O Formalismo GPSG e sua Comparação com DCG e LFG

Thiago Alexandre Salgueiro Pardo
Lucia Helena Machado Rino

Sumário

A abordagem da pesquisa em andamento visa à exploração comparativa de alguns formalismos gramaticais para o processamento de línguas naturais.

Neste relatório, é apresentado o formalismo gramatical "Gramática de Estrutura de Frase Generalizada", ou Gramática Normativa, assim como uma análise comparativa com os formalismos Gramática de Cláusulas Definidas e Gramática Léxico-Funcional e uma interface computacional para redução automática.

1. Introdução

Neste relatório, o formalismo gramatical Gramática de Estrutura de Frase Generalizada (*Generalized Phrase Structure Grammar* – GPSG) (Gazdar et al., 1985) é explorado visando a comparação com os formalismos gramaticais Gramática de Cláusulas Definidas (*Definite Clause Grammar* – DCG) (Pereira and Warren, 1980) e Gramática Léxico-Funcional (*Lexical-Functional Grammar* – LFG) (Kaplan and Bresnan, 1982). Essa comparação, por sua vez, tem o objetivo de verificar possíveis reduções entre tais formalismos, ou seja, dada uma especificação em um formalismo, deseja-se saber se é possível reescrevê-la em outro formalismo. Em especial, discute-se aqui a redução entre DCG e GPSG e entre LFG e GPSG. Ilustrações de possíveis reduções são geradas automaticamente por meio de uma interface construída em Delphi, já apresentada em (Pardo, 1999). Essa interface foi modificada para suportar a redução DCG-GPSG, e, por esse motivo, ela é reapresentada neste relatório.

Nas seções que se seguem são apresentados o formalismo GPSG (Seção 2) e a comparação entre os formalismos DCG, LFG e GPSG, seguindo-se uma avaliação da possibilidade de redução entre os mesmos (Seção 3). A interface computacional para redução entre DCG e os formalismos LFG e GPSG é apresentada na Seção 4. A conclusão desse estudo é apresentada na Seção 5.

2. A Gramática de Estrutura de Frase Generalizada

2.1. Conceitos básicos

Desenvolvida sob a ótica da teoria lingüística, a Gramática de Estrutura de Frase Generalizada (GPSG) é um formalismo flexível e abrangente capaz de representar de forma simples e elegante os fenômenos lingüísticos presentes na maioria das línguas. Originalmente baseada em gramáticas livres de contexto, a GPSG se diferencia pela adoção da Teoria X-Bar (Chomsky, 1970), da existência de não monotonicidade para as categorias gramaticais (Jakobson et al., 1951), da classificação das regras gramaticais em regras de precedência linear e de dominância imediata e pela definição de princípios de boa formação que permitem a estruturação lingüística de maneira não determinística, o que atribui um alto nível de abstração à GPSG.

A teoria de Chomsky sobre a Teoria X-Bar indica a hierarquia existente entre os componentes de uma sentença através da associação de índices aos mesmos, cujo processo se dá da seguinte forma:

- os itens lexicais, como determinante, substantivo e verbo, associar-se-ão ao índice 0;
- os itens pré-terminais, ou seja, aqueles que introduzem os itens lexicais, associar-se-ão ao índice 1;
- os itens não terminais, por sua vez, associar-se-ão ao índice 2.

A Figura 1 ilustra regras gramaticais para a formação de sentenças S-V-O (Sujeito-Verbo-Objeto), enquanto que a Figura 2 exibe a associação de índices aos termos da sentença.

sentença → sintagma_nominal, sintagma_verbal
sintagma_nominal → determinante, substantivo
sintagma_verbal → verbo, sintagma_nominal

Figura 1 – Regras gramaticais para a construção S-V-O

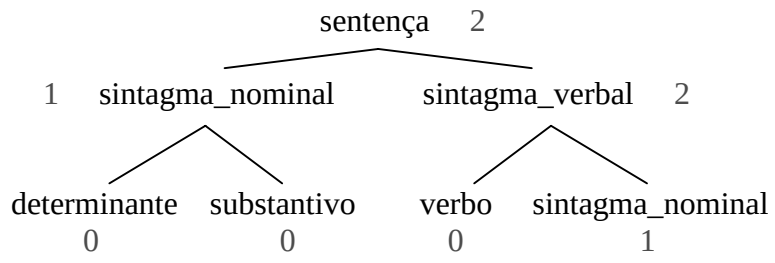


Figura 2 – Associação de índices na Teoria X-Bar

Com relação à não monotonicidade, a GPSG assume que as principais categorias sintáticas, ou seja, substantivo, verbo, adjetivo e preposição, podem ser fragmentadas em pelo menos dois traços (*features*) - nominal e verbal - de acordo com a tabela 1.

Tabela 1 – Classificação das Classes Gramaticais

	[+N]	[-N]
[+V]	**Adjetivo**	**Verbo**
[-V]	**Substantivo**	**Preposição**

A classe adjetivo, por exemplo, pode referenciar tanto substantivos (+N) como verbos (+V), enquanto que a classe substantivo pode somente se referir a substantivos. De acordo com essa teoria de traços, portanto, cada entidade pode ser referenciada tanto por sua categoria sintática quanto por seu conjunto de *features*. Segundo Harman (1963), a fragmentação das categorias sintáticas em partes menores atribui maior potencial descritivo ao formalismo.

