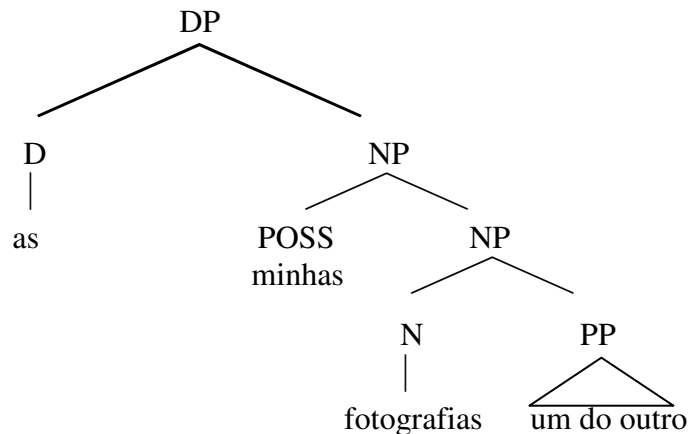
(14) o meu filho



O pronome possessivo está, desse modo, à direita de N°, onde recebe papel temático, e, depois, move-se para SPEC para receber Caso.

Por outro lado, Raposo (1992, p.265) aborda a questão dos possessivos em língua portuguesa da seguinte maneira:

(15) as minhas fotografias um do outro



E diz o seguinte:

Não analisamos o Agente como um especificador do DP na sua posição Spec pelas seguintes razões. Em primeiro lugar, o possessivo *minhas* ocorre à direita da categoria D (*as*), e não à esquerda, contrariamente ao que seria de esperar caso ocupasse essa posição. Em segundo lugar, os determinantes em Português não atribuem Caso<sup>5</sup>, logo, não licenciam uma categoria na sua posição Spec. (RAPOSO,1992, p.265)

<sup>5</sup> Em Raposo (1992) encontramos “kaso” grafado com “k”, já que este baseia-se em Fukui (1986), que utiliza, também, para o inglês *kase*.

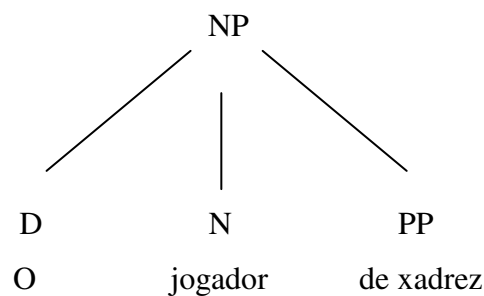
Na moderna abordagem gerativista, como podemos constatar, os pronomes possessivos diferem radicalmente dos artigos, pois não são considerados como determinantes.

Assim como Raposo (1992) e Cerqueira (1996), assumimos, em nossa pesquisa, que os pronomes possessivos não são colocados no mesmo plano dos pronomes pessoais e demonstrativos, que, para nós, podem ser considerados determinantes e núcleos de um DP, diferentemente do que diz, também, a tradição gramatical, que, na verdade, não faz menção nem a essa classe conhecida no âmbito da linha gerativista como determinante.

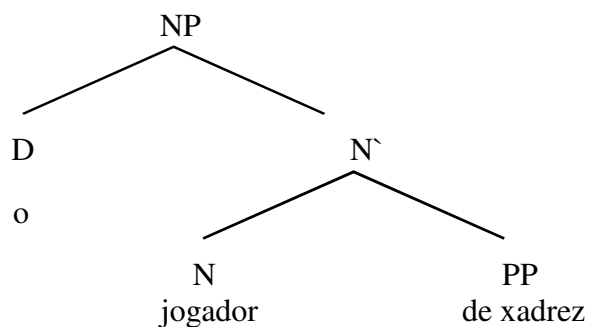
A seguir, iremos comparar a estrutura da expressão nominal *O jogador de xadrez*, levando em conta os três tipos de análise das expressões nominais na gramática gerativa propostos por nós:

(16) o jogador de xadrez

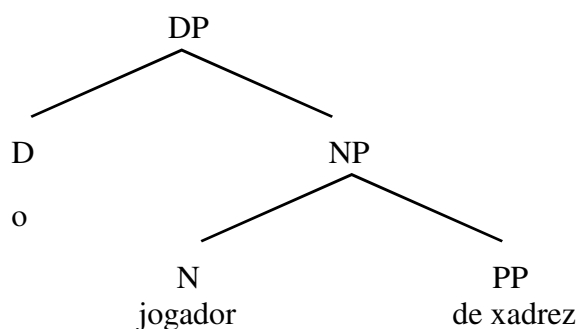
a) Análise gerativa tradicional



b) Teoria X-barras "Clássica"



## Teoria X-barras contemporânea



Não obrigatoriamente deverá haver a presença de um especificador na estrutura sintática das categorias funcionais, e a questão da presença ou ausência destes, depende de propriedades dos itens funcionais particulares. Assim, para cada categoria funcional C, I e D, os itens que a compõe dividem-se em dois grupos: aqueles que permitem a ocorrência de um especificador, e aqueles que não permitem a presença de tal elemento.

De acordo com Raposo (1992) é propriedade relevante para Fukui (1986)<sup>6</sup> a possibilidade de atribuição de Kaso à posição SPEC, por cada núcleo particular. Desse modo, “os núcleos funcionais, capazes de atribuir Kaso, permitem um especificador (ao qual atribuem Kaso), ao passo que os núcleos funcionais que não têm capacidade de atribuir Kaso não permitem um especificador” (RAPOSO, 1992, p.210).

Assim, para Raposo (*op.cit.*), um especificador encontra-se sujeito ao seguinte princípio: “Um especificador é, necessariamente, licenciado por um Kaso atribuído à posição SPEC”. Em inglês, teremos o caso do D que é capaz de licenciar um especificador (de categoria DP), o qual funciona como sujeito do DP. Esse D é o elemento possessivo ‘s:

- (17)            John's picture of Mary  
                  de João o retrato da Maria  
                  'o retrato que o João tirou da Maria'

No entanto, Raposo (1992, p.212) afirma o seguinte: “Em português, não existe nenhum D com as propriedades deste elemento. Nesta língua, a classe dos determinantes (o, este, aquele etc.) reduz-se a elementos que não licenciam especificadores” (RAPOSO, 1992, p.212).

---

<sup>6</sup> Raposo (1992) não cita em seu livro Fukui (1986) mas um artigo de Fukui em co-autoria com Speas publicado, também, em 1986. Ao que tudo indica, esse artigo é baseado na tese de Fukui, publicada anteriormente, no mesmo ano. Porém, consultamos para esse trabalho a tese de Fukui (1986) por ser mais complexa e por não termos conseguido acesso ao artigo supracitado.

A diferença entre o português e o inglês quanto ao licenciamento de um especificador pela categoria D reflete-se na impossibilidade de traduzir literalmente para o português o exemplo (17).

Para Fukui (1986, p.59), quando não há o licenciamento de um especificador por uma determinada categoria funcional, o nível máximo projetado é X', contrariando, desse modo, a posição de Chomsky (1986). Para esse autor a projeção máxima de um núcleo particular é aquela que possui o maior número de barras possível para esse núcleo particular, assim, uma projeção máxima pode conter apenas uma barra (quando a projeção de um núcleo funcional for sem Kaso ou de uma categoria lexical) ou duas barras, se é a projeção de um núcleo funcional com Kaso.

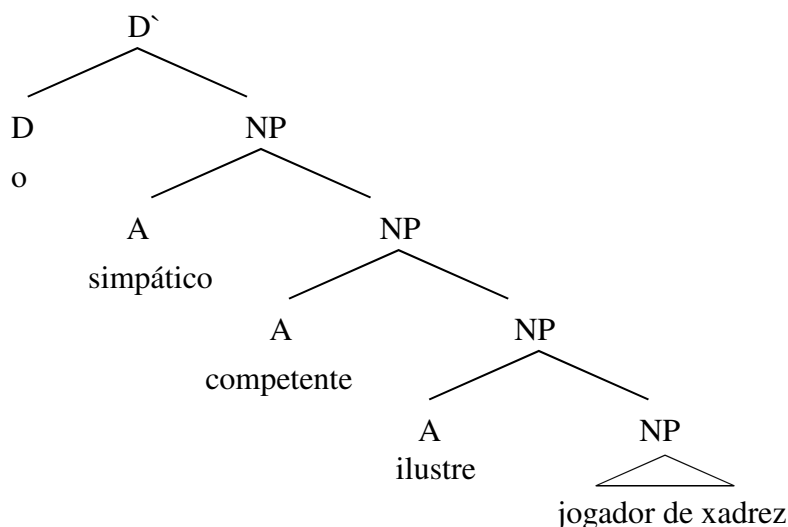
A notação CP, IP e DP é usada para as projeções máximas nesse sentido, independente do número de barras. Para distinguirmos uma projeção X' de uma X" utilizaremos o conceito de *projeção máxima fechada* que é “uma projeção de duas barras (isto é, contendo um especificador)” (RAPOSO, 1992, p.214).

Como já apresentamos no início desta seção, ao contrário das categorias funcionais, as lexicais têm um valor semântico e formam classes abertas. “As categorias lexicais não tomam especificador, projetando unicamente o nível X' por combinação com os seus complementos subcategorizados (X' é pois a projeção máxima para estas categorias)”(RAPOSO, 1992, p.215).

Contudo, para Fukui (1986, p.50), outros elementos podem modificar a categoria lexical, não através da “projeção de um nível adicional da teoria X-barra”, mas sim por uma configuração de adjunção. Desse modo, algumas categorias eram apresentadas, em versões anteriores da teoria X-barra, como especificador, mas, na verdade, são reinterpretadas como adjuntos, como, por exemplo, alguns advérbios e adjetivos. Fukui (*op.cit.*) chama esses especificadores de “pseudo-especificadores”.

Na análise da representação arbórea em (18), baseamo-nos em Raposo (1992) que exemplifica o processo de adjunção sugerido por Fukui (1986). Como podemos ver nesta árvore, os adjetivos não funcionam como especificadores, uma vez que são filhos e, ao mesmo tempo, irmãos de NP (= N'). Note-se que os adjetivos podem ser reiterados, o que não seria possível caso fossem especificadores.

(18) O simpático, competente, ilustre jogador de xadrez.<sup>7</sup>



## 1.2. A teoria da checagem no Programa Minimalista: A concordância dentro do DP

Para elaborarmos um fragmento de gramática de estrutura sintagmática do PB faz-se necessário que se determinem regras capazes de gerar e reconhecer essas expressões nominais como bem formadas ou gramaticais. Para isso, Raposo (1992) apresenta regras de reescrita categoriais, cuja definição já foi dada, também, na seção 1.1.1.

Entretanto, essas regras de reescrita não conseguem assegurar que os traços gramaticais que diferentes palavras trazem em sua estrutura possam ser compatíveis com aqueles de outras palavras da mesma expressão. Assim, podemos dizer que esse modelo de “Regras de Reescrita” apresenta um problema de hipergeração de estruturas agramaticais. Para que isso não ocorra, Chomsky (1995,1999) propôs a Teoria da checagem.

Vejamos, a seguir, o que acontece quando se pensa em gerar estruturas, levando em conta apenas as regras de reescrita:

<sup>7</sup> Exemplo baseado em Raposo (1992)

(19)

a) Léxico

D → ela, uma, as, umas, um, nós, você, esta etc.

N → Maria, mecânico, homem, mecânicos, mulher etc.

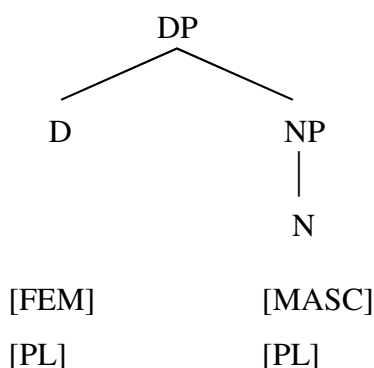
b) Sintaxe

DP → D NP

NP → N

Através da minigramática em (19) são geradas estruturas do tipo: *Uma mecânicos*, *Ela Maria*, *Esta homem*, entre outras, pois o que falta nessa gramática é o acréscimo de traços morfossintáticos aos itens que compõem o léxico, como, por exemplo, na estrutura abaixo:

(20) \*As mecânicos



O traço de gênero do D não concorda com o do NP, ou seja, a estrutura é agramatical, pois, ao checar os traços morfossintáticos, não houve concordância entre eles.

Segundo Radford (1997), para que sintagmas e sentenças sejam formados, é necessário que as palavras combinem-se umas às outras. Os sintagmas e sentenças são formados por uma operação de junção binária que combina pares de categorias para formar estruturas cada vez maiores, podendo essas estruturas ser representadas através de diagramas arbóreos.

As propriedades gramaticais de uma palavra podem ser descritas por meio de traços. No Programa Minimalista, esses traços podem ser classificados em três tipos: i) traços nucleares (os quais descrevem as propriedades gramaticais intrínsecas das palavras); ii) traços de complemento (os quais descrevem os tipos de complemento que as palavras permitem) e iii) traços de especificador (os quais determinam os tipos de especificador/sujeito que essas palavras permitem).

Para garantirmos que os traços gramaticais portados por diferentes palavras numa sentença ou sintagma sejam compatíveis com aqueles das outras palavras que fazem parte da mesma sentença, ou do mesmo sintagma, podemos nos utilizar da teoria da checagem proposta por Chomsky (1995,1999).

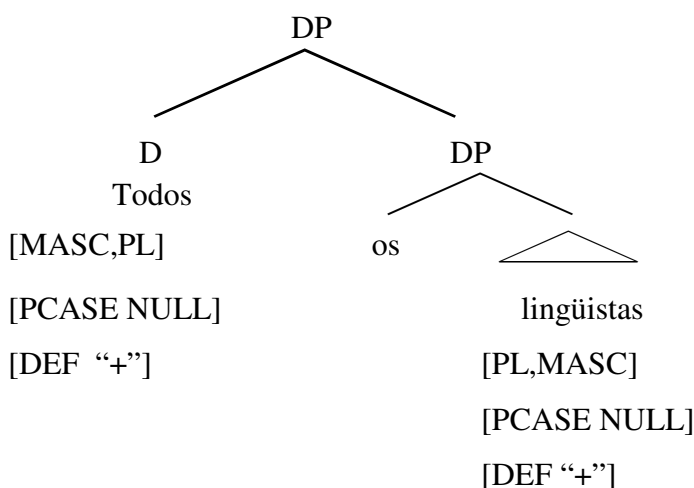
Radford (1997) postula que as estruturas sintáticas geradas por junção de pares de categorias são usadas como base para computar dois tipos de representação estrutural para uma sentença: uma representação da sua forma fonética (PF) e uma representação da sua forma lógica (LF). A representação PF diz respeito a como a sentença é pronunciada e a representação LF descreve aspectos lingüísticos do seu significado.

Outra informação importante é a de que cada item lexical possui traços formais interpretáveis (por exemplo, plural e singular) e não-interpretáveis (por exemplo, nominativo e acusativo). Os traços não-interpretáveis de uma determinada sentença ou expressão precisam ser eliminados quando são computados em LF para que a derivação seja convergente. Os traços interpretáveis, por sua vez, não precisam ser checados porque eles são lidos por LF sem problemas. Alguns traços gramaticais são interpretáveis (em LF) por terem conteúdo semântico, enquanto que outros são não-interpretáveis (em LF) por não conterem nenhum conteúdo desse tipo. Representações do tipo LF devem conter somente traços interpretáveis (semanticamente).

Entretanto, se a derivação gerar uma representação em LF contendo um ou mais traços não-interpretáveis, esta será considerada uma combinação falha, e as sentenças ou sintagmas correspondentes serão agramaticais ou mal-formados.

Vejamos como podemos aplicar o mecanismo de checagem de traços gramaticais em algumas expressões nominais da língua portuguesa:

(21) Todos os lingüistas.

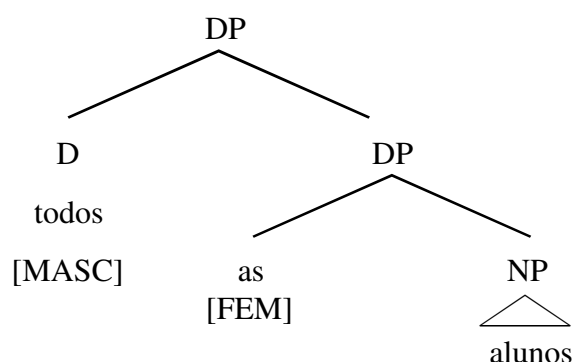




Assumindo a hipótese DP, tal como descrita por Abney (1987), podemos dizer que em (21) *todos* é um determinante com traços de núcleo MASC e PL e traços de complemento MASC, PL, PCASE NULL (o complemento não pode ser introduzido por preposição) e DEF (+) (i.e., o valor para o parâmetro definido deve ser positivo, o que explica a agramaticalidade de “*todos uns alunos*”, por exemplo). A construção (21) é considerada, assim, licenciada porque os traços de núcleo do complemento “*os lingüistas*” são exatamente os exigidos pelo determinante (D) “*todos*”, ou seja, MASC, PL, PCASE NULL e DEF (+).

Vejamos, agora, um outro exemplo:

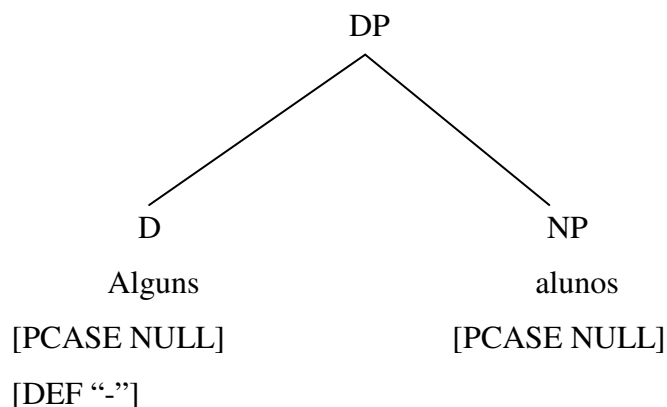
(22) \* *Todos as lingüistas*



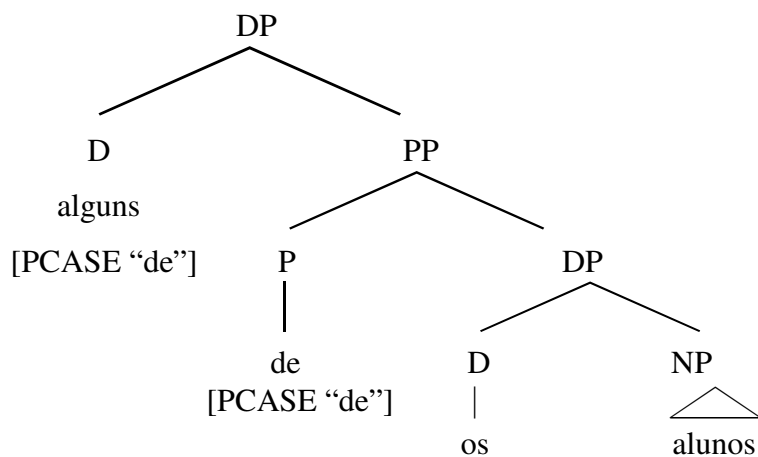
Esta combinação, representada por (22), é falha, pois *todos* possui um traço de núcleo MASC e o seu complemento, desse modo, deverá apresentar, também, esse traço. O que vemos nessa expressão é que no complemento de *todos* há um determinante cujo traço de gênero é FEM, que é um traço interpretável, não sendo compatível com o do núcleo que é MASC. Essa expressão não é licenciada por conta dessa incompatibilidade de traços de gênero.

Agora, observemos as expressões a seguir:

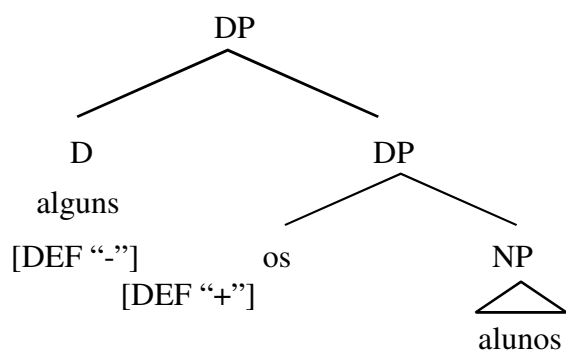
(23) Alguns alunos<sup>8</sup>



(24) Alguns dos alunos<sup>9</sup>



(25) \*Alguns os alunos



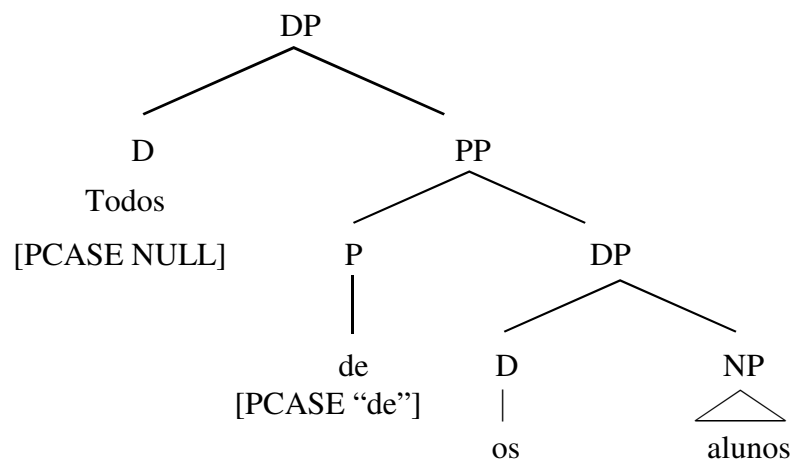
<sup>8</sup> Para o demonstrativo "alguns", assumiremos duas entradas lexicais: Na primeira entrada admite-se o uso da preposição, e, na segunda, não se admite.

<sup>9</sup> Veremos no capítulo 3 desta dissertação, que será revisto o tratamento atribuído às preposições, já que assumimos um formalismo, a DCG, o qual não permite transformações, nem movimentos.

Se observarmos bem, *alguns* tanto em (23) quanto em (24) e (25) é um determinante que possui traços de núcleo MASC e PL. Porém no que diz respeito aos traços de complemento, temos uma diferença entre essas três expressões nominais. Em (23) os traços de núcleo são MASC, PL, PCASE NULL e DEF (-), pois se tornaria agramatical uma expressão como, por exemplo, *alguns uns alunos*. Mas, através do exemplo (25), vemos que, também, é agramatical tal representação, comprovando-se, assim, que *alguns*, também, demanda um traço interpretável de complemento DEF (-). Desse modo, tanto (23) quanto (24) são construções licenciadas; enquanto que (25) é considerada uma representação falha porque o complemento é DEF (+).

Fazendo uma comparação entre os traços de *alguns* e *todos* que, como sabemos, fazem parte da mesma categoria (isto é, D), observamos que aquele, diferentemente deste, é compatível com o uso da preposição *de*. Para isso, observemos a agramaticalidade do exemplo abaixo:

(26) \* Todos dos alunos.



Apesar de “todos” e “alguns” fazerem parte do mesmo grupo categorial, ambos apresentam diferenças em relação aos seus traços de complemento.