A Tabela 2 ilustra as *feature*s consideradas pela GPSG e seus valores possíveis, segundo Gazdar et al. (1985).

Tabela 2 – *Features* e seus valores

Features	**Valores**	**Significado**
CASE	{ACC, NOM}	Indica o tipo do sujeito.
COMP	{*para, que, se,* NIL}	Indica qual a palavra usada para introduzir um complemento.
CONJ	{*e, ambos, mas, ou,* NIL}	Indica qual a conjunção desejada.
GER	{+, -}	Indica se é gerúndio.
NEG	{+, -}	Indica se é negação.
NFORM	{NORM}	Indica o tipo de sintagma verbal. O valor padrão é NORM (normal).
NULL	{+, -}	Indica se é um item nulo.
POSS	{+, -}	Indica se existe relação de posse.
REMOR	{RECP, REFL}	Indica o tipo do pronome: recíproco ou reflexico.
WHMOR	{R, Q, ...}	Indica o tipo do pronome: relativo ou interrogativo.
AGR	CAT	Indica a concordância entre os termos.
ADV	{+, -}	Indica se é advérbio.
AUX	{+, -}	Indica se é verbo auxiliar.
BAR	{0 , 1, 2}	Indica a X-Bar.
INV	{+, -}	Indica a inversão sujeito-verbo.
LOC	{+, -}	Indica a presença de um locativo.
N	{+, -}	Indica se é nominal.
PAST	{+, -}	Indica se o verbo está no passado.
PER	{1, 2, 3}	Indica em qual pessoa está.
PFORM	{*para, por, ...*}	Indica a preposição exigida.
PLU	{+, -}	Indica se está no plural.
PRD	{+, -}	Indica se é predicativo.
SLASH	CAT	Indica as categorias omitidas.
SUBCAT	{1, ..., n} U {*para, que, e, ambos, ...*}	Indica a subcategorização das palavras
SUBJ	{+, -}	Indica se possui sujeito.
V	{+, -}	Indica se é verbal.
VFORM	{BSE, FIN, INF, PAS, PRP, PSP}	Indica a forma verbal.
RE	CAT	Indica a concordância de elementos anafóricos recíprocos e reflexivos.
WH	CAT	Indica a concordância de termos interrogativos.

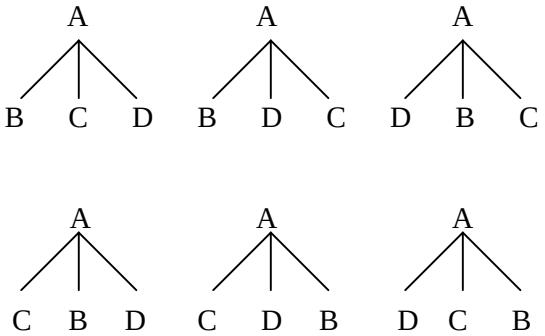
2.2. A sintaxe da GPSG

As regras gramaticais da GPSG se dividem em regras de dominância imediata (regras ID) e regras de precedência linear (regras LP). As regras ID especificam os termos de uma sentença e as relações de dominância entre eles, enquanto que as regras LP determinam a precedência entre os termos do lado direito das regras ID em uma árvore local, ou seja, entre os termos de um mesmo nível da árvore. Essa estruturação de regras é chamada de formato ID-LP.

Supondo a existência da seguinte regra ID:

A → B, C, D

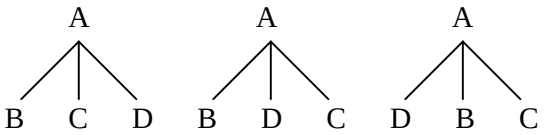
as seguintes derivações são possíveis:



Assim sendo, as regras ID não determinam a ordem dos elementos do lado direito das mesmas, apenas determinam que A domina imediatamente e é formado por B, C e D. Já a seguinte regra LP:

B < C

que especifica que B deve ser derivado antes de C, restringiria as derivações acima, permitindo somente as seguintes derivações:



A GPSG também utiliza o artifício de identificação de núcleos (*heads*) de cada termo de uma sentença. Assim, a GPSG assume basicamente que o núcleo de uma sentença é o sintagma verbal, que o núcleo de um sintagma nominal é um substantivo e que o núcleo de um sintagma verbal é um verbo, os quais são identificados de maneira diferenciada pela GPSG através da letra H. Além disso, as regras gramaticais em GPSG adotam a forma normal de Greibach, ou seja, admitem a forma A → aβ, para 'A' não terminal, 'a' terminal e 'β' uma cadeia de símbolos não terminais.

A Figura 3 ilustra a gramática da Figura 2 exibida no formato ID-LP.

Regras ID:

sentença → sintagma_nominal, H[-SUBJ]
sintagma_nominal → determinante, H^1
N^1 → H[1]
sintagma_verbal → H[2], sintagma_nominal

Regras LP:

[SUBCAT] < ~[SUBCAT]
sintagma_nominal < sintagma_verbal

Figura 3 – Gramática no formato ID-LP

De acordo com a gramática da Figura 3, tem-se que:

- A regra 'sentença → sintagma_nominal, H[-SUBJ]' indica que uma sentença é formada por um sintagma nominal e um núcleo, o qual é obrigatoriamente um sintagma verbal, e caso contrário, deveria ser representado explicitamente, em vez de ser representado em forma de

núcleo. A *feature* [–SUBJ] indica que o sintagma verbal não contém um sujeito, sendo essa característica que o diferencia de uma sentença na GPSG, pois ambas as estruturas possuem a mesma importância na hierarquia de Chomsky;

- A regra 'sintagma_nominal → determinante, H^1' indica que o sintagma nominal é formado por um determinante e um núcleo, o qual é obrigatoriamente um substantivo, caso contrário, deveria ser representado explicitamente por sua categoria gramatical, em vez de ser representado por H. O número sobrescrito ao lado de H é o índice que representa a X-Bar correspondente;
- Como o sintagma nominal introduz obrigatoriamente um substantivo, o termo H^1 irá necessariamente se instanciar com N^1, o qual introduz a regra 'N^1 → H[1]'. O termo H[1] representa, assim, um substantivo, designado pelo número 1 entre colchetes, que expressa a *feature* [SUBCAT 1], ou seja, a subcategorização de substantivos.
- A regra 'sintagma_verbal → H[2], sintagma_nominal' indica que o sintagma verbal é formado por seu núcleo, que é um verbo, e um sintagma nominal. O número 2 entre colchetes expressa a *feature* [SUBCAT 2], ou seja, a subcategorização de verbos transitivos diretos.

Somente os itens lexicais possuem a *feature* [SUBCAT]. Assim, a regra LP [SUBCAT]<~[SUBCAT] indica que os termos lexicais devem ser derivados antes dos termos não lexicais, os quais são expressos por ~[SUBCAT], onde o ~ indica que a *feature* não é definida para este tipo de termo. Desse modo, a forma normal de Greibach será satisfeita. De um modo geral, enquanto *–FEATURE* indica a ausência da propriedade indicada por *feature*, *~FEATURE* expressa a inexistência da definição de *feature* para um dado termo.

A Figura 4 ilustra outro exemplo de gramática, sendo a *feature* [*para*] uma abreviação utilizada para representar [PFORM *para*]. Assim, a regra 'sv --> H[3], sn, sp[*para*]' exige que a preposição do sp (sintagma preposicional) seja a preposição para. Neste exemplo, os verbos com subcategorização 2 são os verbos transitivos diretos, enquanto os com subcategorização 3 são transitivos indiretos.

Regras ID:

sentença --> sn, H[-SUBJ]
sn --> det, H^1
N^1 --> H[1]
sv --> H[2], sn
sv --> H[3], sn, sp[*para*]

Regras LP:

[SUBCAT] < ~[SUBCAT]
$N^2 < P^2 < V^2$

Figura 4 – Exemplo de gramática em GPSG

A associação de *feature*s aos termos das regras gramaticais deve seguir os princípios da Restrição de Co-ocorrência das *Features* (*Feature Co-occurrence Restriction* - FCR) e da Especificação de *Defaults* de *Features* (*Feature Specification Default* - FSD). Enquanto as FCRs representam a possibilidade de combinação entre as *feature*s, as FSDs inicializam os valores das *feature*s. A Figura 5 e 6 exibem, respectivamente, algumas FCRs e FSDs.

FCR1: [VFORM] ⊃ [+V, -N]
FCR2: [NFORM] ⊃ [-V, +N]
FCR3: [PFORM] ⊃ [-V, +N]
FCR4: [SUBCAT] ⊃ ~[SLASH]
FCR5: [BAR 0] ⊃ [SUBCAT]
FCR6: [BAR 1] ⊃ ~[SUBCAT]
FCR7: [BAR 2] ⊃ ~[SUBCAT]
FCR8: [+SUBJ] ⊃ [+V, -N, BAR 2]
FCR9: [+NULL] ⊃ [SLASH]

Figura 5 – Exemplos de FCRs

FSD1: [-INV]
FSD2: ~[CONJ]
FSD3: ~[NULL]
FSD4: ~[NOM]
FSD5: [+N, -V, BAR 2] ≡ [ACC]

Figura 6 – Exemplos de FSDs

Por exemplo, a FCR1 pode ser interpretada como:

"se o termo for definido para a *feature* VFORM, o termo deve ser do tipo [-N][+V]"

enquanto a FSD1 inicializa a *feature* INV com -, a qual indica a inversão sujeito-verbo.

Além das regras no formato ID-LP, das FCRs e das FSDs, ainda existem as meta-regras, as quais são responsáveis por alterar a estrutura das regras gramaticais. Por exemplo, a seguinte meta-regra expressa a transformação da voz ativa para a voz passiva:

SV --> W, SN
⇓
SV[PAS] --> W, (SP[*por*])

Neste caso, para 'SV --> W, SN', W é qualquer conjunto de termos. A transformação expressa por 'SV[PAS] --> W, (SP[*por*])' indica a passivização do verbo (*feature* [PAS]) que implica a mudança de ordem de W e a inclusão de um sintagma preposicional sp, opcional (indicado pelos parênteses). Ao aplicar essa meta-regra à gramática da Figura 4, a gramática da Figura 7 é gerada.