### 1.3. Modelos computacionais do funcionamento da linguagem: DCG e LFG

Gómez Guinovart (2000, p.1) postula que “A implementação computacional de teorias lingüísticas e a elaboração de modelos computacionais da linguagem constituem interesses centrais da investigação em lingüística computacional”. Para este autor, um dos principais objetivos da Lingüística Computacional é o desenvolvimento de teorias lingüísticas formais que sejam implementadas computacionalmente. Essas teorias constituem modelos computacionais do funcionamento da linguagem, razão pela qual se denominam modelos lingüísticos. O autor cita como principais modelos lingüísticos computacionais a

LFG (gramática léxico-funcional), a GPSG (gramática de constituintes imediatos generalizados) e a Gramática Categorial, modelos os quais estão dentro de um só arcabouço, o das gramáticas de unificação, por se utilizarem de recursos matemáticos e técnicas de representação do conhecimento no âmbito da Inteligência Artificial em suas descrições lingüísticas.

Os objetos matemáticos de que fala Gómez Guinovart (2000) são referentes, principalmente, àqueles denominados estruturas de traços, e, por meio desses traços, os fenômenos lingüísticos descrevem-se formando equações com essas estruturas. Nesta seção, apresentaremos algumas noções sobre modelos formais que podem ser implementados computacionalmente. De acordo com Vieira e Lima (2001), uma gramática adotada pode ser escrita através de diversos formalismos, como por exemplo, a PSG (Gramática de Estrutura Sintagmática), a GPSG (Gramáticas de Constituintes Imediatos Generalizados), a HPSG (Gramática Sintagmática Dirigida por Núcleo), as redes de transição e as gramáticas de unificação funcional, que já citamos acima, de acordo com Gómez Guinovart (2000).

Em nossa pesquisa, iremos nos deter sobre o estudo relacionado ao formalismo lingüístico denominado por DCG (Gramática de Cláusula Definida) que, na verdade, não constitui uma teoria lingüística como a LFG. De acordo com Vieira e Lima (2001, p.18) “ela é um formalismo para a representação de gramáticas livres de contexto”. Segundo Pagani (2004, p.11) a DCG é um recurso que foi introduzido na década de 80, por Pereira e Warren (1980).<sup>10</sup> Uma gramática expressa em PSG pode ser transformada facilmente em Prolog através de uma DCG. É através da notação em DCG que será possível acrescentar ao léxico uma nova lista, através da qual se possa atribuir traços morfossintáticos que representem as propriedades de cada item lexical. Melhor dizendo, através da DCG há como garantirmos a unificação entre os traços morfossintáticos dos itens lexicais porque as propriedades gramaticais destes não foram contempladas num tipo de modelo como a PSG.

Eis, então, um dos porquês de a DCG tornar-se uma gramática tão mais poderosa que a gramática de constituintes, por permitir tratar da questão da concordância, através de traços gramaticais.

Como bem colocam Othero e Menuzzi (2005) “a DCG pode ser entendida, então, como um recurso que torna mais fácil ao lingüista a implementação de regras sintagmáticas em Prolog”. É graças ao recurso DCG que podemos fazer programas por meio do Prolog de

---

<sup>10</sup> PEREIRA, F.C.N and WARREN, D.H.D. Definite Clause Grammars for Language Analysis – A Survey of the Formalism and a Comparison with Augmented Transition Networks. **Artificial Intelligence**, 13, pp. 231-278, 1980.

modo muito parecido a uma PSG, só que, melhor ainda, pois conseguimos garantir a unificação entre os itens lexicais de uma determinada sentença, ou, como no caso do nosso estudo aqui, das expressões nominais.

Outro formalismo fortíssimo do qual iremos falar aqui é o da LFG, o qual foi lançado em 1982, por equipe multidisciplinar das áreas de Linguística, Computação e Matemática, ligada ao MIT, à Universidade Stanford e ao Centro de Pesquisa de Palo Alto da Xerox (Xerox Palo Alto Research Center – PARC) (ALENCAR, 2004, p. 174). Segundo Alencar (2004, p. 174) “a Gramática Léxico-Funcional é um dos modelos gramaticais mais difundidos não só no âmbito da Linguística Computacional, mas também, da Tipologia Linguística”.

De acordo com Alencar (2006a) a LFG é um formalismo matemático-computacional que possui o papel de descrever línguas naturais. Através deste formalismo poderão ser desenvolvidos tradutores, corretores e sistemas de diálogo homem-máquina de uma maneira bastante eficiente. Contudo a LFG não é somente um formalismo, mas também, uma vertente linguística gerativa diferindo, porém, bastante dos modelos gerativos da TRL e do PM, principalmente porque propõe uma arquitetura da Faculdade da Linguagem (gramática universal) onde não há transformações sintáticas e as informações gramaticais se distribuem em vários níveis paralelos de representação, relacionados por um sistema de correspondências, sendo os níveis mais importantes: estrutura constitucional e estrutura funcional. Na LFG toda ênfase é dada ao léxico, teoria esta conhecida por Lexicalista. Como constata Carnie (2002) é preconizada uma relação biunívoca entre palavras e os nódulos terminais das estruturas sintáticas.

Em nossa pesquisa, na verdade, utilizaremos para a construção e representação de nossa minigramática do português do Brasil o formalismo da DCG, e não da LFG, todavia, faremos uso de princípios desta teoria, como por exemplo, o da unificação de traços, em nossas análises.

Além disso, não utilizaremos transformações sintáticas, como propõem, por exemplo, Cerqueira (1993) e Magalhães (2004) na análise do DP em português. Contudo, salientamos que através de um formalismo como a LFG podemos de forma bem mais econômica e elegante representar, por exemplo, as propriedades morfossintáticas de uma expressão nominal.

Alencar (2006b) mostra que por meio da LFG, representamos os traços de uma determinada expressão por meio de AVMS (Matrizes de atributo e Valores), através da qual a ordem dos pares de atributos e valores torna-se irrelevante. Já através de uma DCG

precisamos estar atentos para a ordem na qual os traços são listados nas entradas lexicais, pois, qualquer falha, poderá fazer com que as regras sintáticas não funcionem com eficácia.

Em LFG, independentemente da ordem, conforme dissemos acima, podemos representar uma expressão nominal como a em (27), a seguir, da seguinte forma:

(27) Nós dentistas brasileiras.<sup>11</sup>

GEND	FEM
NUM	PL
PERS	1

**Nós**

NUM	PL
PERS	1

**Dentistas**

NUM	PL
-----	----

**Brasileiras**

NUM	PL
GEN	FEM

Como é de se observar, de uma forma bastante econômica, podemos representar os traços dos três itens lexicais que formam a expressão nominal, apresentando apenas os traços que realmente são marcados para cada um. Observamos que no item *dentista* como não há marca de gênero, este não foi representado por meio de nenhum atributo, e, o programa, por meio da LFG, leria isso sem nenhum problema.

É possível, fazermos em DCG, conforme (28), uma representação que se assemelhe em alguns aspectos a uma representação por meio da AVM em (27), contudo, de maneira um tanto quanto diferente no que se refere à elegância e economia de como as propriedades gramaticais são representadas através do modelo da LFG:

---

<sup>11</sup> Exemplo extraído de Alencar (2006b).

## (28) Representação de traços na DCG

*/\*a\*/* n(dentistas,[\_,pl]).

*/\*b\*/* adj(brasileiras,[fem,pl]).

*/\*c\*/* pron(nós,[\_,pl]).

Através de uma representação por meio de uma DCG é obrigatório que todo e qualquer nome seja marcado para gênero e número. Já que o substantivo *dentistas*, por sua vez, é classificado como um substantivo comum de dois gêneros, é preciso representar o gênero por meio de uma variável anônima ( \_ ), uma espécie de curinga, em Prolog. Na LFG, como constatamos, é preferível se dizer que o substantivo em evidência não possui atributo para gênero.

Nessa pesquisa, nos propomos a trabalhar com o formalismo DCG, assim como Araribóia (1989), Pagani (2004), Othero e Menuzzi (2005) Othero (2004,2006), Pardo (2005) e Alencar (2006). Vejamos, a esse respeito, o que afirmam Blackburn *et al.* (2001, p. 114):

Extra arguments can be used to cope with some complex syntactic problems. DCGs are no longer the state-of-art grammar development tools they once were, but they're not toys either. Once you know about writing DCGs with extra arguments, you can write some fairly sophisticated grammars.

### **1.4. Uma análise das expressões nominais do português segundo uma abordagem computacional.**

Nesta seção, apresentaremos, de forma sucinta, como podemos implementar através de uma linguagem computacional, por meio do formalismo DCG em linguagem Prolog, algumas expressões nominais em língua portuguesa.

#### **1.4.1. Construção de fragmentos computacionais em língua portuguesa no formalismo DCG em Prolog**

A seguir, apresentaremos dois modelos de gramática conhecidos por PSG e DCG e adotaremos a DCG como modelo que satisfaz a questão de unificação de traços de expressões nominais em língua portuguesa ao gerar e reconhecer tais estruturas.

A PSG (gramática de estrutura sintagmática) é um modelo gramatical capaz de gerar e reconhecer as expressões bem formadas de uma língua. Essa gramática é constituída por dois tipos de regras: as regras sintagmáticas e as regras lexicais. Ela é, na verdade uma família de

muitos formalismos, dentre eles a CFG (*context free grammar*), a qual é conhecida como gramática livre de contexto, com a qual trabalharemos a partir de então nesta pesquisa.

Como já é de nosso conhecimento, as regras sintagmáticas são aquelas que são capazes de gerar um conjunto infinito de frases de uma língua através de uma representação estrutural. Já as regras lexicais são as responsáveis pela introdução de elementos terminais (categorias lexicais).

Da mesma maneira que uma CFG pode gerar e reconhecer as expressões bem formadas de uma língua, poderemos aplicar as mesmas regras utilizadas por ela a “um interpretador ou compilador da linguagem de programação Prolog, acrônimo da expressão francesa *programmation en logique*, i.e., programação em lógica” (ALENCAR, 2006a, p.4), por meio de uma DCG, conforme já vimos na seção 1.3.

Através da DCG podemos transformar uma minigramática como a de (29) abaixo, expressa em CFG, em um reconhecedor gramatical como em (30):

(29) Fragmento de gramática de estrutura sintagmática do português CFG:

<p>NP → (D) N (PP) PP → P NP N → jogador, jogadora, jogadores, jogadoras, xadrez... P → com, de, ... D → um, uma, uns, umas, ...</p>
--

(30) Minigramática computacional do português no formato DCG sem traços

<p>np --&gt; n. np --&gt; d,n. np --&gt; d,n,pp. np --&gt; n,pp. pp --&gt; p,np. d --&gt; [um]. d --&gt; [uma]. d --&gt; [uns]. d --&gt; [umas]. p --&gt; [com].</p>
--



p --> [de].
n --> [jogador].
n --> [jogadora].
n --> [jogadores].
n --> [jogadoras].
n --> [xadrez].

Nesse novo modelo, não teremos apenas um gerador de estruturas, mas também um reconhecedor de expressões nominais como pertencentes ao português do Brasil.

Conforme Pagani (2004), para a geração de estruturas a partir do Prolog, basta que carreguemos o programa com uma diretiva do tipo “np (X, [])”. Nesse momento, aparecerão alguns dos sintagmas possíveis que o programa consegue gerar, como por exemplo: [um, jogador, de, xadrez], [um, jogador, de jogador], [um, xadrez, de xadrez] etc.

Para o reconhecimento desse sintagma “Um jogador de xadrez” aplicaremos outra diretiva “np ([um,jogador, de xadrez],[])”, para a qual se obtém uma resposta afirmativa, significando que o interpretador reconheceu “Um jogador de xadrez” como sendo uma sentença aceita por aquela gramática.

Para gerarmos um fragmento de gramática de estrutura sintagmática do PB se faz necessário que se determinem regras capazes de gerar e reconhecer expressões nominais pertencentes a essa língua como bem formadas ou gramaticais. Contudo as regras sintagmáticas não conseguem assegurar que os traços gramaticais que diferentes palavras trazem em sua estrutura possam ser compatíveis com aqueles de outras palavras da mesma expressão. Assim, podemos dizer que esse modelo de gramática apresenta um problema de hipergeração de estruturas agramaticais.

Como já constatamos, uma minigramática elaborada a partir de uma CFG é capaz de gerar todas as possíveis combinações gramaticais, sendo, também, capaz de reconhecer se as construções formadas são gramaticais. Porém isso não é tudo, o problema proveniente dessa minigramática é o de adotar uma visão simplista do léxico<sup>12</sup>, pois como sabemos, o léxico não é uma mera lista de palavras, ele possui, também, informações categoriais destas formas, entretanto, o que ela não traz em suas entradas são informações referentes às propriedades gramaticais que cada item possui, como gênero, número, pessoa, definitude, posição etc.

---

<sup>12</sup> Ver Alencar 2006 b

Seria obrigatório que uma determinada gramática que se propõe a apresentar um modelo do que o falante conhece sobre a sua língua materna, incluísse no léxico, por exemplo, a informação de que *umas* é um determinante de gênero feminino e número plural; e que *jogador* é um nome cujas propriedades são singular e masculino, para que, com isso, não fossem geradas expressões do tipo *umas jogador, um jogadora* etc.

Uma solução proposta pela DCG, que a diferencia relevantemente da CFG, é a de acrescentar aos itens lexicais uma segunda lista que possa atribuir traços morfossintáticos, que representem as propriedades gramaticais de cada item lexical, não deixando de alterar, também, as regras sintagmáticas para que entrem em sintonia, digamos assim, com as regras lexicais.

É por isso, que podemos notar que em um modelo do tipo CFG existem regras lexicais e sintagmáticas, porém não há como garantir a unificação entre os traços morfossintáticos dos itens lexicais porque as propriedades gramaticais destes não foram contempladas nesse tipo de modelo.

Vejamos em (31) um exemplo de como aplicamos as mesmas regras utilizadas em CFG por meio de uma DCG em Prolog, considerando os traços morfossintáticos dos itens lexicais:

(31) Minigramática no formato DCG com traços

```
%REGRAS SINTAGMÁTICAS
np([GEN,NUM]) --> n([GEN,NUM]).
% um jogador de xadrez
np([GEN,NUM]) --> d([GEN,NUM]), n([GEN,NUM]), pp.
pp --> p, np[GEN,NUM].

%LÉXICO
n([masc,sing]) --> [jogador].
n([fem,sing]) --> [jogadora].
n([masc,sing]) --> [xadrez].
d([masc,sing]) --> [um].
d([fem,sing]) --> [uma].
p --> [de].
```

Para transformar um reconhecedor em analisador gramatical, basta acrescentar um argumento em todos os predicados usados, de forma a construir nesse argumento uma representação da estrutura gramatical.

Observemos, a seguir, um modelo de analisador gramatical, em (32):

(32) Analisador Gramatical

```
% analisador
np(np(X,Y),[G,N]) --> d(X,[G,N]), n(Y,[G,N]).
n(n(jogador),[masc,sing]) --> [jogador].
n(n(jogadores),[masc,pl]) --> [jogadores].
n(n(jogadora),[fem,sing]) --> [jogadora].
n(n(jogadoras),[fem,pl]) --> [jogadoras].
d(d(o),[masc,sing]) --> [o].
d(d(os),[masc,pl]) --> [os].
d(d(a),[fem,sing]) --> [a].
d(d(as),[fem,pl]) --> [as].
```

O analisador apresentado, não só gera e reconhece estruturas da língua portuguesa, como também, representa (em forma de parentização etiquetada e, também, de árvores) tais estruturas.

# CAPÍTULO 2

## HIPÓTESES E OPÇÕES METODOLÓGICAS

---

### 2.1. Hipóteses

Nessa pesquisa, partimos do pressuposto de que em lingüística gerativa os fenômenos gramaticais de uma língua são explicados a partir de uma linguagem matemática ou lógica e de técnicas de representação do conhecimento no âmbito da inteligência artificial em suas descrições lingüísticas. Atestamos, ainda, que uma expressão nominal é um DP, cujo núcleo pode ser representado por um pronome pessoal, assim como artigos definidos e indefinidos. Assim, baseamo-nos nesses pressupostos para fazermos os seguintes questionamentos:

- a) O Determinante (D) na estrutura sintagmática do português do Brasil sempre precede o NP?
- b) Os pronomes demonstrativos e possessivos se enquadram na categoria dos determinantes no português do Brasil?
- c) De que maneira se deve analisar uma estrutura nominal em que apareçam dois determinantes em seqüência?
- d) Os pronomes pessoais da língua portuguesa podem ser considerados núcleos de um DP que tomam NPs enquanto complemento?
- e) Quais são os traços morfossintáticos do pronome pessoal enquanto determinante de uma estrutura nominal do português do Brasil?

E propomos as seguintes hipóteses:

- a) No português do Brasil, o determinante poderá vir tanto pré-nominalmente como pós-nominalmente em uma determinada expressão nominal.
- b) Os pronomes demonstrativos são determinantes, pois não podem aparecer em uma posição seguidos ou precedidos por um artigo. Já os pronomes possessivos são compatíveis tanto com um determinante vazio, representado por /e/, quanto com um determinante lexical (um artigo definido, por exemplo). Conseqüentemente, o possessivo não toma a posição de núcleo, mas, sim, faz parte do NP, como é o caso dos exemplos: “/e/ meu filho” e “o meu filho”.
- c) Em uma expressão nominal em que nos deparamos com dois ou mais determinantes, como é o caso da expressão *todos os alunos*, já que no português do Brasil um DP não

licencia especificador, o D “todos” é o núcleo de um DP que tem como complemento outro DP, i.e. *os alunos*.

- d) Todos os pronomes pessoais, contrariamente ao que preconiza Radford para o inglês, podem, em língua portuguesa, assumir o papel de núcleo de um DP, admitindo um NP como complemento, já que, em língua portuguesa, o pronome pessoal sozinho já poderá ser um núcleo, como por exemplo em *Ele saiu*.
- e) Os traços morfossintáticos que definem os pronomes pessoais são: definitude, gênero, número, pessoa, posição. Por meio desses traços, é que os pronomes pessoais, enquanto núcleos de um DP, licenciam complementos, que, com eles, possam se unificar, ou seja, ambos se combinam para formar uma estrutura gramatical.

Consideramos como questões fundamentais para nosso trabalho as referentes aos itens (c) e (d) e foi através delas que desenvolvemos grande parte da discussão para esta pesquisa.

## 2.2. Opções Metodológicas

Para nossa pesquisa, utilizamos uma amostra de construções do português do Brasil – PB), extraídas do *corpus* eletrônico NILC (Núcleo Interinstitucional de Linguística Computacional), disponível no Projeto Linguateca do Ministério da Ciência e da Tecnologia de Portugal.<sup>13</sup>

O NILC/São Carlos, como foi batizado pela Linguateca, é o maior *corpus* eletrônico do português do Brasil, o qual vem sendo acessado pela Linguateca desde 1999, através do projeto AC/DC, que significa “acesso à corpora/disponibilização de corpora”. O AC/DC é um projeto o qual tem por objetivo maior disponibilizar em rede uma quantidade significativa de *corpora* em língua portuguesa (PINHEIRO e ALUÍSIO, 2003).

O *corpus* NILC-São Carlos foi primeiramente batizado em 1998, apenas por NILC, e foi um dos primeiros *corpora* equilibrado pois continha vários gêneros de texto para a língua portuguesa. Em 1999, esse *corpus* foi batizado, como já dissemos, por NILC-São Carlos e, em seguida, chamado por NILC Anotado, porém, atualmente é chamado apenas por NILC<sup>14</sup>.

A partir de 1999, ano em que foi acessado pela primeira vez, por meio da Linguateca, vários outros *corpora* dos quais já temos conhecimento, foram sendo anexados a ele. O

---

<sup>13</sup> Este corpus está disponível no site <www.linguateca.pt>

<sup>14</sup> Para maiores detalhes a respeito do corpus NILC, consultar relatório de Pinheiro & Aluísio (2003), o qual apresenta a história, o conteúdo, a vastidão e alguns problemas deste corpus, cuja fonte encontra-se em nossas referências bibliográficas.

material mais significativo, por sua vez, aproveitado pelo *corpus* NILC é o da “Folha de São Paulo”.

Para a coleta de dados, utilizamos o programa computacional “O Constructor”,<sup>15</sup> que é uma ferramenta interativa e “amigável”, para a construção dos comandos de busca nos *corpora* do PB, disponíveis no âmbito da Linguateca .

Este programa auxilia, sobretudo, usuários leigos a pesquisarem em *corpora* que foram codificados em ambiente IMS-CWB, que exige da parte de quem vai usá-lo o conhecimento de uma linguagem chamada de sintaxe de interrogação<sup>16</sup> desse sistema.

Uma das vantagens oferecidas por este programa é que mesmo usuários mais familiarizados com a linguagem computacional poderão, também, utilizar-se dele, para minimizar, assim, o seu trabalho de “construção de determinadas expressões regulares” (ALENCAR, 2000).

Esse programa é um tradutor que converte a pesquisa feita pelo usuário leigo, ou não, a partir de um formulário HTML (que já está no programa), “em expressões na linguagem de interrogação do corpus *Query Processor* (CPQ) do ambiente de processamento de corpora do IMS”. (ALENCAR, 2000, p. 5)

Para a análise de dados de nossa pesquisa, primeiramente, fizemos um levantamento intuitivo de expressões nominais possivelmente geradas por falantes da língua portuguesa, através das quais, a seguir, levantamos dados de mesma estrutura interna no corpus NILC.

A lista de expressões nominais elencadas, por nós, foi a seguinte:

1. Todos lingüistas
2. Alguns alunos
3. Esta teoria
4. Nós gerativistas
5. Vocês funcionalistas
6. Ambos os lingüistas
7. Todos os lingüistas
8. Todos nós lingüistas
9. Todos vocês funcionalistas
10. O meu filho
11. Este meu livro

---

<sup>15</sup> Para maiores detalhes sobre esta ferramenta ver Alencar, 2000.

<sup>16</sup> Do inglês *query language*

12. Eu meu Deus
13. Quase todos nós alunos
14. Quase todos vocês médicos
15. Todos nós lingüistas brasileiros
16. Quase todos nós lingüistas brasileiros
17. Lingüista aquele que
18. Teoria esta que
19. Alguns dos alunos
20. Nenhuma das filhas
21. Nenhuma das minhas filhas
22. Alguns dos meus amigos
23. Oito de nós
24. Cada um dos livros
25. Um filho meu
26. Alguns amigos meus
27. Filho meu
28. Homem nenhum
29. Mulher alguma
30. Aquele imbecil do porteiro
31. Este idiota deste velho
32. Menina rica
33. As meninas mais belas
34. As meninas as mais belas
35. A outra irmã minha
36. A safada da minha irmã

O nosso próximo passo foi dirigir um comando, através do programa “O Constructor”, ao *corpus* NILC, e aguardar que fossem levantadas todas as ocorrências que existiam naquele *corpus* que fossem utilizadas por falantes do PB. Um exemplo do que fizemos pode ser representado pela expressão nominal *nós gerativistas* que encontramos na lista acima.

Acessamos o programa “O Constructor” e solicitamos, através dos comandos que nos foram apresentados por esse programa, que aparecessem expressões nominais que possuíssem

o pronome pessoal “*nós*”, seguido por nome, de acordo com o modelo intuído. Em seguida, o programa nos apresentou como resposta o seguinte<sup>17</sup>:

## RESULTADOS DA PROCURA

Tue Jan 17 17:15:25 CET 2006

Procura: "**nós**" [ ] {0,0} [pos="N.\*"&pos!="NUM.\*"] within 1 s; Pedido de uma concordância em contexto Corpus: NILC/São Carlos anotado v. 4.5 77 ocorrências.

### 1.1.1. Concordância

Procura: "**nós**" [ ] {0,0} [pos="N.\*"&pos!="NUM.\*"] within 1 s;

par 22950: Todos **nós atores** brasileiros estamos loucos para trabalhar num filme lá fora.par

42465: Além disto, os antípodas não conseguem entender por que **nós terráqueos** insistimos nestes termos mentalistas: para eles, a única coisa que produz experiência é a estimulação de partes do cérebro .

par Agrofolha-94b-agr-1: Cabe agora aos demais poderes e principalmente a **nós agricultores** darmos continuidade ao processo .

par Agrofolha-94b-agr-1: Será que não cabe a **nós brasileiros**, que produzimos café sem poluir as águas, proclamar aos consumidores as virtudes de nosso café natural e ecológico?

par Brasil-94b-pol-2: Se o governador não tivesse cedido, todos **nós reféns** teríamos sido mortos e eles mesmos teriam se suicidado .

par Cotidiano-94b-soc-2: «Se as Comunidades Eclesiais de Base são assim tão perigosas, o secretário deveria prender primeiro a **nós bispos**, a começar pelo do Ceará, porque desde 1962 começamos no Brasil a promover as CEBS e incentivá-las .

par Cotidiano-94a-soc-1: Por essa esperança, é preciso parabenizar, nesse 15 de outubro, a todos **nós professores** que temos honra de «professar» o que sabemos .

par Dinheiro-94a-eco-2: Lembro-me de um jornalista ter escrito, na época, que **nós brasileiros** continuávamos a ser índios e trocávamos riquezas por bugingangas .

par Dinheiro-94b-eco-1: Estamos no início de um ano novo, ocasião em que **nós brasileiros** nos quedamos a meditar sobre os rumos de nosso futuro político e socioeconômico .

A partir desses dados, fizemos, no capítulo 3, uma descrição lingüística seguindo pressupostos da teoria X-barras “contemporânea” conforme apresentada por nós na seção

---

<sup>17</sup> Estamos transcrevendo exatamente o que nos foi fornecido pelo AC/DC.



1.1.3. do capítulo 1 dessa pesquisa e, a seguir, no capítulo 4 elaboramos um fragmento de gramática computacional do PB capaz de gerar, reconhecer e analisar possíveis expressões licenciadas por esta língua, conforme modelos computacionais descritos na seção 1.3. do capítulo 1. Este último passo foi o objetivo principal de nossa pesquisa.

### **2.3. Descrição da análise**

Descreveremos, nesta seção, um percurso de como foram feitas as análises que se encontram nos capítulos 3 e 4 desta dissertação.

No capítulo 3, apresentamos a análise de 36 expressões nominais da língua portuguesa, as quais foram intuídas por nós, conforme a seção 2.2., enquanto falantes desta língua, seguindo pressupostos da teoria X-barra “contemporânea”.

Primeiramente apresentamos as expressões nominais representadas em forma de estrutura arbórea, conforme encontramos na maior parte da literatura que traz como perspectiva a teoria gerativa. Após esta representação, mostramos um fragmento de gramática livre de contexto (CFG) elaborado por nós, o qual traz em sua estrutura regras sintagmáticas e um léxico o qual comporta itens lexicais encontrados nas 36 expressões nominais analisadas nesta pesquisa.

No capítulo 4, a partir dessa CFG aplicamos as mesmas regras a um compilador da linguagem de programação Prolog, sobre o qual já falamos na seção 1.4.1. Utilizamos o formalismo DCG para transformar a nossa minigramática em CFG em um reconhecedor gramatical implementado computacionalmente.

A partir desse novo formalismo sintático computacional, vimos que ele não só gera, como também reconhece estruturas do PB como gramaticais ou não, através de comandos que são dados por nós ao programa. Através de um modelo como a CFG temos a possibilidade de gerar e reconhecer, também, estruturas de uma determinada língua como gramaticais. Porém, por meio desse formalismo, o léxico é visto, apenas, como uma mera lista de palavras, deixando de se apresentar informações categoriais de cada uma das suas formas, como, por exemplo, gênero, número, pessoa, definitude, entre outras propriedades que se tornarão relevantes no momento em que estas peculiaridades de cada item forem ser unificadas dentro do DP. Assim, foi através de uma gramática implementada em DCG, que pudemos acrescentar a essas entradas lexicais uma nova lista capaz de atribuir traços morfossintáticos aos itens lexicais os quais vinham representar as particularidades gramaticais de cada item, também, não deixando de alterar as regras sintagmáticas para que houvesse uma sintonia entre estas e o léxico.

O nosso terceiro passo foi acrescentar à nossa gramática em DCG traços morfossintáticos aos itens lexicais, para que, assim, o programa, posteriormente, fornecesse expressões nominais cujos traços morfossintáticos pudessem ser unificados, gerando e reconhecendo, assim, tais estruturas como gramaticais.

No entanto, nesta pesquisa, propomo-nos não só a elaborar um fragmento de gramática computacional que gerasse e reconhecesse um subconjunto de expressões nominais da língua portuguesa. O nosso foco principal foi, sobretudo, analisar essas expressões. Analisar significa representar tais estruturas ou em forma de parentização etiquetada ou por meio de árvores.

Para isso, o nosso quarto passo foi transformar o reconhecedor em DCG em um analisador gramatical. A partir deste analisador, não conseguimos apenas gerar e reconhecer expressões nominais da língua portuguesa, mas também analisar tais estruturas.

# CAPÍTULO 3

## DESCRIÇÃO SINTÁTICA DA ESTRUTURA INTERNA DAS EXPRESSÕES NOMINAIS NO PORTUGUÊS DO BRASIL

---

Apresentaremos na seção 3.1. um estudo a respeito do estatuto dos pronomes pessoais dentro da estrutura interna do DP seguido do resultado da análise que fizemos através do *corpus* NILC com expressões nominais em língua portuguesa que trazem pronomes pessoais em sua estrutura.

Em seguida, na seção 3.2. apresentamos a descrição sintática da estrutura interna das 36 expressões nominais em língua portuguesa e as regras geradas a partir dessas estruturas que serviram para a construção de um fragmento de gramática livre de contexto (CFG), que apresentamos na seção 3.3.

Resolvemos separar a análise dos pronomes pessoais da análise das outras expressões nominais já que um dos objetivos principais para essa pesquisa é constatar se todos os pronomes em PB são núcleos de um DP que tomam NP como complemento. Assim, para dar uma maior ênfase ao assunto, julgamos que fosse melhor explicitar em uma seção separada.

### **3.1. Os pronomes pessoais em língua portuguesa**

Em nossa pesquisa, consideramos que a projeção de uma expressão nominal como, por exemplo, *o jogador de xadrez*, se dá a partir de um D, e não de um N conforme é postulado na teoria X-barras de Chomsky (1970). Para isso, nesta seção, apresentaremos o modelo de Radford (1997) para mostrarmos como este autor analisa as estruturas sintagmáticas nominais da língua inglesa no âmbito do Programa Minimalista. A seguir, baseados neste autor, a título de exemplificação, nos deteremos sobre expressões nominais da língua portuguesa que trazem pronomes pessoais em sua estrutura, pois postulamos, assim como Radford postula para a língua inglesa, que os pronomes pessoais são núcleos de um DP que selecionam NP como complemento.

Porém, antes de nos remetermos a Radford (1997), faremos uma exposição a respeito de como os pronomes pessoais são tratados segundo a perspectiva de alguns gramáticos tradicionais e até mesmo lingüistas mais recentes.

### 3.1.1. A crítica de Monteiro (1994) à análise tradicional dos pronomes pessoais

De acordo com Monteiro (1994) as designações e conceitos atribuídos aos pronomes não correspondem à realidade, pois nem todos os pronomes são meros substitutos, e, quando são, não necessariamente substituem nomes e, ainda mais, existem expressões que têm esse aspecto, mas não são pronomes.

Cunha e Cintra (1985) constituem um exemplo de autores que adotam essa análise tradicional dos pronomes. Com efeito, eles definem os pronomes como elementos que "desempenham na oração as funções equivalentes às exercidas pelos elementos nominais", pois servem para representar e acompanhar um substantivo (pronomes substantivos e pronomes adjetivos, respectivamente).

O contraste nos exemplos em (33) mostra que, sintaticamente, o pronome pessoal não pode ser considerado um substantivo ou substituto do nome, como lembra Monteiro (1994), citando Postal (1969)<sup>18</sup> que diz que os pronomes pessoais não trazem os mesmos traços semânticos de nomes e, também, mantêm um comportamento sintático distinto, por não se articularem com os artigos. Raposo (1992) também chama a atenção para essa diferença de comportamento entre substantivos e pronomes.

- (33) a. uma aluna apenas faltou  
b. \*uma ela apenas faltou

Por outro lado, um pronome pessoal pode ser seguido de substantivo, como salienta Barrenechea (1962, *apud* MONTEIRO, 1994)<sup>19</sup>, conforme exemplo (34). Essa observação é corroborada por exemplos do português do Brasil, extraídos do *corpus* NILC, conforme o exemplo(34). Note-se que, em (35), ao contrário do que ocorre em (34), o substantivo *atores* segue, sem vírgula, o pronome, não podendo, desse modo, ser classificado como um simples epíteto como em (34) e (36).

- (34) Ego, César. (Latim)  
'Eu, César'

- (35) Todos **nós** **atores** brasileiros estamos loucos para trabalhar num filme lá fora.  
(NILC)

---

<sup>18</sup> POSTAL, Paul. On so-called Pronouns in English. In: REIBEL, D.A. & SCHANE, S. A. (eds.). **Modern Studies in English. Readings in Transformacional Grammar**. New Jersey, Englewood Cliffs, Prentice-Hall Inc., 1969. p.177-206.

<sup>19</sup> BARRENECHEA, Ana Maria. El pronombre y su inclusión en un sistema de categorías semánticas. **Filología**. Buenos Aires, 8 (1/2), p. 241-72, 1962.

(36) Tu, meu amor, estás cada vez mais magra. (PE) (Mateus et al. 1989:188).

Epítetos são “certas expressões parentéticas isoladas por pausas no interior do SN e colocadas à direita do núcleo e integram-se nos complementos do nominal. Os epítetos podem ser SNs, SADJs e Fs’’. (MATEUS *et al*, 1989, p.188)

Desse modo, em nossa pesquisa, partimos da hipótese de que os pronomes pessoais podem ser núcleos de um DP, tomando um NP como seu complemento, conforme comprovamos através de dados extraídos do *corpus* NILC.

### 3.1.2. Os pronomes pessoais no Programa Minimalista: Um exemplo de análise da expressão nominal em língua inglesa

Radford (1997) classifica as palavras tradicionalmente chamadas por pronomes pessoais como determinantes, considerando exemplos em inglês como os seguintes:

(37) [We psychologists] don't trust [you linguists].  
'Nós psicólogos não confiamos em vocês lingüistas.'

(38) [The psychologists] don't trust [the linguists].  
'Os psicólogos não confiam nos lingüistas.'

(39) [We] don't trust [you]  
'Nós não confiamos em vocês.'

Para Radford, tanto em (37) quanto em (39), o *we* 'nós' funciona como determinante. Porque, assim como outros elementos considerados determinantes, como os artigos, por exemplo, conforme (38), o pronome pessoal *we* tem também propriedades de seleção de complemento. Os exemplos a seguir mostram que o pronome *you* 'você' ou 'vocês', a exemplo do determinante *several* 'vários', licencia apenas um complemento no plural.

(40) You politicians are liars.  
'Vocês políticos são mentirosos.'

(41) \*You politician are a liar. (Radford 1997:69)  
'você político é um mentiroso.'

(42) He bought several chairs.  
'Ele comprou várias cadeiras.'

(43) \*He bought several chair. (Radford 1997:40)  
'ele comprou vária cadeira.'

Além disso, é o pronome que, exatamente como o artigo, determina o traço de pessoa da expressão nominal, como vemos em (37) e (38). O traço de pessoa não é inerente aos substantivos. Ele é determinado pelo pronome em (37) e pelo artigo em (38). Isso mostra que o pronome funciona, do mesmo modo que o artigo, como núcleo do sintagma.

De um modo geral, os determinantes do inglês, como *all* 'todos' e *we* 'nós', por exemplo, podem ser usados tanto pré-nominalmente quanto pronominalmente:

- (44) a. All guests are welcome.  
'Todos os hóspedes são bem-vindos.'
- b. All are welcome  
'Todos são bem-vindos.'

O determinante *the* 'o', 'os' etc., porém, só pode ser usado pré-nominalmente, conforme exemplo (45). Já o pronome pessoal *you* 'você' não admite o uso pré-nominal, conforme exemplo (41).

- (45) \*[The] don't trust [the]  
'os não confiam nos'

Esses exemplos mostram o paralelismo de comportamento entre os pronomes pessoais e o que se considera, como determinantes. Todos esses elementos possuem traços que determinam se podem ocorrer pré-nominalmente, pós-nominalmente ou em ambas as posições.

Radford (1997), seguindo propostas de Postal (1966)<sup>20</sup>, Abney (1987) e Longobardi (1994)<sup>21</sup>, analisa uniformemente a sintaxe dos nominais. De acordo com sua análise, todo nominal constitui projeção de um constituinte nuclear D, isto é, é um DP. Trata-se, como vimos, da hipótese DP. Os pronomes pessoais, assim como os artigos, enquadram-se, para este autor, na categoria dos determinantes.

### 3.1.3. Um estudo dos pronomes pessoais em língua portuguesa enquanto núcleos de um DP : Uma breve análise

---

<sup>20</sup> POSTAL, Paul. On So-Called "Pronouns" in English..., In: Dinneen, F. (ed.) **The 19<sup>th</sup> Monograph on Languages and Linguistics**. Georgetown University Press, Washington D.C. Reimpresso em Reibel, D.A. e S.A. Schane (eds.), 1966.

<sup>21</sup> LONGOBARDI, G. Reference and Proper Names: A Theory of N-Movement in Syntax and Logical Form. **Linguistic Inquiry**, v. 25, n.4, 1994.

Nesta seção, apresentaremos os resultados do levantamento de dados a partir de uma amostra de construções do português do Brasil extraídas do *corpus* eletrônico NILC, e, faremos, em seguida, uma discussão lingüística acerca do assunto.

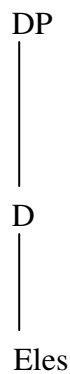
<b>Pronomes Pessoais em Língua Portuguesa</b>	<b>O pronome funciona como um D que toma NP como complemento</b>	<b>O pronome não toma NP como complemento; a palavra seguinte está categorizada erroneamente como N</b>	<b>O substantivo não forma DP com o pronome</b>
<b>EU (61 ocorrências)</b>	1	46	14
<b>TU (5 ocorrências)</b>		5	
<b>VOCÊ (32 ocorrências)</b>	6	20	6
<b>ELE (158 ocorrências)</b>		87	71
<b>ELA (84 ocorrências)</b>		42	42
<b>NÓS (132 ocorrências)</b>	115	6	11
<b>VÓS (1 ocorrência)</b>			01
<b>VOCÊS (12 ocorrências)</b>	12		
<b>ELES (181 ocorrências)</b>		69	112
<b>ELAS (85 ocorrências)</b>		41	44

Tabela 1. Resultado do levantamento dos pronomes pessoais no corpus NILC.

Conforme postulamos do mesmo modo que Radford (1997) os pronomes pessoais *nós* e *vocês* são, assim como os artigos, núcleos de um DP que selecionam NP enquanto complemento em língua portuguesa.

Os dados da tabela 1 mostram claramente que os pronomes pessoais *você*, *vocês*, *nós* e até mesmo o *eu*, apesar de poucas ocorrências encontradas, funcionam como núcleo de um DP, tomando um NP como complemento. Não foi confirmada, contudo, a hipótese de que o pronome *eles* tem esse mesmo comportamento sintático, corroborando assim a hipótese, também, de Radford (1997) em relação a expressões nominais em língua inglesa. De fato, não se constatou nenhuma ocorrência onde esse pronome é o núcleo de um DP com NP complemento. Para os pronomes de terceira pessoa é adotada, portanto, uma representação nos moldes da estrutura a seguir, que mostra a estrutura do DP *eles* numa frase como:

(46) [DP Eles] chegaram.



Durante nossa análise, o pronome pessoal que apresentou um maior número de ocorrências enquanto núcleo de um DP foi o pronome pessoal *nós*. De 132 ocorrências, tivemos 115, nas quais este pronome toma um NP como complemento. Vejamos uma delas a seguir:

(47) Estão querendo levar para o real a inflação que acabou com o valor do cruzeiro e do dinheiro que havia antes, mas **nós brasileiros** não vamos permitir que isto aconteça .

Não há nenhuma dúvida de que temos nessa ocorrência um pronome pessoal tomando um NP como complemento. Observemos como o traço de pessoa deste pronome reflete na concordância até mesmo com o verbo. Desse modo, assim como no inglês, como postulou Radford (1997), em língua portuguesa, o *nós*, também é um núcleo de um DP.

Quanto ao pronome *vocês*, encontramos apenas 12 ocorrências, porém, todas estas 12 ocorrências trazem um pronome pessoal *vocês* enquanto núcleo de um DP:

(48) A percepção, como **vocês psicólogos** a estudam, não pode afinal de contas ser diferente da observação na física, pode?

Mais uma vez estamos de acordo com Radford (1997) de que o pronome pessoal *vocês*, em língua portuguesa, assume a posição de D em uma expressão nominal, selecionando um NP como complemento.

Na seção 3.1.2., de acordo com o exemplo (41) de Radford (1997), o qual representamos a seguir em (49), o *you*, enquanto singular, em língua inglesa, não admite o uso pré-nominal, como o artigo e licencia complemento apenas no plural:

(49) \*You politician are a liar.

Contudo não foi isso que encontramos nos nossos dados do *corpus* NILC. Em relação ao pronome *você*, em língua portuguesa, foram encontradas 32 ocorrências, das quais, seis destas traziam o pronome pessoal *você* pré-nominalmente, conforme podemos ver a seguir:



(50) Há muita coisa que o Governo pode fazer e que **você cidadão** também pode.

No exemplo (50) está bastante claro que o traço de pessoa não é inerente ao substantivo cidadão, mas sim ao pronome *você*, o qual licencia um complemento no singular. Isso nos mostra que este pronome em língua portuguesa funciona, do mesmo modo que o artigo, como núcleo do sintagma determinante.

Em relação ao pronome pessoal *eu*, encontramos 61 ocorrências, das quais, apenas uma apresenta este pronome realizando o papel de núcleo de uma expressão nominal. Observemos a ocorrência encontrada:

(51) A bondade especial de minha mãe tinha sido a de amor constando com a justiça, que **eu menino** precisava.

Não tivemos dúvidas ao classificarmos esta ocorrência *eu menino* como um DP, que tem como núcleo o determinante *eu*. Por meio desse exemplo, fica claro para nós o paralelismo entre o pronome pessoal *eu* e o que consideramos como determinantes. Se substituíssemos o pronome pessoal *eu* pelo artigo *o*, por exemplo, teríamos:

(52) A bondade especial de minha mãe tinha sido a de amor constando com a justiça, que **o menino** precisava.

Ou seja, o *eu* está no mesmo plano que o artigo definido *o*. É por meio desse pronome que em (51), as propriedades morfossintáticas refletem-se no complemento, o NP *menino*, e no verbo *precisar* que passa para a primeira pessoa do singular, assim como *eu*.

No mais, em relação aos pronomes pessoais *tu* e *vós*, o que constatamos, diante de nosso levantamento de dados, foi que estes não têm a possibilidade de aparecerem em uma expressão nominal pré-nominalmente, mas, apenas, pronominalmente, ou seja, assim como o *ele*, *ela*, *eles* e *elas*, enquanto um DP, que não selecionam NP como complemento.

### 3.2. As expressões nominais analisadas segundo a teoria X-barras contemporânea

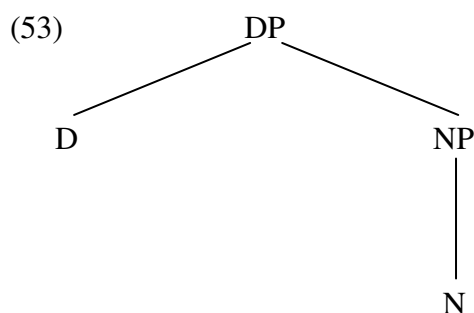
Nesta seção, apresentaremos a descrição linguística das 36 expressões nominais sugeridas por nós e apresentaremos as regras através das quais são geradas e suas respectivas representações arbóreas, para que, posteriormente, no capítulo 4, possamos construir um fragmento de gramática composto por expressões nominais correntes no PB.

## ANÁLISE 1

Observemos os exemplos de expressões nominais a seguir:

- 1- Todos lingüistas
- 2- Alguns alunos
- 3- Esta teoria
- 4- Nós gerativistas
- 5- Vocês funcionalistas

Para estas expressões lingüísticas teremos a seguinte representação arbórea:



A partir da representação arbórea em (53), podemos, assim, produzir as expressões nominais de 1 a 7. Assim, para construirmos nossa minigramática das expressões nominais do PB, incluímos em nosso *parser* as seguintes regras:

(54)  $DP \rightarrow D \ NP$

(55)  $NP \rightarrow N$

A regra (54) diz que um DP consiste de um D e de um NP, e a regra (55) que um NP consiste de um N. Veremos por meio de análises posteriores, que nem sempre um DP, por exemplo, irá constituir-se de D e de um NP, e que nem sempre um NP consistirá apenas de um N, pois isto dependerá da estrutura interna de cada expressão apresentada por nós.

Outra informação que não podemos esquecer é a de que deveremos subcategorizar as palavras no léxico, porém, deixaremos para implementar um léxico geral somente quando construirmos todo o nosso analisador.

Nessa análise já consideramos os pronomes pessoais *nós* e *vocês* como categorias D, assim como os quantificadores *todos* e *alguns*, que segundo Radford (1997), constituem uma subcategoria dos determinantes em inglês. Assim, equiparamos esses pronomes pessoais aos artigos, determinantes prototípicos em língua portuguesa que possuem propriedades de seleção de complemento.