Regras ID:

sentença --> sn, H[-SUBJ]
sn --> det, H^1
N^1 --> H[1]
sv --> H[2], sn
sv --> H[3], sn, sp[*para*]
sv[PAS] --> H[2], (sp[*por*])
sv[PAS] --> H[3], sp[*para*], (sp[*por*])

Regras LP:

[SUBCAT] < ~[SUBCAT]
$N^2 < P^2 < V^2$

Figura 7 – Aplicação da Meta-regra de Passivização

Na primeira regra, "sv[PAS] → H[2], (sp[por])", o termo W se instancia com H[2] e, após a conversão, o SN é excluído da regra introduzida por SV e o SP opcional é incluído. Já na segunda regra, "sv[PAS] → H[3], sp[para], (sp[por])", W se instancia com H[3] e SP[*para*], também excluindo o SN e adicionando o SP opcional. Deste modo, regras que permitem a derivação/interpretação de sentenças na forma passiva são acrescentadas a gramática.

Segundo Gazdar et al. (1985), as meta-regras não alteram o poder expressivo da GPSG, apenas expressam generalizações de fenômenos lingüísticos, como a alteração do tópico sentencial, no caso da regra de passivização apresentada acima, que pode ser diretamente extraída de sua respectiva regra na voz ativa.

Para que uma gramática possa ser representada no formato ID-LP, ela não deve violar a ECPO (*Exhaustive Constant Partial Ordering*), que é uma restrição do poder expressivo da GPSG. Isso se deve ao fato das regras gramaticais da GPSG serem classificadas em regras de dominância imediata e de precedência linear, sendo as últimas responsáveis por essa restrição.

A ECPO restringe as gramáticas que apresentam uma ordem parcial dos termos. Por exemplo, sejam as gramáticas 1 e 2 abaixo:

Gramática 1: A → B, C, D
B → E
C → B, D
Gramática 2: A → B, C, D
B → E
C → D, B

A Gramática 1 poderia ser escrita no formato ID-LP, pois não infringe a ECPO, enquanto a Gramática 2 não poderia, por infringir a ECPO na regra C → D, B. Isso ocorre porque C introduz os termos ordenados D e B, enquanto A introduz os termos ordenados B e D. Para isso ser solucionado, a primeira ou a terceira regra teria que admitir a ordenação dos termos da outra também, ou seja, a Gramática 2 teria que ser reescrita gerando a Gramática 3 ou 4 para que fosse aceita no formato ID-LP, como segue:

Gramática 3: A → B, C, D
A → D, C, B
B → E
C → D, B

Gramática 4: A → B, C, D
B → E
C → D, B
C → B, D

O problema surge devido ao conflito gerado nas regras de precedência que, de acordo com a primeira regra da Gramática 2, admitiria B < D, enquanto que, de acordo com a terceira regra, admitiria D < B. Nas Gramáticas 3 e 4 esse problema é solucionado pela inexistência de ordenação entre B e D (imposta pelas regras 1 e 2 da gramática 3 ou pelas regras 3 e 4 da gramática 4). Quanto à Gramática 1, somente uma ordem é imposta pelas regras 1 e 3 (sendo B < D), e, deste modo, não há conflito.

Dependendo do grau de complexidade da gramática, reescrevê-la pode alterá-la de forma a introduzir erros ou incoerências, sendo aconselhável, neste caso, desconsiderar a ECPO ou reescrever a gramática em outro formalismo gramatical.

2.3. O léxico na GPSG

A GPSG assume que uma entrada lexical deve conter basicamente informações sobre a forma fonológica, a categoria sintática, as formas amorfas e o significado de cada item lexical. A Figura 8 ilustra um exemplo de léxico simples que adota as características citadas acima.

< *comprar*, [[-N][+V][BAR 0][SUBCAT 1]], {}, **comprar'** >

< *dar*, [[-N][+V][BAR 0][SUBCAT 2]], {}, **dar'** >

< *menino*, [[+N][-V][BAR 0][SUBCAT 3]], {}, **menino'** >

< *brinquedo*, [[+N][-V][BAR 0][SUBCAT 3]], {}, **brinquedo'** >

< *belo*, [[+N][+V][BAR 0][SUBCAT 4]], {}, **belo'** >

< *o*, Det, {}, **o'** >

Figura 8 – Exemplo de Léxico na GPSG

Considerando-se, por exemplo, a entrada

< *brinquedo*, [[+N][-V][BAR 0][SUBCAT 3]], {}, **brinquedo'** >

tem-se que brinquedo é a forma fonológica, cuja subcategorização é 3 (substantivo) e é nominal e não verbal, possuindo X-Bar 0, ou seja, é um item lexical e, além disso, não possui nenhuma morfologia irregular, o que é indicado pelas chaves vazias {}. O significado de brinquedo será utilizado para construir sua forma semântica, aqui representado pela própria palavra em sua forma canônica, em negrito.

2.4. Princípios de boa formação de sentenças na GPSG

Durante o processamento de uma sentença a partir de regras gramaticais em GPSG (para síntese ou análise), as *feature*s da estrutura da sentença podem ser herdadas das regras gramaticais ou instanciadas. Caso as *feature*s sejam herdadas das regras gramaticais, assume-se que as *feature*s das regras estão corretas e que a sentença é bem formada. Caso as *feature*s sejam instanciadas, são necessários princípios de boa formação que garantam a correta associação de

tais *features* aos termos das sentenças. Esses princípios incluem a Convenção das *Head Features* (*Head Feature Convention* - HFC), o Princípio das *Foot Features* (*Foot Feature Principle* - FFP) e o Princípio de controle de concordância (*Control Agreement Principle* - CAP).