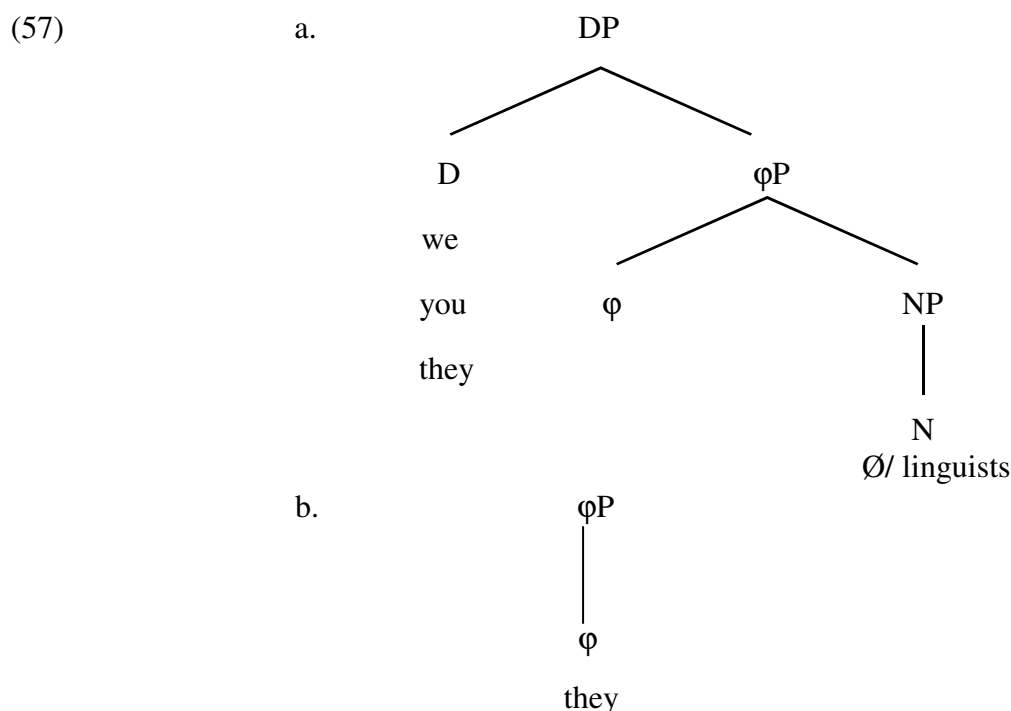
Constatamos ao levantar os dados do *corpus* NILC, que o pronome pessoal *eles* diferentemente do *nós* e *vocês* não é um núcleo que possa tomar um NP como complemento. Assim, concluímos que no português do Brasil acontece algo como no inglês, como apresentamos na seção 3.1.2., em que Radford (1997) constata que, em língua inglesa, tanto o *we* como o *you* funcionam como determinantes em expressões nominais dessa língua, constituindo uma categoria funcional. Porém este autor mostra que “os pronomes pessoais de 3ª pessoa restringem-se, no inglês padrão, ao uso pronominal, daí a agramaticalidade de expressões como *\*they boy* (aceitável em variantes não-padrão do inglês como a de Bristol, Inglaterra)” (Alencar, 2006 p.3).

Em Déchaine e Wiltschko (2003, p.80) são apresentados os seguintes exemplos para a língua inglesa referentes ao *nós*, *vocês* e *eles*:

- (56) a. *we linguists*      *us linguists*  
 b. *you linguists*      *you linguists*  
 c.\* *they linguists*      *\*them linguists*

O que as autoras apontam é que enquanto 1ª e 2ª pessoas, no que se refere a sujeito e objeto, há a possibilidade desses pronomes assumirem o lugar de um D que tomam NP enquanto complemento, contudo em 3ª pessoa (pessoa de quem se fala), não há esta possibilidade, nem para sujeito, nem para objeto.

Vejamos, a seguir, a representação arbórea apresentada por Déchaine e Wiltschko (2003, p.81):



Por meio da análise da estrutura 57a., vemos que o *they* é um pró-φP, uma categoria funcional, que corresponde à categoria AGR (concordância) em outras abordagens .

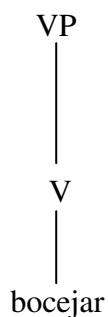
Para nossas análises, referente aos pronomes pessoais, seguiremos Radford (1997), que postula que tanto *them*, como *they* são um D, o qual diferem de *vocês* por não admitirem complemento:

(58)



Observemos que esta representação é semelhante à de um verbo intransitivo, por exemplo:

(59)

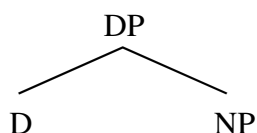


O fenômeno da subcategorização é geral, não envolvendo apenas D (determinante) e V (verbo), mas também categorias do tipo A (adjetivo), N (nome), P (preposição), C (complementador) etc.

Comparemos essas outras duas construções a seguir:

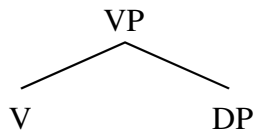
(60) a. Nós gerativistas

b.



(61) a. Comprei um livro

b.



O que constatamos é que 60 a. e 60b são duas representações idênticas.

Após nossa análise, podemos constatar que Déchaine e Wiltschko (2003) divergem de Radford (1997) tão somente por conta de uma mera questão notacional ao analisarem o *they* como *phi-agreement*, e Radford, analisa o *they* como D.

Em nossa pesquisa, através de dados do *corpus* NILC, pudemos comprovar que, de fato, os pronomes pessoais *nós* e *vocês* subcategorizam um NP como complemento, assim como postulam esses autores para a língua inglesa.

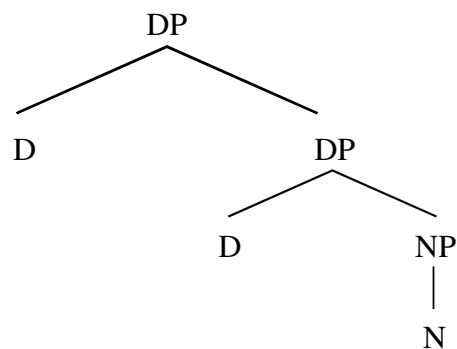
## ANÁLISE 2

Analise as expressões a seguir:

- |   |
|---|
| <ol style="list-style-type: none"><li>6. Ambos os lingüistas</li><li>7. Todos os lingüistas</li><li>8. Todos nós lingüistas</li><li>9. Todos vocês funcionalistas</li></ol> |
|---|

Para estas expressões teremos a seguinte representação arbórea:

(62)



Por meio dessa representação arbórea, notamos que nos exemplos de 6 a 9 há um DP gerado pelas regras de (63), o qual é complemento de um D. Essa construção é licenciada pela seguinte regra:

(63)

DP → D DP

Acompanhando com bastante atenção, notamos que já houve uma mudança nas regras em relação às da análise das estruturas das expressões nominais apresentadas na análise 1. O que vemos aqui é que teremos duas regras para o DP. Qual o motivo disso? A melhor explicação para esse caso, a nosso ver, está relacionada à presença dos quantificadores *todos* e *ambos*, pois, nesses casos, o DP passa a ser complemento de outro DP, e, teremos, assim, dois determinantes, e o complemento do D *os*, *nós* e *vocês* é um NP.

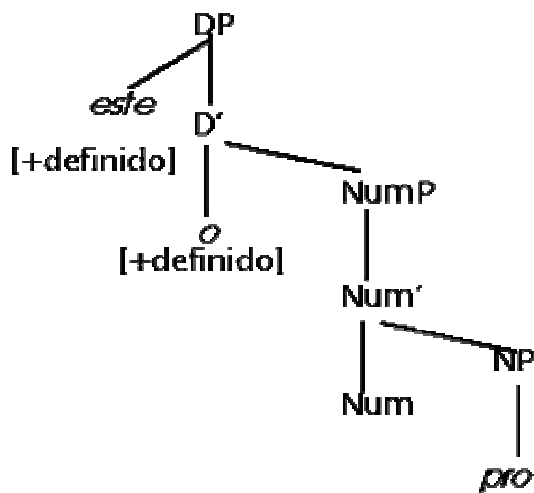
Como vimos na análise 1, o quantificador *alguns* opera diferentemente de *todos* e *ambos*, pois aquele não seleciona um DP como complemento, mas, sim, um NP ou um PP, como por exemplo em : *Alguns alunos* e *Alguns dos alunos*, respectivamente.

Analisando as expressões que são encabeçadas por *todos*, como, por exemplo, *Todos nós lingüistas* e *Todos vocês lingüistas*, observamos que os pronomes pessoais *nós* e *vocês* possuem a mesma distribuição dos artigos definidos, podendo, assim, licenciar, da mesma forma que eles, complementos do tipo NP.

Aproveitando o gancho, expomos aqui o que Martinho (1998) acrescenta para a literatura a respeito dos quantificadores e também dos demonstrativos. Este autor postula, primeiramente, que, em relação aos quantificadores, existe uma projeção funcional dedicada a elementos quantificadores, como numerais e cardinais, que, segundo Abney (1987), podem ser designadas por QP. Essa projeção funcional é complemento de D°, precede os adjetivos e recebe no seu especificador o elemento quantificador que lhe é associado. Esses quantificadores são projetados numa posição de especificador de uma categoria funcional.

Para este autor, da mesma forma, os demonstrativos, assim como os quantificadores, são especificadores, porém não são gerados nem em [Spec, DP], posição que reserva para elementos como o quantificador universal, com o qual o demonstrativo pode co-ocorrer livremente, como em: *todos estes meus amigos* (cf. Martinho, 1998, p.89); nem em D°. Vejamos em (64) a representação de estrutura sugerida por Martinho (1998, p.88), em que o artigo é gerado em D° e o demonstrativo ocupa a posição de especificador:

(64)

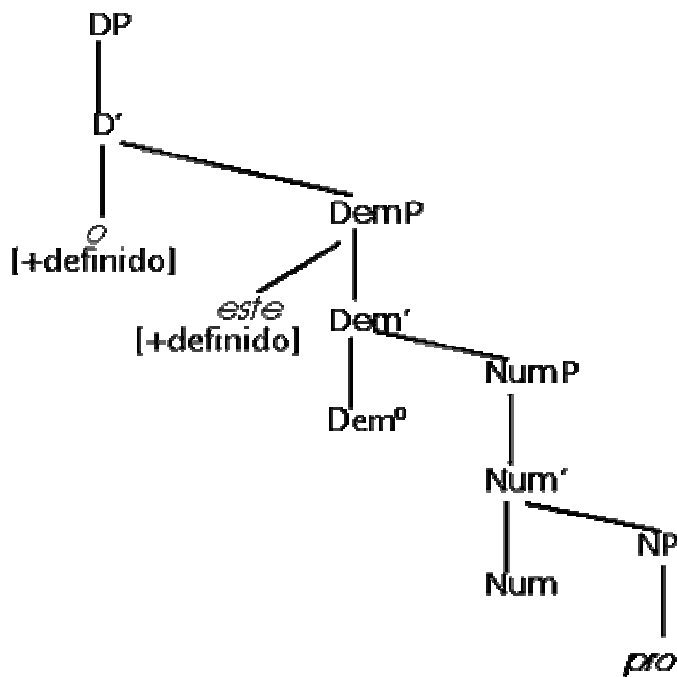


Apontamos um problema nesta representação pelo fato de haver co-ocorrência entre duas formas definidas, o artigo *o* e o demonstrativo *este*. O autor propõe que, em português, o demonstrativo, assim como o possessivo e os demais adjetivos atributivos, sejam também gerado em especificador de uma projeção funcional específica.

Martinho (1998) denomina DemP esta projeção funcional, e considera que é projetada acima de QP e dos possessivos, mas, abaixo de DP, de modo a explicar, por exemplo, a ordem de uma expressão do tipo: *Todos + Dem+Poss, Dem+Q e Dem +Poss+Q* (Martinho, 1998, p.89).

Vejamos esta nova representação feita por ele (Martinho, 1998, p.91):

(65)



(p.91)

Os mesmos comentários que fizemos à representação (64) devem ser feitos a (65), também, pois dois elementos definidos não podem co-ocorrer numa expressão nominal.

Para Martinho (1998), esta estrutura proposta por ele deve ser associada a mecanismos de restrição semântica, de modo a explicar a proibição de ocorrência de formas como *o este menino*, pois tanto *o* como *este* têm o traço [+definido], sendo assim, o artigo definido não pode ser projetado, pois, se for, a estrutura poderá ser rejeitada, talvez ao nível de Forma Lógica.

Por isso, para nossa pesquisa, é inviável nos utilizar destas representações propostas por Martinho, pois através do formalismo da DCG, haveria hipergeração de estruturas do tipo *o este menino*, que não é aceita para o português.

### ANÁLISE 3

Vejamos as expressões a seguir:

- |                    |
|--------------------|
| 10. O meu filho    |
| 11. Este meu livro |
| 12. Eu seu filho   |



Ao analisarmos as expressões nominais de 10 a 12, levaremos em conta, aqui, a questão referente aos pronomes possessivos.

No capítulo 1 da presente pesquisa, mostramos que Raposo (1992) e Cerqueira (1996) assumem que os pronomes possessivos não estão distribuídos do mesmo modo que os pronomes pessoais e demonstrativos, pois aqueles podem aparecer em uma determinada posição acompanhados por um artigo definido, o que nos faz, assim, dizer que estes pronomes não concorrem a uma posição de determinante.

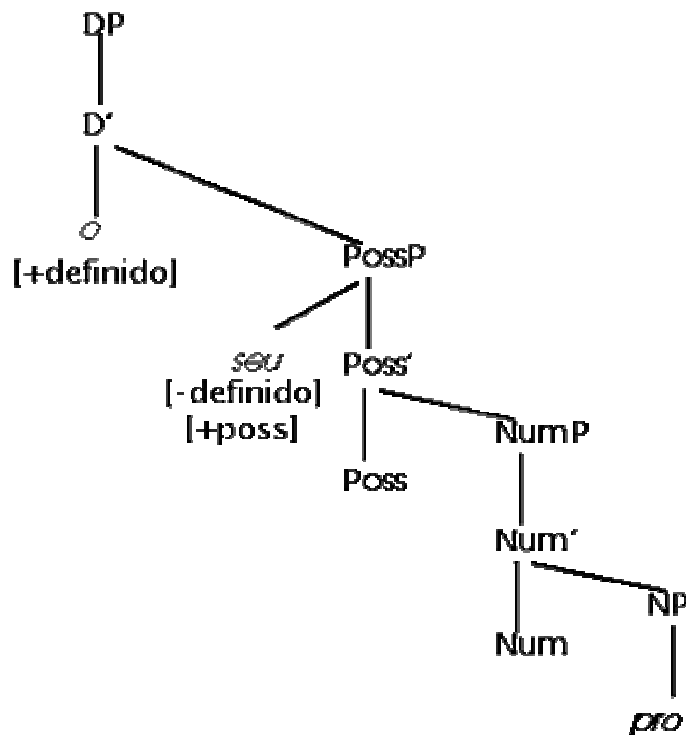
Ao leitor leigo, pode surgir a seguinte indagação: E em relação a construções do tipo *Meu amigo*, por exemplo, não teríamos aí o pronome pessoal “meu” assumindo a posição de determinante? A resposta é negativa.

Seguindo na esteira de Radford (1997), constatamos que a expressão nominal *meu amigo* é um DP, cujo núcleo está sendo representado por uma categoria vazia, ou seja, há, nesse caso, um determinante nulo, que pode ser representado por  $\emptyset$ , o qual, do mesmo modo que outras categorias possui também, propriedades gramaticais de seleção de complemento (definição, indefinição, número, pessoa etc.), apesar de esta categoria não estar representada foneticamente.

Em Martinho (1998) encontramos uma representação para os possessivos bem diferente da de Raposo (1992), por exemplo. Para aquele autor, o possessivo, em francês, por exemplo, parece variar entre o estatuto de especificador e o de núcleo. Para ele, os possessivos podem variar de língua para língua, sendo que, em umas é representado como núcleo de determinante, e em outras como especificadores funcionais ou até adjetivos, até porque os adjetivos não se distinguem morfologicamente dos pronomes. Em formas como *o meu*, por exemplo, em português, sugere que o artigo definido seja gerado em D<sup>o</sup>, e a forma possessiva *meu* no especificador da projeção funcional correspondente. Para ele, essa posição funcional é PossP, a qual está situada acima de QP e NumP, mas abaixo de DemP. Assim, em português, o possessivo pode ser visto exclusivamente como especificador.

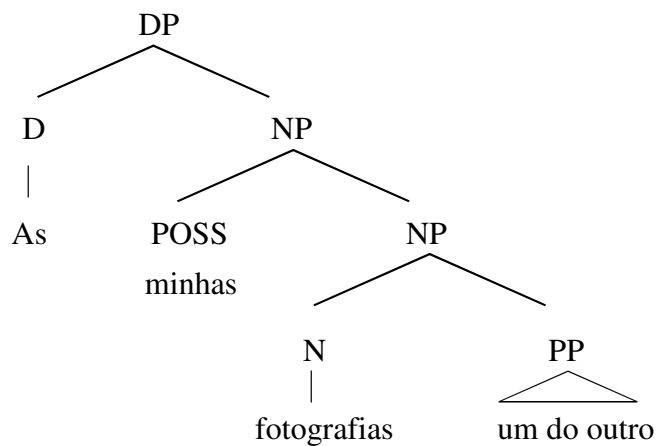
Observemos a representação dada por Martinho (1998, p.93), que representa o que dissemos acima:

(66)



Comparemos a representação (66) com a (67) de Raposo (1992) apresentada por nós no capítulo 1, como (15):

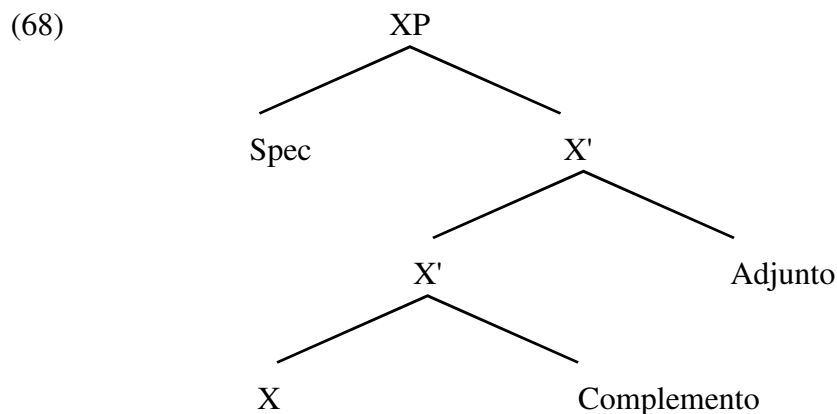
(67) As minhas fotografias um do outro



O que podemos dizer de tais representações? Primeiramente, deixamos claro que, em partes, optamos pela representação de Martinho, para a nossa presente análise, por este apresentar o pronome possessivo com um estatuto de categoria funcional e, não apenas, como um mero adjunto de NP.

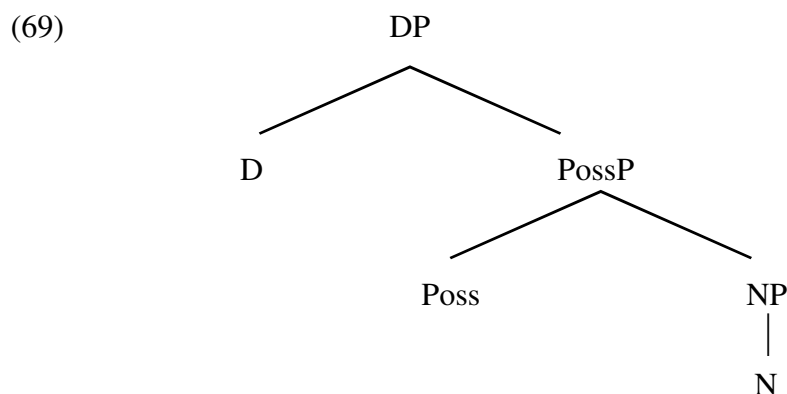
A importância de não analisarmos o possessivo como adjunto é que quando implementarmos nossas regras em DCG, será mais prático tomarmos o possessivo enquanto categoria funcional para evitar uma hipergeração no programa computacional. Se implementássemos em nosso programa uma regra do tipo, conforme a representação de Raposo (1992),  $NP \rightarrow Poss NP$ , o programa geraria estruturas agramaticais como: *minhas, tuas, nossas, suas... filhas*, pois temos aí um caso de regra recursiva, ou seja, conforme Othero (2006,p.54), “as regras recursivas são aquelas que permitem a repetição de um constituinte *ad infinitum*, já que a gramática de um língua pode gerar sentenças infinitas a partir de um conjunto infinito de regras”. Porém, tal fato não acontece, por exemplo, com a regra com a qual estamos lidando  $DP \rightarrow D PossP$ , pois esta regra não hipergera estruturas agramaticais.

O nosso programa computacional que mostraremos no capítulo 4 não faz distinção entre complementos e adjuntos. De acordo com a teoria X-barra, há uma diferença na estrutura sintagmática entre complementos e adjuntos que pode ser percebida nas relações de dominância:



O complemento sempre estará mais próximo ao núcleo, sendo irmão de X, enquanto que o adjunto vem ao lado do constituinte intermediário, sendo irmão de X'.

Assim, em nossa pesquisa, fica claro para nós que o NP é complemento da categoria funcional Poss, que projeta PossP, por sua vez, um complemento de D:



Teremos para essa representação, duas novas regras:

(70)  $DP \rightarrow D \text{ PossP}$

(71)  $\text{PossP} \rightarrow \text{Poss NP}$

#### ANÁLISE 4

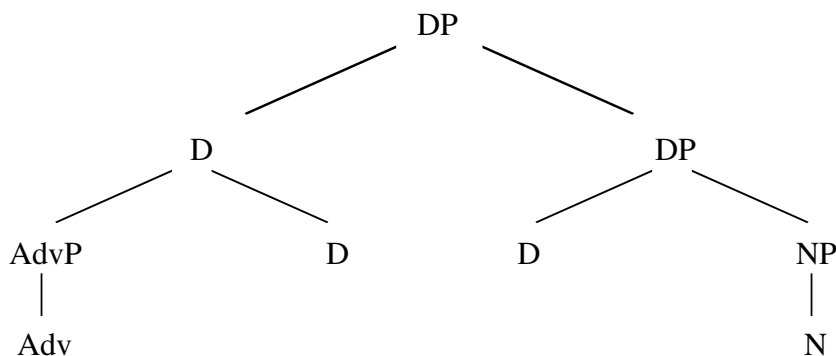
O que dizermos das expressões seguintes?

13. Quase todos nós alunos  
14. Quase todos vocês médicos

Nesta análise, teremos o acréscimo do advérbio *quase* à estrutura da análise 2. Para Raposo (1992) como vimos na sessão 1.1.3 deste trabalho, existem elementos que podem modificar a categoria lexical não através da projeção de um nível adicional da teoria X-barra, mas sim por uma configuração de adjunção. Em versões anteriores da teoria X-barra, algumas dessas categorias eram apresentadas como especificador, mas, na verdade, são adjuntos, como, por exemplo, advérbios e adjetivos.

Nesta análise, postulamos que o *quase* vem à esquerda de *todos*, e o produto não passa de um processo de adjunção:

(72)



A diferença entre as regras desta análise para a análise 2, é que o D mais alto, agora, consiste de um D adjungido a um advérbio. Neste caso acrescentaríamos apenas uma regra para o D, que tenha a possibilidade dessa adjunção, vejamos:

(73)  $D \rightarrow \text{Adv D}$

## ANÁLISE 5

Na expressão nominal seguinte, não encontraremos nenhuma novidade em relação ao que já foi exposto na análise de 2. O que teremos é a ocorrência do sintagma adjetival (AP) após o N *lingüistas*. Este AP não é um complemento de N, mas sim um adjunto de NP, conforme Raposo (1992, p.199) e Mioto *et al* (1999, p.95):

15. Todos nós lingüistas brasileiros

As regras que acrescentamos, a partir dessa expressão, para a nossa minigramática, são as seguintes:

(74) NP → N AP

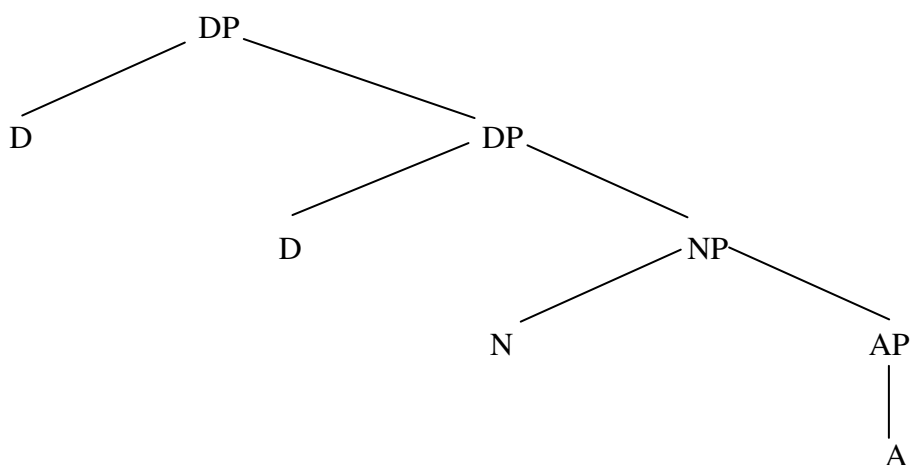
(75) AP → A

A regra de (74) não é encontrada, por exemplo, em Raposo (1992), em quem nos temos baseado nessa pesquisa, mas o formalismo com o qual estamos trabalhando, o da DCG, impõe o uso desse tipo de regra, para que não haja hipergeração, do mesmo modo que na análise 3.

Compactuamos com Othero (2006, p.61) que ainda faltam explicações e descrições mais claras para a distinção entre complemento e adjunto. Assim, não iremos fazer, posteriormente, essa distinção em nossa gramática.

Vejamos a representação arbórea para essa estrutura:

(76)



## ANÁLISE 6

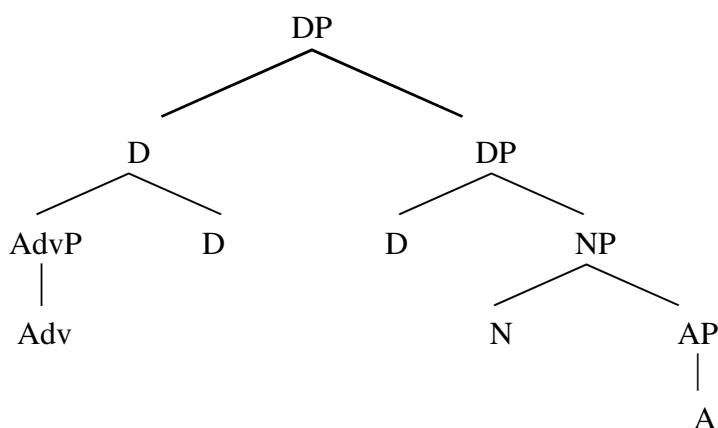
Ao observarmos as análises 4 e 5, obteremos, com a junção das regras de uma e outra a seguinte expressão nominal:

16. Quase todos nós lingüistas brasileiros

Aqui a partir dessa estrutura, não teríamos nenhuma regra nova para acrescentar ao nosso *parser*, pois tais regras já se encontram nas análises 4 e 5.

Eis a representação arbórea:

(77)



## ANÁLISE 7

De acordo com Mateus *et al* (1989, p. 294), as orações relativas apositivas de uma sentença F são aquelas que constituem um comentário acerca de uma proposição anterior. Porém, muitas vezes, aquilo sobre o que a relativa apositiva asserta não é em relação a tudo o que foi dito antes, mas, sim, a uma parte da proposição.

Para essas autoras, estas orações podem ser introduzidas por *N+DET+ que*, como é o caso do exemplo de nossas expressões nominais a seguir:

17. Lingüista aquele que...

18. Teoria esta que...

Mateus *et al* (1989, p.294) dizem que nesses casos os Ns são determinados simultaneamente pelo DET (presumivelmente movido para a direita de N) e pela relativa restritiva.

O que encontramos em nossas expressões é uma inversão bem restrita para a língua portuguesa pois, apesar de esta ser uma língua de núcleo inicial, encontramos esta construção excepcional. Esta estrutura, na verdade, está presa a uma estrutura maior CP, pois há um DP inserido em uma oração relativa.

Há, assim, uma limitação em nosso programa ao analisar expressões nominais desse tipo, pois não entraremos aqui em detalhes a respeito de estruturas maiores como CP. Porém investigações futuras poderão vir a eliminar essa deficiência.

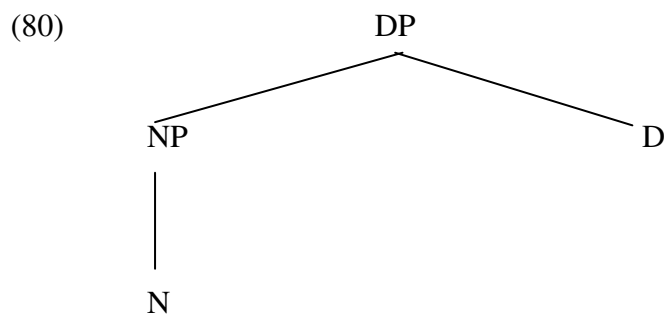
Encontramos, ainda, um exemplo do mesmo tipo, em Martinho (1998, p.89), ao propor para os demonstrativos um estatuto de projeção funcional DemP:

(78) O Nome da Rosa foi o livro que mais vendeu em 1982, **livro este que** tornou o seu autor mundialmente conhecido.

Para nossa minigramática, teremos uma nova regra:

(79) DP → NP D

Observemos a estrutura arbórea a seguir:



### ANÁLISE 8

Nessa análise teremos uma subcategorização diferente para o *alguns* e *nenhuma*, os quais podem vir seguidos ou não por preposição. Na análise 2, encontramos, por exemplo, o *alguns* seguido por NP, e não por PP.

- |                        |
|------------------------|
| 19. Alguns dos alunos  |
| 20. Nenhuma das filhas |

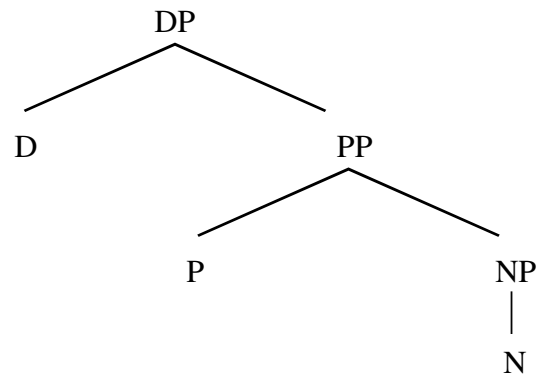
Assim, teríamos, a partir dessa análise, novas regras para nossa minigramática:

(81) DP → D PP

(82) PP → P NP

A representação arbórea para essa estrutura é a seguinte:

(83)



Para a elaboração dessas regras, seguimos uma alternativa de Pagani (2004, p.18) de que no léxico já existem cinco versões para as preposições: *de, da, das, do e dos*. Para esse autor “normalmente, se costuma explicar essas últimas quatro formas da preposição como se elas fossem o resultado da amalgamação da preposição *de* com uma das quatro versões do artigo definido”(PAGANI, 2004, p.18). Assim, é como se as informações veiculadas pelo D, já se encontrassem em P.

Pagani (2004, p.20) ainda sugere o seguinte, levando-se em conta todas as preposições:

Esse tipo de solução para a distribuição das preposições sugere que existem dois grandes subconjuntos de preposições: as que são completamente inflexionáveis e as que aceitam flexão. Esse segundo subconjunto, por sua vez, se subdivide em outros dois subconjuntos: aquelas marcadas explicitamente com a flexão nominal e aquelas que não são marcadas, mas que, por isso mesmo, são compatíveis com qualquer concordância.

## ANÁLISE 9

Vejamos as expressões seguintes:

21. Nenhuma das minhas filhas

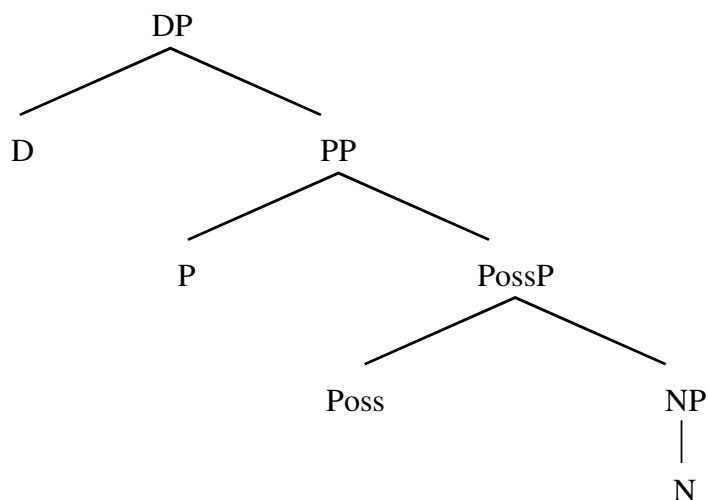
22. Alguns dos meus amigos

Ao analisar tais estruturas, a novidade que encontramos é a inclusão de uma projeção funcional PossP entre o PP e o NP. Esse PossP passa a ser complemento do P *das*. A seguir, o Poss, núcleo de PossP toma um NP como complemento.



Vejam a representação em árvore:

(84)



Temos a nova regra para essa estrutura;

(85)  $PP \rightarrow P \text{ PossP}$

### ANÁLISE 10

A expressão que iremos analisar agora é a seguinte:

23. Oito de nós

Para esta análise, teremos três novas regras para a nossa gramática:

(86)  $DP \rightarrow D \text{ QP}$

(87)  $QP \rightarrow Q \text{ PP}$

(88)  $PP \rightarrow P \text{ DP}$

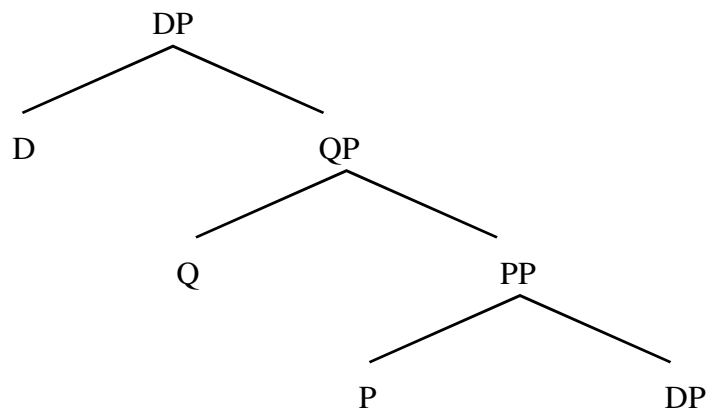
Consideramos o *oito* como um quantificador o qual não co-ocorre com um artigo, pois vejamos:

Os oito alunos  
Uns oito alunos.

Observemos que o artigo e o numeral co-ocorrem sem nenhum problema. Desse modo *oito* é um quantificador e não um determinante. Mais uma vez, teremos para a estrutura 23 um

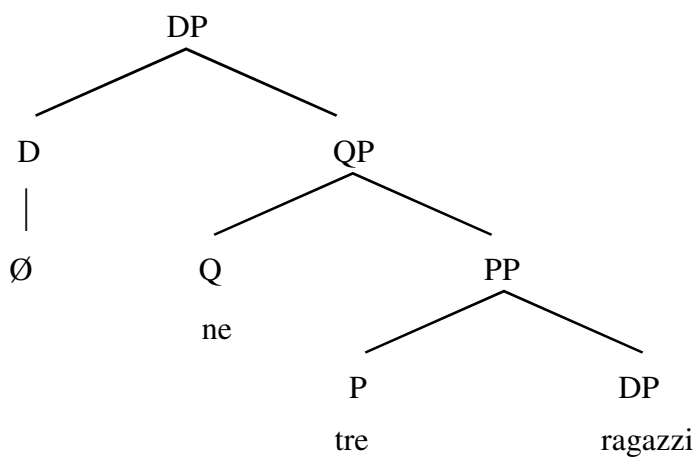
determinante vazio, anteriormente ao quantificador *oito*, que determina os traços de complemento para essa expressão nominal:

(89)



Quem também utiliza-se de representação semelhante para estruturas de línguas como o italiano, é Coene e D'hulst (2003, p.12);

(90)



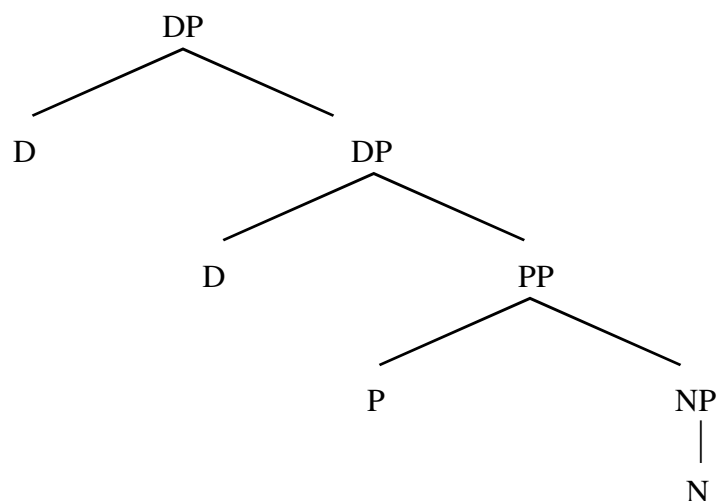
### ANÁLISE 11

Na expressão abaixo, teremos apenas a inclusão do determinante *cada* à regra semelhante à da análise 8. Assim, o D *cada* passa a selecionar outro DP enquanto complemento, do mesmo modo que *todos* e *ambos* na análise 2.

24. Cada um dos livros

Assim, como era de se esperar, não teremos novas regras para essa análise, pois é a junção das regras 2 e 8:

(91)



## ANÁLISE 12

Para essas estruturas, temos uma pequena diferença em relação a exemplos da análise 3:

25. Um filho meu  
26. Alguns amigos meus

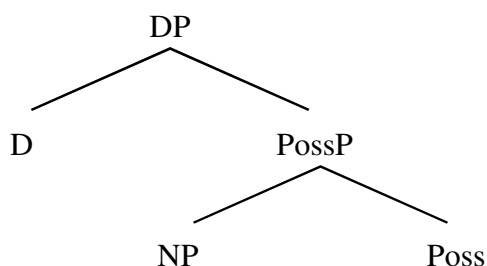
O que mostramos aqui é a mudança de posição do possessivo, que passa de pré-nominal para pós-nominal. Ou seja, algo inverso à regra gerada através da análise de *O meu filho*.

Para isso é gerada a seguinte regra:

(92) PossP → NP Poss

A representação arbórea, pois, é a seguinte:

(93)



### ANÁLISE 13

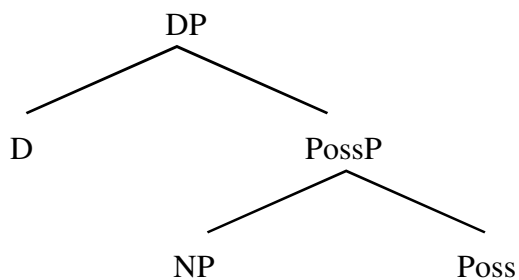
Qual a diferença de estrutura e representação entre as expressões 25 e 26 , e a seguinte? Para o leitor atento, está claro que a questão dessa análise é referente à omissão do D. O que fazer então? Nesse caso, postulamos, a partir de Radford (1997), que há um D vazio não representado foneticamente.

Teríamos, então, um D foneticamente vazio que toma uma categoria PossP como seu complemento:

27. Filho meu

Vejam a representação em árvore:

(94)



Observamos que é a mesma árvore que representa as expressões nominais 25 e 26 da análise 12, por isso, não teremos introdução de novas regras.

### ANÁLISE 14

Vejam o que temos a seguir:

28. Homem nenhum  
29. Mulher alguma

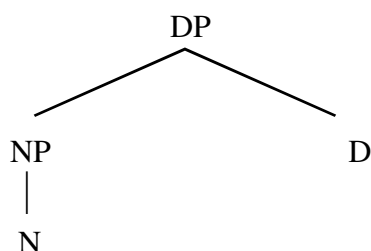
O bom leitor deve ter observado que, agora, partimos para uma espécie de análise “às avessas”. Pois, até então, as expressões analisadas possuem uma estrutura cujos determinantes ocorrem pré-nominalmente.

No caso das expressões nominais 28 e 29, o D ocorre pós NP, formando, assim, uma construção bastante rara em língua portuguesa.

A regra que temos para essa representação é a mesma utilizada para a representação das expressões nominais da análise 7, *lingüistas aqueles* e *teoria esta*. Assim, não teremos regra nova para acrescentar á nossa minigramática.

Observemos a estrutura arbórea a seguir:

(95)



O que foi gerado aqui não é nada mais, nada menos que a representação gerada na análise 1, contudo, o NP passa para o lado esquerdo e o D para o lado direito, gerando estruturas do tipo: *Homem nenhum, Mulher alguma, etc.*

### ANÁLISE 15

Conforme Mateus *et al* (1989, p.195), a posição ocupada pelos determinantes não é exclusiva dos artigos, quantificadores e demonstrativos. Observemos que as expressões qualitativas analisadas logo abaixo têm a seguinte estrutura: *DET + ADJ + de + N*:

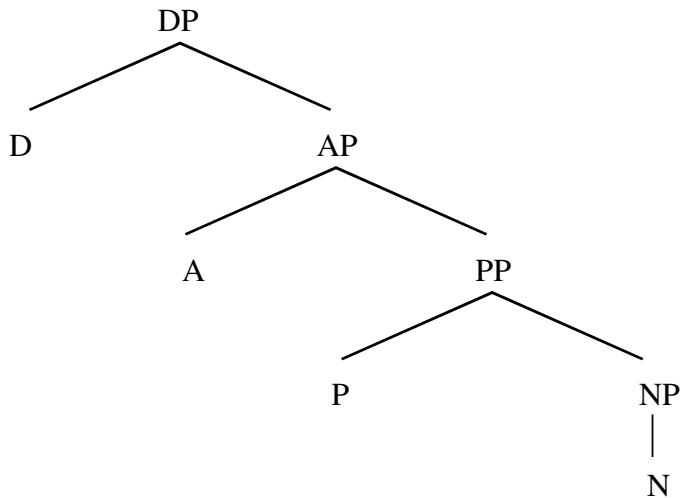
- |   |
|---|
| 30. Aquele imbecil do porteiro<br>31. Este idiota deste velho |
|---|

O que podemos observar é que há uma alteração de sentido quando o adjetivo da construção qualitativa é colocado após o núcleo: *O porteiro imbecil* e *Este velho idiota*.

Algo importante a ser observado, é que, o D, de um modo geral, não só toma NP como complemento, mas também outros constituintes, como constatamos aqui deparamo-nos com um D cujo complemento é um AP. O DP passa a ser rico do mesmo modo como o VP em relação a sua valência.

Vejam a representação arbórea:

(96)



Temos, assim, novas regras para nosso parser:

(97)  $DP \rightarrow D \ AP$

(98)  $AP \rightarrow A \ PP$

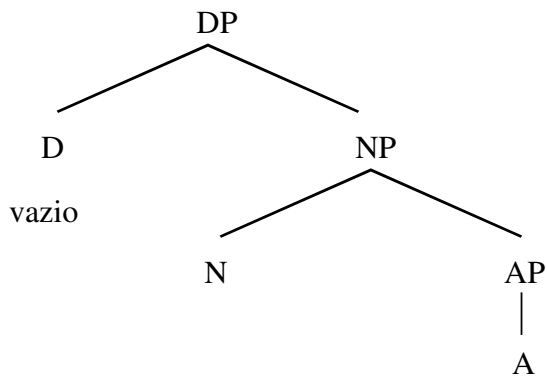
### ANÁLISE 16

Apresentamos aqui uma estrutura restrita para o PB:

32. Menina rica

Observamos que nessa expressão nominal, há propriedades semânticas em um D vazio que não permitem que essa estrutura seja gerada para uma menina “qualquer”. Refere-se a uma classe e não a um indivíduo propriamente.

(99)



Para essa representação, não teremos novas regras para nossa minigramática, já que não passa da junção entre as regras das análises 1 e 5, já apresentadas anteriormente.

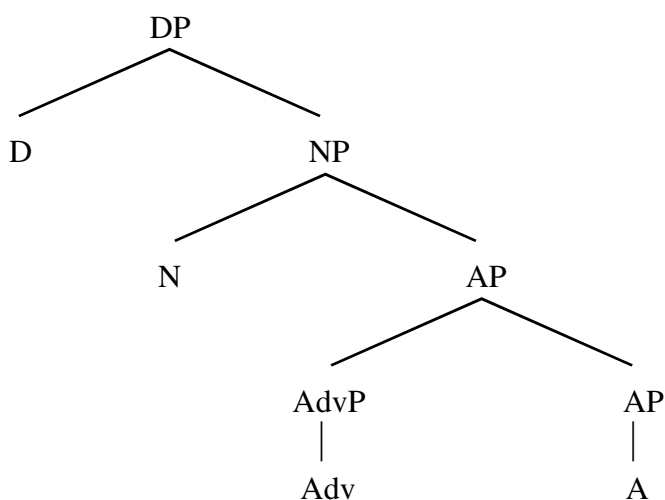
## ANÁLISE 17

Nessa análise, há a introdução de um adjunto representado por um AdvP *mais*, o qual é denominado, tradicionalmente, como intensificador e modifica, nesta representação, o AP *belas*. Tradicionalmente, os intensificadores são denominados como adjuntos, assim como os adjetivos. Conforme Mioto *et al* (2005, p.104), esses intensificadores pendem do AP pelo lado esquerdo, diferentemente do que acontece com os adjuntos de VP ou NP, que normalmente pendem pelo lado direito:

33. As meninas mais belas

Em um processo de adjunção, um dos filhos deve ser sempre idêntico ao pai, conforme Raposo (1992, p.207). Observando a representação arbórea, a seguir, teremos AP, que é o pai, cujos filhos são AdvP e AP:

(100)



As novas regras são as seguintes:

(101) AP → AdvP A

(102) AdvP → Adv

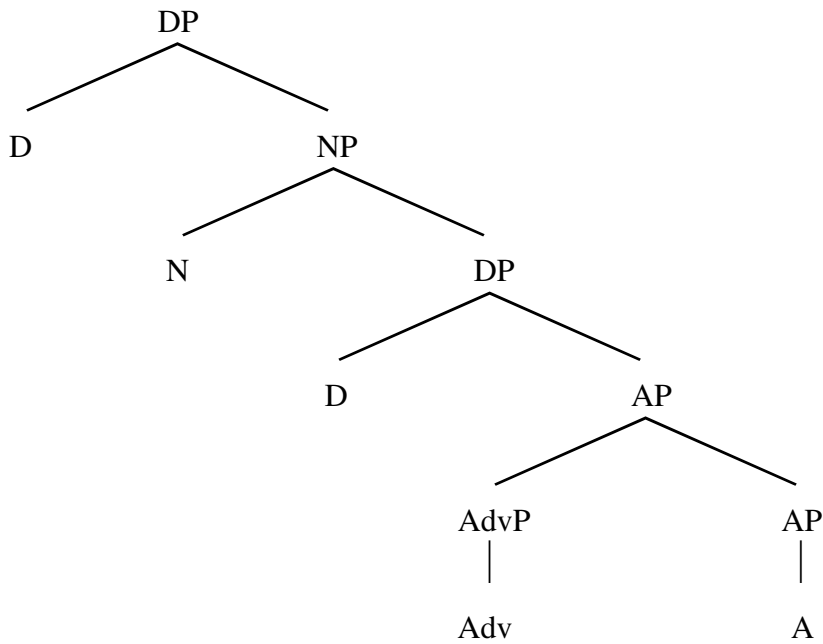
## ANÁLISE 18

O que encontramos a *mais* na expressão nominal seguinte é a inclusão de um determinante *as*, entre o NP e o AP, dando-nos, assim, uma nova regra:

(103) NP → N DP

34. As meninas as mais belas

(104)



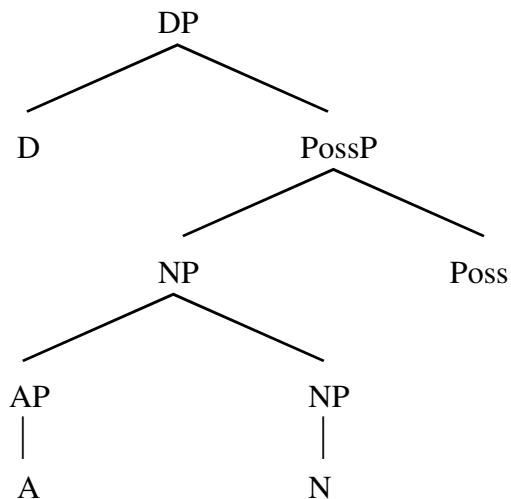
### ANÁLISE 19

Consideramos, nesta análise, o item *outra* como um AP o qual vem adjungido ao NP *irmã*.