A convenção HFC controla a instanciação das *head feature*s, sendo elas N, V, PLU, PER, VFORM, SUBJ, PFORM, AUX, INV, PAST, PRD, ADV, SLASH, AGR, SUBCAT, BAR e LOC (refira-se à Tabela 2 para mais detalhes sobre essas *features*). A HFC garante que toda *head feature* instanciada em um nó filho será instanciada também no nó pai. Assim, considerando-se as seguintes regras gramaticais:

S → SN, H[-SUBJ]
SN → DET, H^1
N^1 → H[1]
SV[PAST] → H[2]

SN < SV

tem-se as seguintes derivações [1]:

Derivação 1: S
    SN
    H[-SUBJ]

Derivação 2: S[-SUBJ]
    SN
    H[-SUBJ]

Derivação 3: S[FIN, -PLU]
    SN[-PLU]
    H[-SUBJ, FIN, -PLU]

Derivação 4: S[PER 3, +PLU]
    SN[PER 2, +PLU]
    H[+PLU]

Derivação 5: SV[PAST]
    H[SUBCAT 2]

Essas derivações indicam o seguinte:

- A derivação 1 está de acordo com a HFC, pois foi projetada a partir da primeira regra, ou seja, a *feature* [-SUBJ] foi herdada da primeira regra da gramática, logo não pode ser instanciada no nó pai S;
- A derivação 2 é o contrário do que foi dito anteriormente, ou seja, violou a HFC, logo é uma derivação incorreta;
- A derivação 3 instanciou as *feature*s [FIN] e [-PLU] nos nós filhos, e segundo a HFC, também instanciou no nó pai. A *feature* [-SUBJ] encontra-se corretamente em H e não em S por ser herdada da regra. Essa derivação está correta;
- A derivação 4 é incoerente, pois a *feature* [-SUBJ] que deveria ser herdada da primeira regra da gramática não se encontra presente na derivação e o número da pessoa representado pela *feature* [PER] está erroneamente instanciada, ou seja, enquanto o nó filho SN possui a *feature* [PER 2], o nó pai possui [PER 3];

[1] A representação em colunas indentadas corresponde à representação hierárquica arbórea, isto é, cada coluna corresponde a um nível distinto na árvore de representação.

- A derivação 5 está correta, pois herdou da regra 4 as *feature*s [PAST] e [SUBCAT 2];

O princípio FFP controla as relações impostas pelas *foot feature*s, sendo elas SLASH, WH e RE (vide Tabela 2). Ele garante, da mesma forma que a HFC, que todas as *foot feature*s instanciadas em nós filhos durante a derivação de uma sentença sejam instanciadas no nó pai também.

As *foot feature*s são normalmente utilizadas na presença de dependências de longa distância, as quais são analisadas na Seção 5.

O último dos princípios de boa formação de uma sentença é o CAP, o qual controla as relações de concordância entre os componentes de uma sentença. O CAP adota os preceitos de Keenan (1974), ou seja, a noção de que argumentos devem concordar com os termos principais de uma sentença. A Tabela 3 ilustra alguns exemplos de concordância:

Tabela 3 – Associação de Argumentos e Termos

Argumento	**Termo**
Det	N^1
N^1	SN
SV	SN
SN	S

Assim, o CAP garante que os traços de concordância (gênero, número, pessoa, ...) sejam instanciados no argumento da relação de concordância através da *feature* [AGR], por exemplo:

Derivação 1: SN
Det[AGR Subst[-PLU]]
Subst[-PLU]

Derivação 2: SN
Det[AGR Subst[+PLU]]
Subst[-PLU]

Derivação 3: S
SN[PER 3, -PLU]
SV[AGR SN [PER 3, -PLU]

Derivação 4: S
SN[PER 3, -PLU]
SV[AGR SN [PER 1, -PLU]

Derivação 5: SV[AGR SN[PER 3, +PLU]
H[10]
SV[AGR SN [PER 3, +PLU]

Neste caso, as derivações ilustram o seguinte:

- A derivação 1 indica que o determinante concorda com o substantivo. O CAP garante que o traço de concordância ([-PLU]) também esteja presente no determinante através da *feature* AGR;
- A derivação 2 está inconsistente, pois a determinante concorda com um substantivo no plural, enquanto que o substantivo existente está no singular. Esse tipo de relação permitira sentenças erradas como "*Os menino gastou todo o dinheiro*";

- A derivação 3 indica que o SV concorda com o SN. O CAP garante que os traços de concordância ([PER 3],[-PLU]) também estejam presentes em SV dentro da *feature* AGR;
- A derivação 4 está inconsistente, pois a SV concorda com o SN na primeira pessoa do singular, enquanto que o SN existente está na terceira pessoa do singular. Esse tipo de relação permite sentenças erradas como "*O menino gastei todo o dinheiro*";
- A derivação 5 está correta e é responsável por sentenças em que uma conjunção verbal concorda com um sintagma nominal. Essa derivação produz sentenças como "*Eles insistem em se queixar*".