35. A outra irmã minha

Esta representação, gera, assim, uma nova regra para nossa gramática:

(105)



A nova regra é:

(106) NP → AP NP

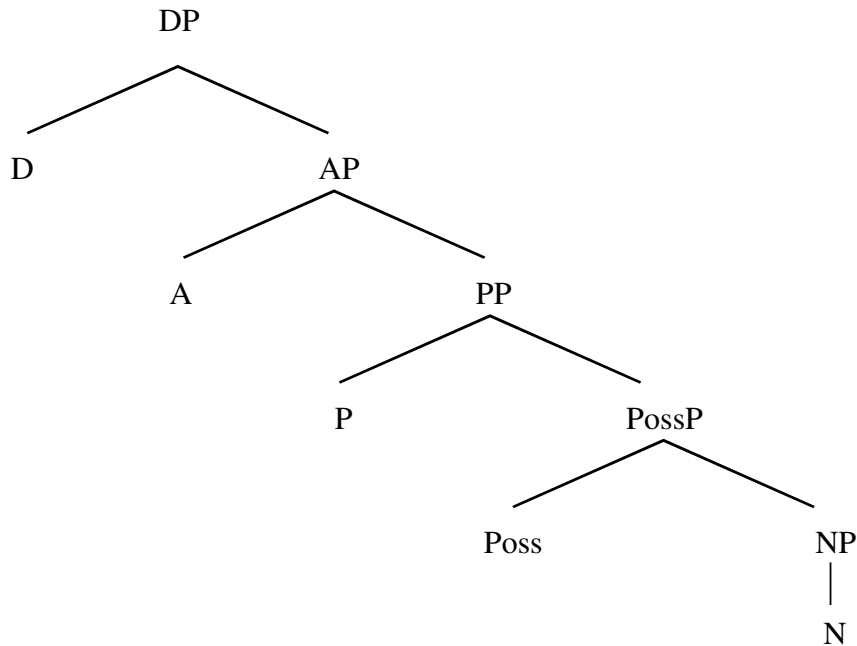


## ANÁLISE 20

A análise a seguir acrescenta, tão somente, em relação à análise 15, um PossP entre o PP e o NP:

36. A safada da minha filha

(107)



A nova regra para implementar em nossa minigramática é a seguinte:

(108)  $PP \rightarrow P \text{ PossP}$

A seguir, mostraremos a gramática livre de contexto, ou seja a CFG que construímos a partir das regras geradas pelas expressões nominais analisadas:

### 3.3. Fragmento de gramática livre de contexto CFG:

#### **Regras**

$NP \rightarrow N$

$NP \rightarrow N \text{ AP}$

$NP \rightarrow N \text{ DP}$

$NP \rightarrow \text{AP NP}$

DP → D

DP → D NP

DP → D DP

DP → D PossP

DP → NP D

DP → D PP

DP → D QP

DP → D AP

QP → Q PP

PP → P NP

PP → P PossP

PP → P DP

PossP → Poss NP

PossP → NP Poss

AdvP → ADV

AP → A

AP → A PP

AP → AdvP A

### **Léxico**

D → todos, alguns, esta, eu, tu, ele, ela, nós, vocês, eles, elas, ambos, os, eles, o, este, tu, aquele, alguns, nenhuma, um, uns, nenhum, alguma, as, a, cada.

N → lingüista, lingüistas, alunos, teoria, gerativistas, funcionalistas, médicos, filho, livro, livros, filhas, filha, amigos, mulher, homem, rapaz, porteiro, menina, meninas, irmã, velho

A → brasileiros, rica, belas, idiota, safada, imbecil, outra

Adv → quase, mais

P → de, da, das, do, dos, deste

Poss → meu, meus, minha, minhas, seu

Q → oito

Para gerarmos esse fragmento de gramática de expressões nominais do português do Brasil, fizemos o seguinte: primeiramente inserimos as regras que foram sendo geradas durante a análise das 36 expressões nominais acima, como, por exemplo:

(109)  $DP \rightarrow D NP$

A regra (109) quer dizer que um Sintagma Determinante consiste de um determinante e de um sintagma nominal, que, por sua vez é complemento de D. A partir dessa regra, devemos elaborar outras para o NP, conforme a representação (110):

(110)  $NP \rightarrow N$

Para nosso *parser* ser eficiente, contudo, não podemos parar por aí, pois teremos também que implementar os itens lexicais, conforme a seguir:

(111)

$D \rightarrow a, o, as, os$

$N \rightarrow médicos, lingüistas, aluno$

Essa primeira versão que apresentamos de nossa gramática não permite que entremos em detalhes de programação em Prolog, mas, com base nessa gramática, já conseguimos reconhecer algumas expressões nominais do tipo:

(112) Nós lingüistas

(113) Vocês funcionalistas

(114) Eu meu Deus

A hipótese de que os pronomes pessoais são núcleos de DP já foi implementada nessa primeira versão dessa gramática, garantindo, assim, a geração e reconhecimento dessas expressões como gramaticais. Mais à frente, mostraremos a deficiência desse formalismo no que se refere à questão de hipergeração de estruturas, como, por exemplo:

(115) \* As menino

Contudo, no decorrer desse trabalho, mostraremos como resolver tais problemas provenientes de limitações da nossa gramática esboçada no quadro acima.

Um grande problema que enfrentaremos, conforme se observa no exemplo (115) é o referente à concordância nominal, que, de acordo com o modelo apresentado acima, é uma expressão considerada gramatical, apesar de sabermos que em língua portuguesa é agramatical.

Para resolvermos esse tipo de problema, teremos que alterar as regras sintagmáticas para que possam levar em consideração elementos referentes a concordância nominal, ou seja, teremos que atribuir traços morfossintáticos às regras sintagmáticas, e, também, modificar o léxico para que haja uma coerência entre ambos.

Ainda no que diz respeito ao abastecimento do léxico para nossa gramática, vejamos os exemplos abaixo:

(116) Nós lingüistas

(117) Nós alunos

(118) Nós artistas

Por que o nosso *parser* aceitaria as expressões nominais (116) e (117) e não a (118)? Aparentemente, essas três expressões deveriam ser aceitas, pois têm estruturas sintáticas idênticas, porém, para nosso reconhecedor gramatical a palavra *artistas* não consta em nosso léxico, por isso essa expressão não é reconhecida, por não se encontrar em nossa base de dados.

# CAPÍTULO 4

## IMPLEMENTAÇÃO COMPUTACIONAL DE EXPRESSÕES NOMINAIS DO PB EM PROLOG

---

A seguir, mostraremos, passo a passo, como transformar a CFG apresentada no capítulo 3 em um analisador gramatical em DCG capaz de gerar, reconhecer e analisar essas estruturas nominais em língua portuguesa.

O que faremos agora, é representar as regras acima que se encontram descritas em modelo de CFG em linguagem Prolog, por meio de um formalismo, o qual já apresentamos no capítulo 1 desta dissertação, conhecido por DCG.

Aqui mostraremos apenas a conversão propriamente dita de um formalismo para outro, sem mexer, ainda, nas regras e no léxico, pois, como dissemos no capítulo 3, aparecerão alguns obstáculos durante o processo de desenvolvimento de um *parser* em Prolog. Porém, as estratégias para resolução desses problemas que estão, principalmente, relacionadas às regras de descrição dos sintagmas em português, somente apresentaremos quando mostrarmos o analisador final com traços, que é nossa última etapa.

Passemos para o reconhecedor:

### 4.1. Reconhecedor gramatical do português no formato DCG sem traços

```
%RECONHECEDOR GRAMATICAL SEM TRAÇOS DE EXPRESSÕES  
NOMINAIS EM LP  
%AUTORA: KARINE ALVES DAVID  
%DATA: 29/01/2007.  
  
%corpus 1  
/*  
Expressões nominais em Língua portuguesa  
*/  
corpus(1,[todos,lingüistas]).  
corpus(2,[alguns,alunos]).  
corpus(3,[esta,teoria]).  
corpus(4,[nós,gerativistas]).  
corpus(5,[vocês,functionalistas]).
```

corpus(6,[ambos,os,lingüistas]).  
corpus(7,[todos,os,lingüistas]).  
corpus(8,[todos,nós,lingüistas]).  
corpus(9,[todos,vocês,funcionalistas]).  
corpus(10,[o,meu,filho]).  
corpus(11,[este,meu,livro]).  
corpus(12,[eu, seu, filho]).  
corpus(13,[quase,todos,nós,alunos]).  
corpus(14,[quase,todos,vocês,médicos]).  
corpus(15,[todos,nós,lingüistas,brasileiros]).  
corpus(16,[quase,todos,nós,lingüistas,brasileiros]).  
corpus(17,[lingüista,aquele]).  
corpus(18,[teoria,esta]).  
corpus(19,[alguns,dos,alunos]).  
corpus(20,[nenhuma,das,filhas]).  
corpus(21,[nenhuma,das,minhas,filhas]).  
corpus(22,[alguns,dos,meus,amigos]).  
corpus(23,[oito,de,nós]). %problema  
corpus(24,[cada,um,dos,livros]).  
corpus(25,[um,filho,meu]).  
corpus(26,[alguns,amigos,meus]).  
corpus(27,[filho,meu]). %problema  
corpus(28,[homem,nenhum]).  
corpus(29,[mulher,alguma]).  
corpus(30,[aquele,imbecil,do,porteiro]).  
corpus(31,[este,idiota,deste,velho]).  
corpus(32,[menina,rica]).  
corpus(33,[as,meninas,mais,belas]).  
corpus(34,[as,meninas,as,mais,belas]).  
corpus(35,[a,outra,irmã,minha]).  
corpus(36,[a,safada,da,minha,filha]).

%SINTAXE%

np--> n.

np--> n,ap.

np--> n,dp.

np--> ap,np.

dp--> d.  
dp--> d,np.  
dp--> np..  
dp--> d,dp.  
dp--> d,poss.  
dp--> poss.  
dp--> np,d.  
dp--> d,pp.  
dp--> d,qp.  
dp--> qp.  
dp--> d,ap.

qp--> q,pp.

pp--> p,np.  
pp--> p,poss.  
pp--> p,dp.

poss--> poss,np.  
poss--> np,poss.

advp--> adv.

ap--> a.  
ap--> a,pp.  
ap--> advp,a.

%LÉXICO%

d-->[o].  
d--> [os].  
d--> [a].  
d--> [as].  
d--> [um].  
d--> [um].  
d--> [uns].  
d--> [este].

d--> [esta].

d--> [aquele].

d--> [nenhum].

d--> [nenhuma].

d--> [alguma].

d--> [alguns].

d--> [todos].

d--> [ambos].

d--> [cada].

d--> [eu].

d--> [tu].

d--> [você].

d--> [ele].

d--> [ela].

d--> [nós].

d--> [vocês].

d--> [elas].

d--> adv,d.

poss--> [meu].

poss--> [meus].

poss--> [minhas].

poss--> [minha].

poss--> [seu].

q--> [oito].

a-->[rica].

a-->[safada].

a-->[idiota].

a-->[belas].

a-->[imbecil].

a-->[brasileiros].

a-->[outra].



adv--> [quase].

adv--> [mais].

p--> [de].

p--> [do].

p--> [dos].

p--> [das].

p--> [da].

p--> [deste].

n--> [homem].

n--> [mulher].

n--> [filho].

n--> [filhas].

n--> [filha].

n--> [menina].

n--> [meninas].

n--> [irmã].

n--> [porteiro].

n--> [lingüista].

n--> [alunos].

n--> [lingüistas].

n--> [funcionalistas].

n--> [gerativistas].

n--> [amigos].

n--> [médicos].

n--> [livro].

n--> [livros].

n--> [teoria].

n--> [velho].

De acordo com o que se pode observar, construímos um reconhecedor gramatical, levando em conta a sintaxe e o léxico da CFG de 3.3. Porém, anteriormente à sintaxe e ao léxico desse reconhecedor, inserimos dois *corpora*: um que consiste das 36 expressões nominais analisadas por nós, e outro que consiste de 16 expressões nominais que consideramos agramaticais em língua portuguesa.

Essa inserção desses *corpora* justifica-se pelo fato de termos um melhor manuseio com o programa ao solicitarmos informações a respeito dessas expressões.

Seguimos modelos de gramáticas propostos por Othero e Menuzzi (2005), Othero (2006), Pagani (2003), e Vieira e Lima (2001), para podermos desenvolver nosso *parser* de expressões nominais do português do Brasil. O formalismo que utilizaremos, como já discutimos na sessão 1.4. é chamado de DCG e está presente em quase todas as versões de Prolog atualmente distribuídas.

É graças à DCG que poderemos converter as regras da CFG apresentadas no quadro 4.1. para a notação de Prolog, pois é um formalismo que serve para representar gramáticas livres de contexto (Vieira e Lima, 2001, p.18).

Vejam como Othero e Menuzzi (2005, p.77) explicam de forma bem didática como essas regras são convertidas:

Tudo o que se tem de fazer é escrever os rótulos dos constituintes em letras minúsculas (caso contrário, o Prolog os trataria como variáveis), substituir a flecha “→” por “-->”, separar os constituintes à direita da flecha por uma vírgula e, finalmente, lembrar de colocar um ponto final em cada regra.

Na verdade, o Prolog não vai operar diretamente com as regras, pois, na verdade, há um processo de tradução dessas regras para um formato de listas, e é com esse formato que o Prolog vai operar diretamente.

O que acabamos de apresentar foi um reconhecedor gramatical. Como se dá, então, esse reconhecimento? Conforme apresentamos na sessão 1.4.1., faremos ao Programa uma consulta, através de perguntas ou diretivas, como bem coloca Pagani (2004). Por exemplo: `dp([todos,os,lingüistas],[,])`, que trocando em miúdos, temos o seguinte questionamento ao programa: *Todos os linguistas é um DP?* O programa nos dará uma das respostas: *Yes ou No*, ou seja, reconhecerá, ou não, a expressão nominal em evidência como um DP.

Vejam, detalhadamente, como isso acontece:

(119)

<p>?- dp([todos,lingüistas],[,]). Yes.</p>
--

O programa responde afirmando, porque apresenta, dentre outras, a seguinte regra para o DP:  $DP \rightarrow D NP$ .

Durante esse processo de reconhecimento, o reconhecedor observa o DP, vai à procura do NP, depois de N. Esse processo é conhecido como *topdown*, pois é descendente, analisa da esquerda para a direita, e isto se torna um problema quando nos deparamos com regras recursivas à esquerda, pois as regras gramaticais em Prolog não conseguem lidar com a recursão à esquerda. Conforme Othero (2006, p.118)

Não é nada que vá causar um erro sintático, mas se o sistema encontra uma recursão à esquerda, então um *looping* sem fim irá começar. Depende, portanto, do programador evitar recursão à esquerda em seus programas.”

Para que o nosso reconhecedor gere uma estrutura, além de responder apenas *yes* ou *no*, basta lançarmos a seguinte diretiva: `dp(X,[])`. Então, teremos como resposta o seguinte: *todos lingüistas, esta teoria, nós gerativistas etc.*

Veremos mais adiante que a DCG é uma gramática bem mais poderosa que a CFG, pois através dela podemos acrescentar argumentos extras que podem trazer os traços morfossintáticos para serem implementados em uma determinada gramática.

Mostraremos a seguir, como transformar esse reconhecedor, em um analisador gramatical.

#### 4.2. Analisador gramatical do português no formato DCG sem traços

A nova versão da DCG que apresentaremos a seguir é o que chamamos de analisador gramatical.

Para transformar o reconhecedor de 4.1. em um analisador gramatical, basta acrescentarmos um argumento em cada um dos predicados usados, de forma a construir nesse argumento uma representação da estrutura gramatical. Observemos o analisador a seguir:

```
%ANALISADOR GRAMATICAL SEM TRAÇOS DE EXPRESSÕES NOMINAIS
EM LP

%AUTORA: KARINE ALVES DAVID

%DATA: 25/01/2007.

%corpus      1
/*
Expressões nominais em Língua portuguesa      GRAMATICAIS
*/

corpus(1,[todos,lingüistas]).
corpus(2,[alguns,alunos]).
corpus(3,[esta,teoria]).
corpus(4,[nós,gerativistas]).
corpus(5,[vocês,functionalistas]).
corpus(6,[ambos,os,lingüistas]).
corpus(7,[todos,os,lingüistas]).
```

corpus(8,[todos,nós,lingüistas]).  
corpus(9,[todos,vocês,funcionalistas]).  
corpus(10,[o,meu,filho]).  
corpus(11,[este,meu,livro]).  
corpus(12,[eu,seu,filho]).  
corpus(13,[quase,todos,nós,alunos]).  
corpus(14,[quase,todos,vocês,médicos]).  
corpus(15,[todos,nós,lingüistas,brasileiros]).  
corpus(16,[quase,todos,nós,lingüistas,brasileiros]).  
corpus(17,[lingüista,aquele]).  
corpus(18,[teoria,esta]).  
corpus(19,[alguns,dos,alunos]).  
corpus(20,[nenhuma,das,filhas]).  
corpus(21,[nenhuma,das,minhas,filhas]).  
corpus(22,[alguns,dos,meus,amigos]).  
corpus(23,[oito,de,nós]). %problema  
corpus(24,[cada,um,dos,livros]).  
corpus(25,[um,filho,meu]).  
corpus(26,[alguns,amigos,meus]).  
corpus(27,[filho,meu]). %problema  
corpus(28,[homem,nenhum]).  
corpus(29,[mulher,alguma]).  
corpus(30,[aquele,imbecil,do,porteiro]).  
corpus(31,[este,idiota,deste,velho]).  
corpus(32,[menina,rica]).  
corpus(33,[as,meninas,mais,belas]).  
corpus(34,[as,meninas,as,mais,belas]).  
corpus(35,[a,outra,irmã,minha]).  
corpus(36,[a,safada,da,minha,filha]).

%corpus 2

/\*

Expressões nominais AGRAMATICAS

\*/

agram(1,[todos,uns,alunos]).

agram(2,[todos,todos,alunos]).

agram(3,[os,todos,alunos]).

agram(4,[os,este,gerativistas]).

agram(5,[aquele,este,porteiro]).

agram(6,[nenhum,este,meninas]).

agram(7,[alunos,os]).

agram(8,[menina,a]).

agram(9,[homem,ambos]).

agram(10,[os,alunos,os,alunos]).

agram(11,[as,gerativistas,os,funcionalistas]).

agram(12,[o,quase,belas]).

agram(13,[a,filho,meu]).

agram(14,[oito,de,o,filho]).

agram(15,[todos,todos,cada,alunos]).

agram(16,[todos,cada,funcionalistas]).

%SINTAXE%

np(np(X)) --> n(X).

np(np(X,Y)) --> n(X),ap(Y).

np(np(X,Y)) --> n(X),dp(Y).

np(np(X,Y)) --> ap(X),np(Y).

dp(dp(X)) --> d(X).

dp(dp(X,Y)) --> d(X),np(Y).

dp(dp(d(e),Y)) --> np(Y).

dp(dp(X,Y)) --> d(X),dp(Y).

dp(dp(X,Y)) --> d(X),possp(Y).

dp(dp(d(e),Y)) --> possp(Y).

dp(dp(X,Y)) --> np(X),d(Y).

dp(dp(X,Y)) --> d(X),pp(Y).

dp(dp(X,Y)) --> d(X),qp(Y).

dp(dp(d(e),Y)) --> qp(Y).

dp(dp(X,Y)) --> d(X),ap(Y).

qp(qp(X,Y)) --> q(X),pp(Y).

pp(pp(X,Y)) --> p(X),np(Y).

pp(pp(X,Y)) --> p(X),possp(Y).

pp(pp(X,Y)) --> p(X),dp(Y).

possp(possp(X,Y)) --> poss(X),np(Y).

possp(possp(X,Y)) --> np(X),poss(Y).

advp(advp(X)) --> adv(X).

ap(ap(X)) --> a(X).

ap(ap(X,Y)) --> a(X),pp(Y).

ap(ap(X,Y)) --> advp(X),a(Y).

%LÉXICO%

d(d(o)) --> [o].

d(d(os)) --> [os].

d(d(a)) --> [a].

d(d(as)) --> [as].

d(d(uns)) --> [uns].

d(d(este)) --> [este].

d(d(esta)) --> [esta].

d(d(aquele)) --> [aquele].

d(d(nenhum)) --> [nenhum].

d(d(nenhuma)) --> [nenhuma].

d(d(alguma)) --> [alguma].

d(d(alguns)) --> [alguns].

d(d(todos)) --> [todos].

d(d(ambos)) --> [ambos].

d(d(cada)) --> [cada].

d(d(eu)) --> [eu].

d(d(tu)) --> [tu].

d(d(você)) --> [você].

d(d(ele)) --> [ele].

d(d(ela)) --> [ela].

d(d(nós)) --> [nós].

d(d(vocês)) --> [vocês].

d(d(eles)) --> [eles].

d(d(elas)) --> [elas].

%d(d(adv,d)) --> [adv,d].

d(d(X,Y))--> adv(X),d(Y).

poss(poss(meu)) --> [meu].

poss(poss(meus)) --> [meus].

poss(poss(minhas)) --> [minhas].

poss(poss(minha)) --> [minha].

poss(poss(seu)) --> [seu].

q(q(oito)) --> [oito].

a(a(rica)) -->[rica].

a(a(safada)) -->[safada].

a(a(idiota)) -->[idiota].

a(a(belas)) --> [belas].

a(a(imbecil)) -->[imbecil].

a(a(brasileiros)) -->[brasileiros].

a(a(outra)) -->[outra].

adv(adv(quase)) --> [quase].

adv(adv(mais)) --> [mais].

p(p(de)) --> [de].

p(p(do)) --> [do].

p(p(dos)) --> [dos].

p(p(das)) --> [das].

p(p(da)) --> [da].

p(p(este)) --> [este].

n(n(homem)) --> [homem].

n(n(mulher)) --> [mulher].

n(n(filho)) --> [filho].

n(n(filhas)) --> [filhas].

n(n(filha)) --> [filha].

n(n(menina)) --> [menina].

n(n(meninas)) --> [meninas].

n(n(irmã)) --> [irmã].

n(n(porteiro)) --> [porteiro].

n(n(lingüista)) --> [lingüista].

n(n(alunos)) --> [alunos].

n(n(lingüistas)) --> [lingüistas].

n(n(funcionalistas)) --> [funcionalistas].

n(n(gerativistas)) --> [gerativistas].

n(n(amigos)) --> [amigos].

n(n(médicos)) --> [médicos].