Com os princípios de boa formação especificados pela GPSG atuando conjuntamente, é possível garantir a estruturação correta de sentenças.

2.5. Dependências de longa distância

Através da adoção da *feature* SLASH, a GPSG introduz uma forma simples e elegante de representar dependências de longa distância. Sendo a *feature* SLASH tanto uma *head* como *foot feature*, o controle sobre o fenômeno ocorre de forma natural, pois tal *feature* é passada aos diferentes níveis da hierarquia da estrutura da sentença, representando o fenômeno corretamente, como pode ser notado, por exemplo, na Figura 9.

A *feature* SLASH, com o auxílio da *feature* NULL, é capaz de representar termos omitidos/deslocados de uma sentença. Enquanto a *feature* SLASH se associa a partes de sentenças de onde o termo omisso foi extraído, a *feature* NULL se associa ao próprio termo omisso. Por exemplo, considerando a regra:

$S \rightarrow X^2, S/X^2$

X representa uma categoria qualquer e S/X^2 uma abreviação para S[SLASH X^2], ou seja, S que teve um termo qualquer X^2 extraído de si. Assim, conforme indicado pela regra acima, uma sentença S pode ser estruturada a partir de um termo qualquer deslocado. As Figuras 9, 10 e 11 ilustram possíveis dependências de longa distância a partir dessa regra, sendo ε a cadeia vazia. Essas estruturas representam, respectivamente, as estruturas profundas das sentenças "Kim queria estar orgulhosa de Fido" com o termo SA (sintagma adjetivado) deslocado ("*orgulhosa de Fido*") e "Kim quer dar Fido para Sandy" com o sintagma preposicional deslocado ("*para Sandy*") e, posteriormente, o sintagma nominal deslocado ("*Kim*").

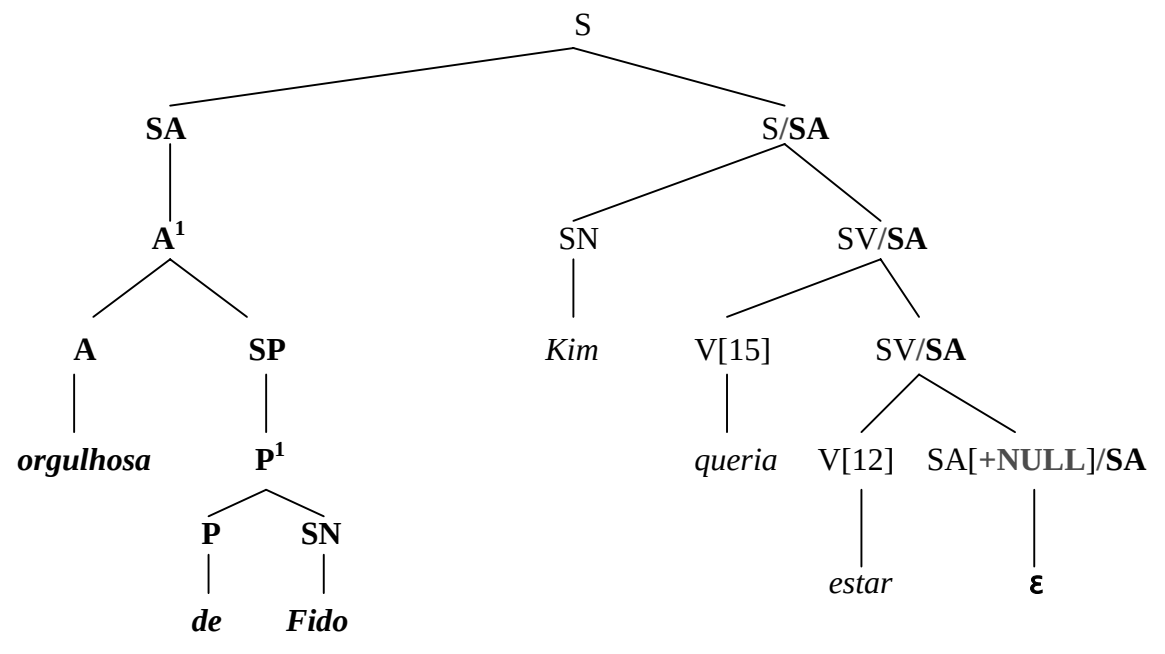


Figura 9 – Exemplo de Gramática com Dependência de Longa Distância

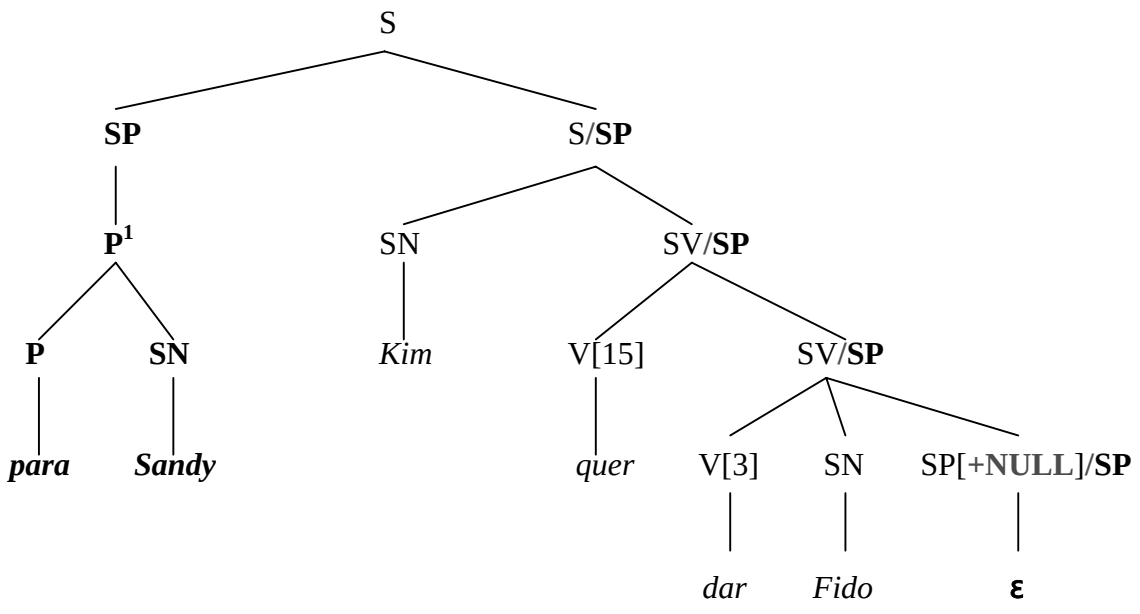


Figura 10 - Exemplo de Gramática com Dependência de Longa Distância

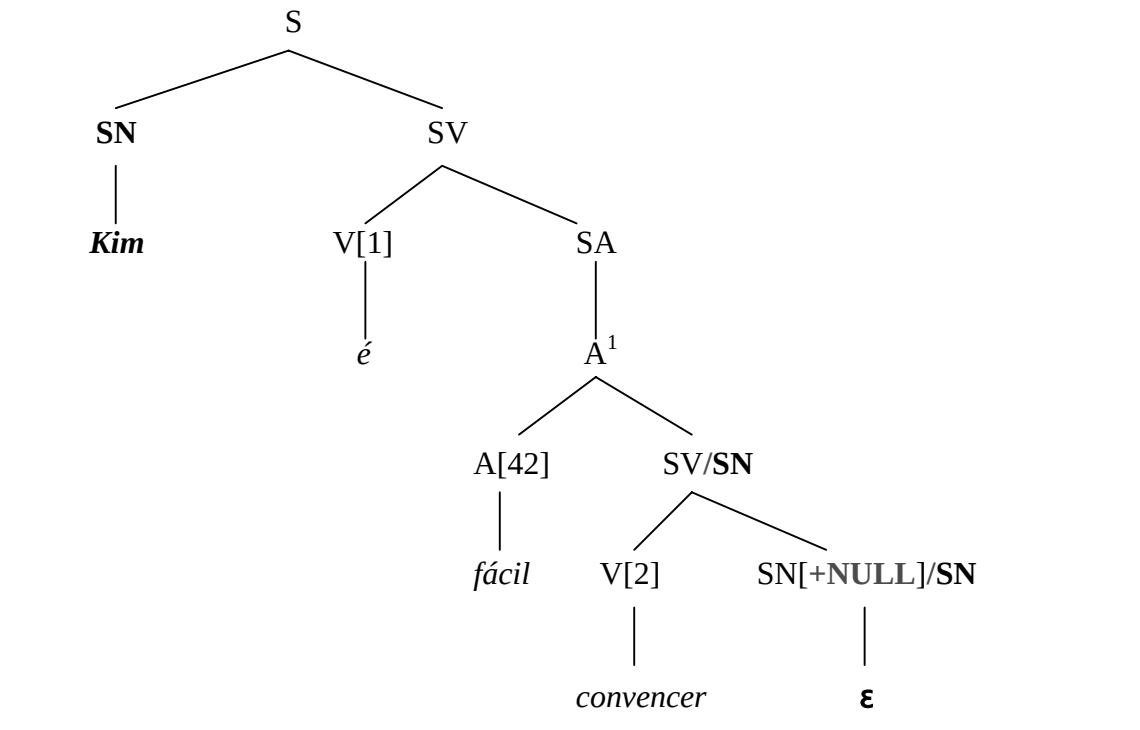


Figura 11 - Exemplo de Gramática com Dependência de Longa Distância

A partir das características principais da GPSG apresentadas nesta seção, serão discutidas as propriedades deste formalismo em comparação com as propriedades dos formalismos DCG e LFG.

3. A GPSG e os Formalismos DCG e LFG

Nesta seção são comparadas as principais características dos formalismos DCG, LFG e GPSG, com o objetivo de verificar as vantagens e desvantagens de cada um deles e averiguar a possibilidade de redução entre os mesmos. Em particular, é abordada neste estudo somente a possibilidade de redução entre DCG e GPSG e entre LFG e GPSG. A redução DCG-LFG já foi amplamente estudada anteriormente (Pardo, 1999) e é reproduzida neste relatório para efeito de síntese da interface de redução apresentada na Seção 4.

As propriedades indicadas na Tabela 4 são extraídas de (Kaplan and Bresnan, 1982) e Shieber (1986) e consistem no conjunto mínimo considerado relevante para a averiguação da possibilidade de redução entre os formalismos.