<p>n(n(livro)) --&gt; [livro]. n(n(livros)) --&gt; [livros]. n(n(teoria)) --&gt; [teoria]. n(n(velho)) --&gt; [velho].</p>
--

Representamos uma estrutura em Prolog por meio de predicados e argumentos.

De acordo com Pagani (2004, p.12): “o efeito de se incluir argumentos nos predicados de uma gramática em DCG é que, quando ela é lida pelo interpretador Prolog, esses argumentos são incluídos antes dos dois argumentos em que a diferença de lista é processada”.

Assim, para podermos solicitar do programa a análise gramatical de uma expressão nominal, usaremos umas das seguintes diretivas para a expressão *todos nós lingüistas*:

(120)  
dp(A,[todos,nós,lingüistas],[]), ou  
corpus(8,E),dp(A,E,[]).

Ou seja, estamos pedindo ao programa que nos apresente a estrutura (A) do DP *todos nós lingüistas*. A resposta para nossa pergunta será a seguinte:

(121) dp(d(todos), dp(d(nós), np(n(lingüistas))))

Para a representação da expressão *teoria esta*, por exemplo, basta substituímos no lugar de *todos os lingüistas*, a nova expressão:

(122) dp(A[teoria,esta],[]).

Teremos a resposta:

(123)  
A = dp(d(e), np(n(teoria), dp(d(esta))))

Como foi visto já temos um *corpus* descrito no programa para facilitar nossas consultas.

Uma novidade que apresentaremos aqui, que não foi proposta, por exemplo, por Vieira e Lima (2003), Pagani (2004), e Othero e Menuzzi (2005) é um programa o qual desenha as estruturas em forma de árvore que foi desenvolvido em 1985 por Gustaf Neumann e este programa recebeu uma versão para a língua portuguesa por Alencar (2006). Esse programa é

capaz de representar as estruturas em língua portuguesa por meio de representações arbóreas, para isso, basta compilar este programa em Prolog, no mesmo momento em que estivermos trabalhando com o analisador.

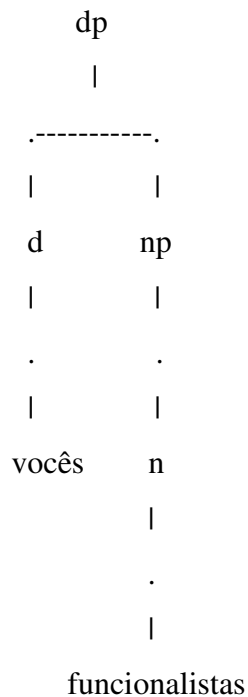
Observemos, a seguir, uma demonstração de como isso pode ser feito. Ao compilar o programa *desenha*, lançamos, não mais uma, mas duas diretivas:

(124)

corpus(5,DP),dp(E,DP,[]),desenha(E).

O programa nos fornecerá o seguinte:

(125)



DP = [vocês, functionalistas]

E = dp(d(vocês), np(n(functionalistas)))

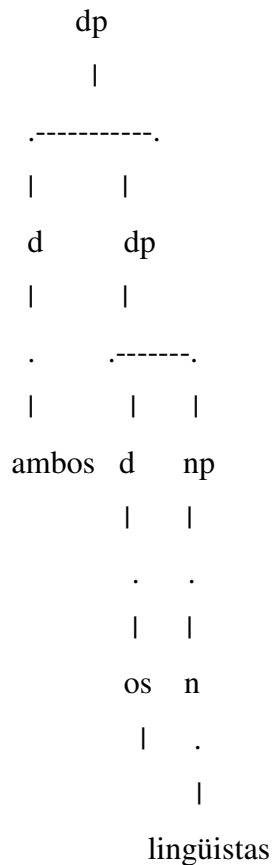
Lancemos outra diretiva:

(126)

corpus(6,DP),dp(E,DP,[]),desenha(E).

A resposta:

(127)



DP = [ambos, os, lingüistas]

E = dp(d(ambos), dp(d(os), np(n(lingüistas))))

Ou seja, o que solicitamos ao programa foi o seguinte: “apresente a estrutura constituinte do item 5 do *corpus*, *vocês funcionalistas* e desenhe em forma de árvore esta estrutura. Foi exatamente isso que o programa fez.

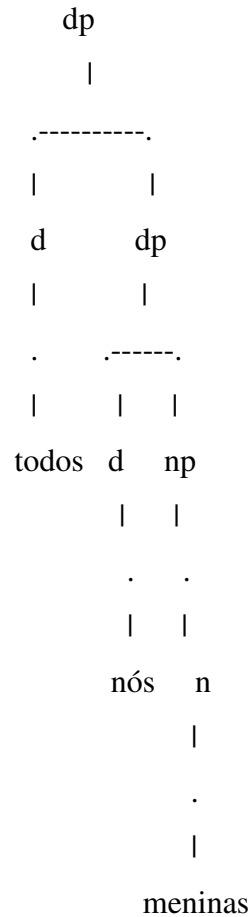
Que tal, se tentássemos, agora, solicitar do programa que nos apresentasse a análise da seguinte estrutura:

(128)

dp(E,[todos,nós,meninas],[[]], desenha(E).

Resposta:

(129)



DP = [todos,nós,meninas].

E = dp(d(todos), dp(d(nós), np(n(meninas))))

O programa reconheceu essa estrutura como gramatical para o PB e, ainda, a analisou. Qual o problema que houve?

O problema que decorreu foi porque desse modo como esse analisador se encontra, por não termos implementado nas regras e no léxico os traços gramaticais, ele está hipergerando estruturas agramaticais, como *este menina, todas nós amigos* etc.

Para entendermos como resolver tal problema, passemos para a próxima seção, através da qual apresentamos como solucionar, em partes, este problema de hipergeração.

### 4.3. Analisador gramatical do português no formato DCG com traços

```
%ANALISADOR GRAMATICAL com traços DE EXPRESSÕES NOMINAIS EM LP
%AUTORA: KARINE ALVES DAVID
%DATA: 29/01/2007.
```

```
%corpus
```

```
/*
```

```
Expressões nominais em Língua portuguesa
```

```
*/
```

```
corpus(1,[todos,lingüistas]).
corpus(2,[alguns,alunos]).
corpus(3,[esta,teoria]).
corpus(4,[nós,gerativistas]).
corpus(5,[vocês,funcionalistas]).
corpus(6,[ambos,os,lingüistas]).
corpus(7,[todos,os,lingüistas]).
corpus(8,[todos,nós,lingüistas]).
corpus(9,[todos,vocês,funcionalistas]).
corpus(10,[o,meu,filho]).
corpus(11,[este,meu,livro]).
corpus(12,[eu,seu,filho]).
corpus(13,[quase,todos,nós,alunos]).
corpus(14,[quase,todos,vocês,médicos]).
corpus(15,[todos,nós,lingüistas,brasileiros]).
corpus(16,[quase,todos,nós,lingüistas,brasileiros]).
corpus(17,[lingüista,aquele]).
corpus(18,[teoria,esta]).
corpus(19,[alguns,dos,alunos]).
corpus(20,[nenhuma,das,filhas]).
corpus(21,[nenhuma,das,minhas,filhas]).
corpus(22,[alguns,dos,meus,amigos]).
corpus(23,[oito,de,nós]).
corpus(24,[cada,um,dos,livros]).
corpus(25,[um,filho,meu]).
corpus(26,[alguns,amigos,meus]).
corpus(27,[filho,meu]).
corpus(28,[homem,nenhum]).
```

corpus(29,[mulher,alguma]).  
 corpus(30,[aquele,imbecil,do,porteiro]).  
 corpus(31,[este,idiota,deste,velho]).  
 corpus(32,[menina,rica]).  
 corpus(33,[as,meninas,mais,belas]).  
 corpus(34,[as,meninas,as,mais,belas]).  
 corpus(35,[a,outra,irmã,minha]).  
 corpus(36,[a,safada,da,minha,filha]).

%SINTAXE%

np(np(X),[gend(G),num(N)]) --> n(X,[gend(G),num(N)]).  
 np(np(X,Y),[gend(G),num(N)]) -->  
 n(X,[gend(G),num(N)],ap(Y,[gend(G),num(N)]).  
 np(np(X,Y),T) --> n(X,T),ap(Y,T).

np(np(X,Y)) --> n(X),dp(Y).  
 np(np(X,Y)) --> ap(X),np(Y).

dp(dp(X),[gend(G),num(N),pers(P),def(D)]) -->  
 d(X,[gend(G),num(N),pers(P),def(D),pron('+'),quant(\_)]).

dp(dp(X,Y),[gend(G),num(N),pers(P),def(D)]) -->  
 d(X,[gend(G),num(N),pers(P),def(D),pron('-'),quant(\_)]),  
 np(Y,[gend(G),num(N)]).

dp(dp(d(e),Y),T) --> np(Y,T).

dp(dp(X,Y),[gend(G),num(N),pers(P),def(D)]) -->  
 d(X,[gend(G),num(N),pers(\_),def(\_),pron(\_),quant(univ)]),  
 dp(Y,[gend(G),num(N),pers(P),def(D)]).

dp(dp(X,Y)) --> d(X),possp(Y).  
 dp(dp(d(e),Y)) --> possp(Y).  
 dp(dp(X,Y)) --> np(X),d(Y).  
 dp(dp(X,Y)) --> d(X),pp(Y).  
 dp(dp(X,Y)) --> d(X),qp(Y).

dp(dp(d(e),Y)) --> qp(Y).  
dp(dp(X,Y)) --> d(X),ap(Y).

qp(qp(X,Y)) --> q(X),pp(Y).

pp(pp(X,Y)) --> p(X),np(Y).  
pp(pp(X,Y)) --> p(X),possp(Y).  
pp(pp(X,Y)) --> p(X),dp(Y).

possp(possp(X,Y)) --> poss(X),np(Y).  
possp(possp(X,Y)) --> np(X),poss(Y).

advp(advp(X)) --> adv(X).

ap(ap(X,[gend(G),num(N)]) --> a(X,[gend(G),num(N)]).  
ap(ap(X,Y)) --> a(X),pp(Y).  
ap(ap(X,Y)) --> advp(X),a(Y).

%LÉXICO%

d(d(o),[gend(masc),num(sing),pers(3),def('+'),pron('-'),quant(null)]) -->[o].  
d(d(os),[gend(masc),num(pl),pers(3),def('+'),pron('-'),quant(null)]) --> [os].  
d(d(a),[gend(fem),num(sing),pers(3),def('+'),pron('-'),quant(null)]) --> [a].  
d(d(as),[gend(fem),num(pl),pers(3),def('+'),pron('-'),quant(null)]) --> [as].  
d(d(um),[gend(masc),num(sing),pers(3),def('-'),pron('-'),quant(null)]) --> [um].  
d(d(uns),[gend(masc),num(pl),pers(3),def('-'),pron('-'),quant(null)]) --> [uns].  
d(d(este),[gend(masc),num(sing),pers(3),def('+'),pron(\_),quant(null)]) --> [este].  
d(d(esta),[gend(fem),num(sing),pers(3),def('+'),pron(\_),quant(null)]) --> [esta].  
d(d(aquele),[gend(masc),num(sing),pers(3),def('+'),pron(\_),quant(null)]) -->  
[aquele].  
d(d(nenhum),[gend(masc),num(sing),pers(3),def('-'),pron(\_),quant(null)]) -->  
[nenhum].  
d(d(nenhuma),[gend(fem),num(sing),pers(3),def('-'),pron(\_),quant(null)]) -->  
[nenhuma].  
d(d(alguma),[gend(fem),num(sing),pers(3),def('-'),pron(\_),quant(null)]) -->  
[alguma].

d(d(alguns),[gend(masc),num(sing),pers(3),def('-'),pron(\_),quant(null)]) --> [alguns].  
 d(d(todos),[gend(masc),num(pl),pers(3),def('-'),pron(\_),quant(univ)]) --> [todos].  
 d(d(ambos),[gend(masc),num(pl),pers(3),def('-'),pron(\_),quant(univ)]) --> [ambos].  
 d(d(cada),[gend(\_),num(\_),pers(3),def('-'),pron(\_),quant(null)]) --> [cada].  
 d(d(eu),[gend(\_),num(sing),pers(2),def('+'),pron(\_),quant(null)]) --> [eu].  
 d(d(tu),[gend(\_),num(sing),pers(2),def('+'),pron('+'),quant(null)]) --> [tu].  
 d(d(você),[gend(\_),num(sing),pers(2),def('+'),pron(\_),quant(null)]) --> [você].  
 d(d(ele),[gend(masc),num(sing),pers(3),def('+'),pron('+'),quant(null)]) --> [ele].  
 d(d(ela),[gend(fem),num(sing),pers(3),def('+'),pron('+'),quant(null)]) --> [ela].  
 d(d(nós),[gend(\_),num(pl),pers(1),def('+'),pron(\_),quant(null)]) --> [nós].  
 d(d(vocês),[gend(\_),num(pl),pers(3),def('+'),pron(\_),quant(null)]) --> [vocês].  
 d(d(eles),[gend(masc),num(pl),pers(3),def('+'),pron('+'),quant(null)]) --> [eles].  
 d(d(elas),[gend(fem),num(pl),pers(3),def('+'),pron('+'),quant(null)]) --> [elas].  
 %d(d(adv,d)) --> [adv,d].  
 d(d(X,Y))--> adv(X),d(Y).

poss(poss(meu)) --> [meu].  
 poss(poss(meus)) --> [meus].  
 poss(poss(minhas)) --> [minhas].  
 poss(poss(minha)) --> [minha].  
 poss(poss(seu)) --> [seu].

q(q(oito)) --> [oito].

a(a(rica),[gend(fem),num(sing)]) -->[rica].  
 a(a(safada),[gend(fem),num(sing)]) -->[safada].  
 a(a(idiota),[gend(\_),num(sing)]) -->[idiota].  
 a(a(belas),[gend(fem),num(pl)]) --> [belas].  
 a(a(imbecil),[gend(\_),num(sing)]) -->[imbecil].  
 a(a(brasileiros),[gend(masc),num(pl)]) -->[brasileiros].  
 a(a(outra),[gend(fem),num(sing)]) -->[outra].

adv(adv(quase)) --> [quase].  
 adv(adv(mais)) --> [mais].



p(p(de)) --> [de].  
p(p(do)) --> [do].  
p(p(dos)) --> [dos].  
p(p(das)) --> [das].  
p(p(da)) --> [da].  
p(p(este)) --> [este].

n(n(homem), [gend(masc), num(sing)]) --> [homem].  
n(n(mulher), [gend(fem), num(sing)]) --> [mulher].  
n(n(filho), [gend(masc), num(sing)]) --> [filho].  
n(n(filhas), [gend(fem), num(pl)]) --> [filhas].  
n(n(filha), [gend(fem), num(sing)]) --> [filha].  
n(n(menina), [gend(fem), num(sing)]) --> [menina].  
n(n(meninas), [gend(fem), num(pl)]) --> [meninas].  
n(n(irmã), [gend(fem), num(sing)]) --> [irmã].  
n(n(porteiro), [gend(masc), num(sing)]) --> [porteiro].  
n(n(lingüista), [gend(\_), num(sing)]) --> [lingüista].  
n(n(alunos), [gend(masc), num(pl)]) --> [alunos].  
n(n(lingüistas), [gend(\_), num(pl)]) --> [lingüistas].  
n(n(funcionalistas), [gend(\_), num(pl)]) --> [funcionalistas].  
n(n(gerativistas), [gend(\_), num(pl)]) --> [gerativistas].  
n(n(amigos), [gend(masc), num(pl)]) --> [amigos].  
n(n(médicos), [gend(masc), num(pl)]) --> [médicos].  
n(n(livro), [gend(masc), num(sing)]) --> [livro].  
n(n(livros), [gend(masc), num(pl)]) --> [livros].  
n(n(teoria), [gend(fem), num(sing)]) --> [teoria].  
n(n(velho), [gend(masc), num(sing)]) --> [velho].

Conforme podemos constatar, através do analisador sem traços representado na seção 4.2, apesar de fazer análises adequadas, ele também permite algumas inadequações, como vimos em (129), a respeito do exemplo *todos nós meninas*, estrutura a qual foi gerada, reconhecida e analisada por aquela DCG.

O que aconteceu foi uma espécie de descontrole da gramática em relação aos traços morfossintáticos os quais não foram mencionados naquele analisador. Para isso, a solução que encontramos é acrescentar um argumento a mais tanto para a sintaxe, quanto para o léxico, para que haja a unificação entre os traços dos itens que formam cada expressão nominal.

Primeiramente, fizemos um levantamento dos traços que seriam acrescentados, que foram os referentes a *gênero, número, pessoa, definitude, posição e quantificação*. Veremos, mais adiante, que somente esses traços não deram conta para que nossa gramática se tornasse perfeita, porém, por uma questão de tempo, não pudemos, infelizmente, acrescentar outros traços como *pcase* (referente ao uso da preposição), localização do núcleo, etc.

Em nosso programa, apesar de não termos estabelecido todos os traços pertinentes para a perfeição desta gramática já anulamos a possibilidade da análise de uma expressão como a de (129), a qual representaremos abaixo em (130), a diretiva e a resposta que foi dada pelo programa, depois de os traços serem implementados:

(130) ?- dp(E,T,[todos,nós,meninas],[]).

Resposta: *No*

O que fizemos em (130) foi solicitar ao programa que reconhecesse e analisasse a expressão nominal *todos nós meninas*. O que se pode observar é que por já termos implementado os traços referentes a gênero e número, o programa anula a possibilidade de analisar estruturas agramaticais nesse nível.

Para os substantivos e adjetivos, apenas implementamos traços de gênero e número, pois são estes os traços que achamos que pudessem ser inerentes a estas categorias, *a priori*.

Para

os determinantes, incluímos *gênero, número, posição, definitude e quantificação*. Apresentaremos, a seguir, como foi feita essa implementação passo a passo, por meio de exemplos.

Para todos os determinantes, além dos traços de *gênero e número*, começamos pela inclusão dos traços de *pessoa*. Pois conforme Radford (1997), como vimos na seção 3.1.2., o pronome, exatamente como o artigo, determina o traço de *pessoa* da expressão nominal, pois o traço de *pessoa* não é inerente aos substantivos. Ele é determinado pelo pronome e pelo artigo. Isso mostra que o pronome funciona, do mesmo modo que o artigo, como núcleo do sintagma. Observemos a diretiva abaixo:

(131) ?- dp(E,T,[todos,nós,lingüistas],[]).

E = dp(d(todos), dp(d(nós), np(n(lingüistas))))

T = [gend(masc), num(pl), pers(1), def(+)]

Nessa análise, feita pelo programa, podemos constatar que os traços referentes a esta expressão nominal são: masculino, plural, 1ª pessoa e definido. Se observarmos bem os traços de *todos*, iremos constatar que seu traço para pessoa é representado por *pers* ( $\_$ ), pois como já vimos, temos um traço representado por uma variável anônima ( $\_$ ), uma espécie de curinga, em Prolog, que vai unificar tanto com *pers(1)* e *pers(3)*. Então, o traço que prevalece é o do *nós* que é o de 1ª pessoa.

Analisemos como seria uma diretiva com o pronome pessoal *vocês* que possui traço de 3ª pessoa:

(132) dp(E,T,[todos,vocês,lingüistas],[,]).

E = dp(d(todos), dp(d(vocês), np(n(lingüistas))))

T = [gend(masc), num(pl), pers(3), def(+)]

Exatamente, teremos como traço de pessoa, o de 3ª pessoa, já que *vocês* traz esse traço em sua estrutura.

Vamos ver, agora, o traço referente à definitude. Analisemos novamente o determinante *todos*, que não co-ocorre, por exemplo, com determinantes indefinidos. Então, para ele, implementamos o traço *def* ('+') para que não unifique, por exemplo, com determinantes do tipo: *alguns*, *uns*,  *nenhuns etc.*

Para testar o programa, lançamos a seguinte diretiva:

(133) ?- dp(E,T,[todos,uns,funcionalistas],[,]).

Resposta: No

Tivemos uma resposta negativa do programa, já que o traço de definitude de *todos* não unifica com o de *uns*. Testamos, mais uma vez, o programa:

(134) ?- dp(E,T,[todos,alguns,funcionalistas],[,]).

Resposta: No

Novamente: resposta negativa. Pois o *alguns* possui um traço *def* ('-') o qual não unifica com o *def* ('+') de *todos*.

Observemos a diretiva a seguir:

(135) ?- dp(E,T,[todos,os,lingüistas],[,]).

Resposta:

E = dp(d(todos), dp(d(os), np(n(lingüistas))))

T = [gend(masc), num(pl), pers(3), def(+)]

Agora sim, teremos uma resposta positiva do programa, pois os traços tanto de *todos* como de *os* se unificam, pois são *def*('+').

Outro traço que incluímos, bastante interessante, foi o referente à posição dos determinantes no DP. Para isso, utilizamos as representações *pron* ('+'), para quando o determinante ocorrer pronominalmente e *pron* ('-') para quando este ocorrer apenas pré-nominalmente.

Conforme nossa hipótese de que os pronomes pessoais são núcleos de um DP que tomam NP como complemento, já que, nessa pesquisa, corroboramos apenas com os pronomes *eu, você, nós e vocês*, apenas para eles atribuímos o traço *pron* (\_), pois eles tanto podem aparecer tanto pré-nominalmente como também pronominalmente. Já para os outros pronomes pessoais *tu, ele, ela, eles, elas*, implementamos o traço *pron* ('+'), já que estes, segundo nossos levantamentos, apenas aparecem pronominalmente.

Vamos conferir com o nosso programa:

(136) ?- dp(E,T,[vocês],[]).

E = dp(d(vocês))

T = [gend(\_G382), num(pl), pers(3), def(+)]

Perguntamos ao programa se *vocês* é um DP e solicitamos sua estrutura e traços. Ele nos respondeu que sim e nos deu a análise deste DP. Agora, vamos perguntar ao programa se *vocês* pode aparecer pronominalmente:

(137) ?- dp(E,T,[vocês,gerativistas],[]).

E = dp(d(vocês), np(n(gerativistas)))

T = [gend(\_G467), num(pl), pers(3), def(+)]

Sim, conforme o programa, *vocês*, também, além de possuir o traço *pron* ('+'), ele também possui o traço *pron* ('-'), pois toma o NP *gerativistas* enquanto seu complemento.

Testemos, agora, o programa com um pronome o qual, segundo nossos resultados no corpus NILC, é *pron* ('+'), mas não pré-nominal:

(138) - dp(E,T,[elas,lingüistas],[]).

No

O programa respondeu que não, que a expressão *elas lingüistas* não é gramatical, pois o *elas* não toma NP como complemento. Agora:

(139) - dp(E,T,[elas],[]).

E = dp(d(elas))

T = [gend(fem), num(pl), pers(3), def(+)]

Nesse caso, não há problemas, pois o *elas*, é um DP, contudo, apenas aparece pronominalmente, mas, não, pré-nominalmente.

Por último, implementamos um traço *quant(univ)* e *quant(null)*. Este traço serve para diferenciar os determinantes *todos* e *ambos*, que são quantificadores universais, dos outros que não são, pois são nulos enquanto quantificação.

Na verdade, nosso objetivo inicial nesta pesquisa era construirmos um analisador que trouxesse todos os traços que fizessem com que analisasse todas as estruturas intuídas por nós. Fizemos exatamente isso, ao construirmos o analisador sem traços. Aquele analisador gera, reconhece e analisa as nossas 36 expressões nominais. Porém, por motivo de tempo, o que não conseguimos, foi implementar no analisador com traços, todos os traços precisos na sintaxe e léxico para não haver hipergeração de estruturas agramaticais em língua portuguesa.

De acordo como colocamos anteriormente, o analisador expresso em 4.4. possui alguns problemas de hipergeração em relação a expressões nominais. Pois, apesar de termos implementados nele os traços de gênero, número, pessoa, definitude, por questão de tempo, não implementamos outros traços que pudessem a vir evitar hipergeração de algumas estruturas agramaticais.

Isso, contudo, não desmerece nosso trabalho, pois, ao final, atingimos nosso objetivo que foi o de construir um analisador que gerasse, reconhecesse e analisasse expressões nominais em língua portuguesa.

Para concluir esta seção, apresentamos, na seção 4.5. um fragmento do fragmento de gramática que conseguimos construir até então:

#### 4.4. Fragmento de analisador gramatical do português no formato DCG com traços.

```
% FRAGMENTO DE ANALISADOR GRAMATICAL com traços DE
EXPRESSÕES NOMINAIS EM LP
%AUTORA: KARINE ALVES DAVID
%DATA: 29/01/2007.

%corpus

/*
Expressões nominais em Língua portuguesa
*/
corpus(1,[todos,lingüistas]).
corpus(2,[alguns,alunos]).
corpus(3,[esta,teoria]).
corpus(4,[nós,gerativistas]).
corpus(5,[vocês,funcionalistas]).
corpus(6,[ambos,os,lingüistas]).
corpus(7,[todos,os,lingüistas]).
corpus(8,[todos,nós,lingüistas]).
corpus(9,[todos,vocês,funcionalistas]).
corpus(10,[o,meu,filho]).
corpus(11,[este,meu,livro]).
corpus(12,[eu,seu,filho]).
corpus(13,[quase,todos,nós,alunos]).
corpus(14,[quase,todos,vocês,médicos]).
corpus(15,[todos,nós,lingüistas,brasileiros]).
corpus(16,[quase,todos,nós,lingüistas,brasileiros]).
corpus(17,[lingüista,aquele]).
corpus(18,[teoria,esta]).
corpus(19,[alguns,dos,alunos]).
corpus(20,[nenhuma,das,filhas]).
corpus(21,[nenhuma,das,minhas,filhas]).
corpus(22,[alguns,dos,meus,amigos]).
corpus(23,[oito,de,nós]). %problema
corpus(24,[cada,um,dos,livros]).
corpus(25,[um,filho,meu]).
corpus(26,[alguns,amigos,meus]).
```

corpus(27,[filho,meu]). %problema  
 corpus(28,[homem,nenhum]).  
 corpus(29,[mulher,alguma]).  
 corpus(30,[aquele,imbecil,do,porteiro]).  
 corpus(31,[este,idiota,deste,velho]).  
 corpus(32,[menina,rica]).  
 corpus(33,[as,meninas,mais,belas]).  
 corpus(34,[as,meninas,as,mais,belas]).  
 corpus(35,[a,outra,irmã,minha]).  
 corpus(36,[a,safada,da,minha,filha]).

%SINTAXE%

np(np(X),[gend(G),num(N)]) --> n(X,[gend(G),num(N)]).  
 np(np(X,Y),[gend(G),num(N)]) -->  
 n(X,[gend(G),num(N)]),ap(Y,[gend(G),num(N)]).  
 np(np(X,Y),T) --> n(X,T),ap(Y,T).

np(np(X,Y)) --> n(X),dp(Y).  
 np(np(X,Y)) --> ap(X),np(Y).

dp(dp(X),[gend(G),num(N),pers(P),def(D)]) -->  
 d(X,[gend(G),num(N),pers(P),def(D),pron('+'),quant(\_)]).

dp(dp(X,Y),[gend(G),num(N),pers(P),def(D)]) -->  
 d(X,[gend(G),num(N),pers(P),def(D),pron('-'),quant(\_)]),  
 np(Y,[gend(G),num(N)]).

dp(dp(d(e),Y),T) --> np(Y,T). %aqui tem um D vazio CFG: DP->D,NP.

dp(dp(X,Y),[gend(G),num(N),pers(P),def(D)]) -->  
 d(X,[gend(G),num(N),pers(\_),def(\_),pron(\_),quant(univ)]),  
 dp(Y,[gend(G),num(N),pers(P),def(D)]).

ap(ap(X),[gend(G),num(N)]) --> a(X,[gend(G),num(N)]).

%LÉXICO%

d(d(o),[gend(masc),num(sing),pers(3),def('+'),pron('-'),quant(null)]) -->[o].  
d(d(os),[gend(masc),num(pl),pers(3),def('+'),pron('-'),quant(null)]) --> [os].  
d(d(a),[gend(fem),num(sing),pers(3),def('+'),pron('-'),quant(null)]) --> [a].  
d(d(as),[gend(fem),num(pl),pers(3),def('+'),pron('-'),quant(null)]) --> [as].  
d(d(um),[gend(masc),num(sing),pers(3),def('-'),pron('-'),quant(null)]) --> [um].  
d(d(este),[gend(masc),num(sing),pers(3),def('+'),pron(\_),quant(null)]) --> [este].  
d(d(esta),[gend(fem),num(sing),pers(3),def('+'),pron(\_),quant(null)]) --> [esta].  
d(d(aquele),[gend(masc),num(sing),pers(3),def('+'),pron(\_),quant(null)]) -->  
[aquele].  
d(d(nenhum),[gend(masc),num(sing),pers(3),def('-'),pron(\_),quant(null)]) -->  
[nenhum].  
d(d(nenhuma),[gend(fem),num(sing),pers(3),def('-'),pron(\_),quant(null)]) -->  
[nenhuma].  
d(d(alguma),[gend(fem),num(sing),pers(3),def('-'),pron(\_),quant(null)]) -->  
[alguma].  
d(d(alguns),[gend(masc),num(sing),pers(3),def('-'),pron(\_),quant(null)]) -->  
[alguns].  
d(d(todos),[gend(masc),num(pl),pers(3),def('-'),pron(\_),quant(univ)]) --> [todos].  
d(d(ambos),[gend(masc),num(pl),pers(3),def('-'),pron(\_),quant(univ)]) --> [ambos].  
d(d(cada),[gend(\_),num(\_),pers(3),def('-'),pron(\_),quant(null)]) --> [cada].  
d(d(eu),[gend(\_),num(sing),pers(2),def('+'),pron(\_),quant(null)]) --> [eu].  
d(d(tu),[gend(\_),num(sing),pers(2),def('+'),pron('+'),quant(null)]) --> [tu].  
d(d(você),[gend(\_),num(sing),pers(2),def('+'),pron(\_),quant(null)]) --> [você].  
d(d(ele),[gend(masc),num(sing),pers(3),def('+'),pron('+'),quant(null)]) --> [ele].  
d(d(ela),[gend(fem),num(sing),pers(3),def('+'),pron('+'),quant(null)]) --> [ela].  
d(d(nós),[gend(\_),num(pl),pers(1),def('+'),pron(\_),quant(null)]) --> [nós].  
d(d(vocês),[gend(\_),num(pl),pers(3),def('+'),pron(\_),quant(null)]) --> [vocês].  
d(d(eles),[gend(masc),num(pl),pers(3),def('+'),pron('+'),quant(null)]) --> [eles].  
d(d(elas),[gend(fem),num(pl),pers(3),def('+'),pron('+'),quant(null)]) --> [elas].  
%d(d(adv,d)) --> [adv,d]. % isto fica assim? tá dando problema  
d(d(X,Y))--> adv(X),d(Y).  
  
a(a(rica),[gend(fem),num(sing)]) -->[rica].  
a(a(safada),[gend(fem),num(sing)]) -->[safada].  
a(a(idiota),[gend(\_),num(sing)]) -->[idiota].  
a(a(belas),[gend(fem),num(pl)]) --> [belas].



a(a(imbecil),[gend(\_),num(sing)]) -->[imbecil].  
 a(a(brasileiros),[gend(masc),num(pl)]) -->[brasileiros].  
 a(a(outra),[gend(fem),num(sing)]) -->[outra].

n(n(homem), [gend(masc),num(sing)]) --> [homem].  
 n(n(mulher),[gend(fem),num(sing)]) --> [mulher].  
 n(n(filho),[gend(masc),num(sing)]) --> [filho].  
 n(n(filhas),[gend(fem),num(pl)]) --> [filhas].  
 n(n(filha),[gend(fem),num(sing)]) --> [filha].  
 n(n(menina),[gend(fem),num(sing)]) --> [menina].  
 n(n(meninas),[gend(fem),num(pl)]) --> [meninas].  
 n(n(irmã),[gend(fem),num(sing)]) --> [irmã].  
 n(n(porteiro),[gend(masc),num(sing)]) --> [porteiro].  
 n(n(lingüista),[gend(\_),num(sing)]) --> [lingüista].  
 n(n(alunos),[gend(masc),num(pl)]) --> [alunos].  
 n(n(lingüistas),[gend(\_),num(pl)]) --> [lingüistas].  
 n(n(funcionalistas),[gend(\_),num(pl)]) --> [funcionalistas].  
 n(n(gerativistas),[gend(\_),num(pl)]) --> [gerativistas].  
 n(n(amigos),[gend(masc),num(pl)]) --> [amigos].  
 n(n('Deus'),[gend(masc),num(sing)]) --> ['Deus'].  
 n(n(médicos),[gend(masc),num(pl)]) --> [médicos].  
 n(n(livro),[gend(masc),num(sing)]) --> [livro].  
 n(n(livros),[gend(masc),num(pl)]) --> [livros].  
 n(n(teoria),[gend(fem),num(sing)]) --> [teoria].  
 n(n(velho),[gend(masc),num(sing)]) --> [velho].

O que fizemos aqui foi eliminar todas as regras e itens lexicais os quais não conseguimos implementar os traços morfossintáticos.

# CONCLUSÃO

---

Neste trabalho, discutimos a estrutura interna de um fragmento de expressões nominais do português do Brasil, seguindo a hipótese DP postulada, principalmente, por Fukui (1986) e Abney (1987), com base no modelo de descrição sintática mais recente da teoria X-barra.

Ao descrevermos tais expressões, tomamos como foco principal para nosso trabalho aquelas que trazem em sua estrutura interna, pronomes pessoais assumindo o papel de determinantes e tomam como complementos sintagmas nominais (NPs).

Porém, nossa pesquisa não se limitou apenas a uma descrição lingüística. Como visto no capítulo 4, a partir do fragmento de expressões nominais levantadas por nós, construímos um analisador gramatical capaz de gerar, reconhecer e analisar essas expressões nominais em língua portuguesa por meio de um mecanismo computacional.

Como considerações finais dessa discussão, retomamos o que foi tratado no decorrer do trabalho e sintetizamos as principais conclusões delineadas, propondo algumas generalizações. Destacamos pontos que julgamos de maior relevância e comentamos algumas implicações dos resultados para o desenvolvimento da ciência lingüística dentro de uma teoria computacional da linguagem natural e engenharia da linguagem, por consistirem áreas ainda pouco desenvolvidas em nosso país.

A motivação do nosso trabalho proveio de estudos anteriores desenvolvidos tanto no âmbito da lingüística gerativa como da lingüística computacional, como forma de contribuir para o crescente número de investigações nessa área.

No capítulo 3 de nossa pesquisa, fizemos uma análise lingüística das expressões nominais em português apresentando para cada grupo de expressões, sua representação arbórea e a regra por meio da qual foi gerada. Analisamos, ainda, nesse capítulo a hipótese de que todos os pronomes pessoais em língua portuguesa se comportam como D que selecionam NP como complemento, do mesmo modo que Radford (1997) fez em pesquisas anteriores para os pronomes pessoais em inglês. A conclusão a qual chegamos, por meio de dados extraídos do *corpus* NILC, é que em língua portuguesa apenas os pronomes pessoais *eu*, *you*, *nós* e *vocês* assumem esse papel. Apesar de nossa hipótese não ter sido corroborada por completo, pretendemos, ainda, em trabalhos futuros, refazer essa análise nos utilizando de outros *corpora*, por acreditarmos que em língua portuguesa, os pronomes *tu*, *ele* e *elas* também podem vir a ocupar essa posição.

Após essa análise lingüística, conseguimos chegar à outra extremidade, que se faz objetivo geral de nossa dissertação: Implementamos as regras gramaticais que foram geradas a partir do fragmento de expressões nominais analisados no capítulo 3 em linguagem PROLOG, linguagem com a qual desenvolvemos nosso analisador sintático.

Muitas vezes, por limitação da própria linguagem de programação, tivemos de alterar algumas regras, contradizendo, não somente, princípios da teoria X-barras, modelo o qual adotamos para a descrição sintática do português, mas também, outros modelos da Lingüística. Estas soluções *ad hoc* foram necessárias por conta da limitação do nosso programa.

Contudo, após construirmos o analisador com traços capaz de gerar, reconhecer e analisar o fragmento de expressões nominais em português, meta final da nossa pesquisa, encontramos problemas no momento da introdução de traços morfossintáticos, pois, por motivo de disponibilidade de tempo, não implementamos, em nosso parser, traços suficientes que evitassem a hipergeração de estruturas nominais agramaticais.

Nosso *parser*, assim mesmo, com essas limitações, mostrou-se eficiente para a geração, reconhecimento e análise das expressões nominais, objetivo central desta dissertação. Assim, propomo-nos a, futuramente, em uma pesquisa, talvez, em nível de doutorado, implementar outros traços a essa minigramática para aperfeiçoá-la a ponto de não mais hipergerar estruturas agramaticais.

Acreditamos que nossa pesquisa contribuiu em muito para o crescimento dos estudos na área de Lingüística Computacional, até porque, até onde sabemos este trabalho é pioneiro no Ceará, quiçá no Nordeste.

## REFERÊNCIAS BIBLIOGRÁFICAS

---

- ABNEY, Steven Paul. **The English Noun Phrase in its Sentential Aspect**, Dissertação de PhD, Massachusetts Institute of Technology (MIT), 1987.
- ALENCAR, Leonel Figueiredo de. Complementos verbais oracionais – uma análise léxico-funcional. In: **Revista Lingua(gem)/Instituto Latino-Americano de Pesquisas Científicas – ILAPEC** 1(1); 173-218. Santa Maria: Pallotti. ISSN 1807-524X, 2004.
- ALENCAR, Leonel Figueiredo de. Der Constructor – ein interaktives Werkzeug für Recherchen in portugiesischen Korpora auf dem WWW. In: PUSCH, Claus D.; RAIBLE, Wolfgang (Orgs.). **Romanistische Korpuslinguistik. Korpora und gesprochene Sprache**. Tübingen:Narr. p.147-154, 2002.
- ALENCAR, Leonel Figueiredo de. Linguagem e Inteligência Artificial. In: MATTES, Marlene Gonçalves (ORG.) **Linguagens várias: expressões múltiplas do ser**. Fortaleza: EUFC, 2006 a.
- ALENCAR, Leonel Figueiredo de. **O Constructor**. Disponível em: <<http://www.geocities.com/briefaustausch/indexc.htm>>, 2000. Acesso em: 24 jan. 2005.
- ALENCAR, Leonel Figueiredo de. **Teoria da Gramática: uma abordagem computacional**. Manuscrito não publicado. Fortaleza: Universidade Federal do Ceará, 2006b.
- ARARIBÓIA. G. **Inteligência Artificial**. Rio de Janeiro: Livros Técnicos e Científicos Editora Ltda, 1988.
- BLACKBURN, Patrick; BOS, Johan; STRIEGNITZ, Kristina. **Learn Prolog now!** Saarbrücken: Universität des Saarlandes, 2001. Disponível em:< <http://www.coli.uni-saarland.de/~kris/learn-Prolog-now/html/Prolog-notes.pdf>> Acesso em: 29.04.2005.
- BRESNAN, Joan. **Lexical-functional syntax**. Oxford: Blackwell, 2001.
- BUTT, Miriam et al. **A grammar writer's cookbook**. Stanford: CSLI Publications, 1999.
- CARNIE, Andrew. **Syntax: A Generative Introduction**. Oxford: Blackwell, 2002.
- CASTRO, Ana Cristina. **A Categoria Vazia no Sintagma Determinante: licenciamento sintático e implicação semântica**. 1997. 86f. Dissertação (Mestrado em Lingüística)-Faculdade de Letras.Universidade Federal do Rio de Janeiro, 1997.
- CERQUEIRA,Vicente. A forma genitiva ‘dele’ e a categoria de concordâncias (AGR) no português brasileiro. In: ROBERTS, I.; KATO, M. A. (orgs). **Português Brasileiro: Uma viagem diacrônica**. Campinas: Editora da Unicamp, 1993. p.129-157.

CHOMSKY, Noam. Remarks on Nominalization, in Jacobs,R.a e P.Rosenbaun (eds.) **Readings in English Transformational Grammar**, Ginn and Company, Walthan, Massachusetts, 1970.

CHOMSKY, Noam. **The Minimalist Program**. Cambridge, MA: MIT Press, 1995.

CHOMSKY, Noam. **O Programa Minimalista**. Tradução de Eduardo Paiva Raposo. Lisboa: Caminho, 1999.

COENE, Martine (ed). **From NP to DP**. The syntax and semantics of noun phrases, Amsterdam: Benjamins, v.1, 2003.

COENE, Martine (ed). **From NP to DP**. The expression of possession in noun phrases. Amsterdam: Benjamins, v.2, 2003.

CUNHA, Celso; CINTRA, Lindley. **Nova gramática do português contemporâneo**. 2. ed. Rio de Janeiro: Nova Fronteira, 1985.

DAVID, Karine Alves. **O pronome pessoal enquanto núcleo de um DP: uma abordagem computacional**. Semana de Letras da UFC, 2006. Disponível em: <http://www.geocities.com/Athens/Crete/1546/gelfoco-htm#leituras>. Acesso em: 16 fev. 2006.

DAVID, Karine Alves. **Uma análise em Prolog do pronome pessoal enquanto núcleo do DP**. XXI Jornada Nacional de Estudos Lingüísticos. ISBN 85-7539-286-7, 2006.

DÉCHAINED, Rose-Marie e WILTSCSKO, Martina. On pro-nouns and other “pronouns”. In: COENE, M. (ed.). **From NP to DP**, Amsterdam: Benjamins, v.1, 2003.

DEVLIN, Keith. **O gene da matemática**. 2. ed. Tradução de Sérgio Moraes Rego. Revisão técnica de Leo Tenenblat. Rio de Janeiro: Record, 2005.

FALK, Yehuda N. **Lexical-Functional Grammar: An Introduction to Parallel Constraint-Based Syntax**. Stanford: CSLI Publications, 2001.

FUKUI, Naoki. **A theory of category projection and its applications**. Tese de doutorado. Massachusetts Institute of Technology (MIT), 1986.

GALVES, Charlotte. O enfraquecimento da concordância no PB. In: ROBERTS, I.; KATO, M. A. (orgs). **Português Brasileiro: Uma viagem diacrônica**. Campinas: Editora da Unicamp, 1993. p.129-157.

GÓMEZ GUINOVART, Xavier. Lingüística computacional. In: RAMALLO, Fernando; REIDOVAL, Gabriel; RODRÍGUEZ, Xoán Paulo (Orgs.). **Manual de ciencias da linguaxe**. Vigo: Xerais, 2000. p. 221-268.

HERDAN. Simona. Floating Quantifiers and the Structure Romanian. In: NOMURA, M., NIINUMA, F.(eds.). **University of Connecticut Working Papers in Linguistics**. Vol.13,p.79-118,2005.

- KLENK, Ursula. **Generative Syntax**. Tübingen: Narr, 2003.
- LOBATO, Lúcia. A concordância Nominal no Português do Brasil à Luz da Teoria de Princípios-e-Parâmetros e da Sociolinguística Variacionista. **D.E.L.T.A**, v.10, n.2, p. 173-212, 1994.
- MAGALHÃES, Telma M.V. A valoração de traços de concordância dentro do DP. **D.E.L.T.A**. v.20, n.1. p.149-170, 2004. Disponível em: <www.scielo.br/delta>. Acesso em 12 out. 2005.
- MARTINHO, Fernando Jorge dos Santos. **A elipse nominal em português e francês**. Dissertação de mestrado. Faculdade de Letras da Universidade do Porto. Porto,1998.
- MATEUS, Helena Mira. *et al.* **Gramática da língua portuguesa**. 2.ed. Lisboa: Caminho, 1989.
- MIOTO, Carlos; FIGUEIREDO, Maria Cristina; LOPES, Ruth. **Manual de Sintaxe**. Florianópolis: Insular, 1999.
- MONTEIRO, José Lemos. **Pronomes Pessoais**. Fortaleza: EUFC, 1994.
- MÜLLER, Ana. **A sintaxe e a Semântica dos Possessivos no Português do Brasil**, 1996, 193 f. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem. Universidade estadual de Campinas, 1996.. UNICAMP, 1996.
- NEGRÃO, Esmeralda; MÜLLER, Ana L. As mudanças no sistema pronominal do Português do Brasil: substituição ou especialização de formas? **D.E.L.T.A**, v. 12, n. 1, p. 211-22, 1996.
- OTHERO, Gabriel de A. **Grammar Play: um parser sintático em Prolog para língua portuguesa**. 2004. Dissertação (Mestrado em Linguística), Pontifícia Universidade Católica do Rio Grande do Sul,Porto Alegre, 2004.
- OTHERO, Gabriel e MENUZZI, Sérgio. **Linguística Computacional**. São Paulo: Parábola, 2005.
- OTHERO, Gabriel. **Teoria X-barra**. São Paulo: Contexto, 2006.
- PAGANI, Luiz Arthur. Analisador Gramatical em Prolog para gramáticas de estrutura sintagmática. In: **Revista Virtual de Estudos em Linguagem –ReVEL**. Ano 2. n.3.[www.revelhp.cjb.net], 2004.
- PAGANI, Luiz Arthur. Gramática categorial através de estrutura de características. **Revista Letras**, p. 358-410. 2005.
- PARDO, Tiago. **Estudo comparativo dos formalismos gramaticais DCG e LFG: relatório técnico RT-DC 001/99**. Departamento de Computação, Universidade Federal de São Carlos. Disponível em: <http://www.nilc.icmc.usp.br/~thiago/RTDC00199-Pardo.pdf>. Acesso em: 14 nov.2005.

PINHEIRO, Gisele Montilha e ALUÍSIO Sandra. **Corpus Nilc: descrição e análise crítica com vistas ao projeto Lacio-Web**. NILC-TR-03-03, 60 p., 2003.

RADFORD, Andrew. *Syntax: A Minimalist Introduction*. Cambridge: Cambridge University Press, 1997.

RAPOSO, Eduardo. **Teoria da Gramática: A Faculdade da Linguagem**. Lisboa: Caminho, 1992.

SCHWARZE, Christoph. **Introduction à la sémantique lexicale**. Tübingen: Narr, 2001.

TICIO, Emma. Locality Conditions in Spanish DPs. In: **Selected proceedings of the 8<sup>th</sup> Hispanic Linguistics Symposium**, ed. Timothy L.Face and carol A.Klee, Somerville, MA: Cascadilla Proceedings Project, 2006, p.137-153.

VALOIS, Daniel. **The internal syntax of DP**. Doctoral Dissertacion, University of California, L.A., 1991.

VIEIRA, Renata. e LIMA, Vera Lúcia. **Linguística computacional: princípios e aplicações**. In: FERREIRA, Carlos Eduardo (Ed.). **As tecnologias da informação e a questão social: anais**. v.2, 2001. p.47-88.

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)



[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)