Tabela 4 – Quadro comparativo das propriedades da DCG, LFG e GPSG

	DCG	**LFG**	**GPSG**
Concepção	Ferramenta Lingüística	Teoria Lingüística	Teoria Lingüística
Restritividade	Sim	Sim	Sim
Computabilidade	Alta	Alta	Média
Expressividade	Média	Alta	Alta
Simplicidade	Média	Alta	Média
Capacidade de Representação	Alta	Alta	Alta
Suporte Computacional	Sim	Não	Não
Monotonicidade	Sim	Sim	Não
Unificação	Alto Nível	Alto Nível	Baixo Nível
Codificação da Informação Lingüística	+Léxico +Regras	+Léxico -Regras	+Regras -Léxico
Formato das Regras	Único	Único	ID-LP
Meta-regras	Inexistentes	Lexicais	Gerais
Níveis de Estruturação Sintática	Um	Dois	Um
Controle de Formação de Estruturas	Argumentos	Esquemas	*Feature*s, FCRs e FSDs
Garantia de Boa Formação de Sentenças	Argumentos	Condição de Unicidade, Completeza e Coerência	HFC, FFP, CAP
Dependências de Longa Distância	Argumentos	Metavariáveis de Dominância Limitada	Slash *feature*
Poder Matemático	Médio	Alto	Baixo
Nível de Abstração	Baixo	Baixo	Alto
Semântica	Cálculo de Predicados	Formas Semânticas	Denotações

A seguir, cada uma dessas propriedades é comentada, justificando-se a comparação apresentada na tabela acima.

3.1. Concepção

A DCG teve por base o conceito de ferramenta lingüística como uma representação computacional para as línguas naturais, enquanto que a LFG e a GPSG foram desenvolvidas a partir de teorias lingüísticas, procurando caracterizar os princípios universais da linguagem.

3.2. Restritividade

Um formalismo é restrito se ele possui domínio e forma bem definidos, podendo ser, portanto, bem representado e analisado.

Os três formalismos são restritos, sendo sua principal característica a fundamentação em gramáticas livres de contexto que, por sua vez, sugere a definição clara do domínio e da forma gramatical.

3.3. Computabilidade

Um formalismo é computável se pode ser algoritmizado. Por serem formalismos restritos e bem delimitados, tanto a DCG quanto a LFG possuem alto nível de computabilidade. Na GPSG, o alto nível de abstração (para maiores detalhes sobre o que se considera nível de abstração, refira-se a (3.18)) lhe confere maior dificuldade para ser algoritmizada, implicando, portanto, em um menor nível de computabilidade.

3.4. Expressividade

Expressividade é a facilidade de um formalismo para expressar as propriedades da língua ou de um subconjunto de construções lingüísticas. Em geral, utiliza-se, para tanto, uma linguagem matemática. A LFG e a GPSG são mais expressivas que a DCG, pois utilizam, respectivamente, esquemas e *feature*s, enquanto a DCG possui somente o recurso de inclusão de argumentos para aumentar seu poder expressivo.

3.5. Simplicidade

A representação da gramática de uma língua em LFG se faz de forma mais simples que em DCG e GPSG, pelo fato da LFG representar toda a complexidade em seu léxico. Já a DCG, que o faz em suas próprias regras gramaticais através da adição de argumentos, pode se tornar bastante complexa, dependendo do tamanho e do nível de dependência contextual entre os constituintes da gramática em questão.

Devido à não monotonicidade de suas categorias gramaticais, ou seja, à fragmentação das categorias em *features*, a GPSG se torna um pouco mais complexa que os outros formalismos, pois precisa garantir a correta utilização das *feature*s postulando princípios de boa formação.

3.6. Capacidade de representação

A capacidade de representação está relacionada à quantidade de línguas que um formalismo abrange. Neste caso, os três formalismos apresentam uma alta capacidade de representação, apesar da *Exhaustive Constant Partial Ordering* (ECPO) restringir a expressividade da GPSG com a exigência de uma certa estruturação das regras ID para a boa formação das regras LP.

3.7. Suporte computacional

Suporte computacional, neste caso, refere-se à existência de algum recurso computacional já desenvolvido para manipulação de um formalismo. Somente a DCG apresenta suporte computacional, por estar incorporada ao ambiente de programação Prolog.

3.8. Monotonicidade

A monotonicidade é a característica de um formalismo apresentar suas categorias gramaticais principais como forma básica, sem decomposição adicional. Neste caso, a não

monotonicidade é a possibilidade de fragmentação das categorias gramaticais dos componentes lingüísticos em unidades menores.

Tanto a DCG quanto a LFG são monotônicas, enquanto que a GPSG, por ter suas categorias divididas em *feature*s, é não monotônica.

3.9. Unificação

A operação de unificação na GPSG ocorre em um nível mais baixo que na DCG ou na LFG: enquanto a DCG e a LFG se baseiam nas categorias gramaticais para a unificação, a GPSG se baseia nas *feature*s, implicando, por um lado, a utilização de princípios e convenções para regulamentação de seu uso e, por outro, o alto nível de abstração do formalismo.

3.10. Codificação da informação lingüística

Na DCG a codificação da informação necessária para a derivação e/ou interpretação de uma sentença é representada integralmente nas regras. Assim, mesmo os itens lexicais são introduzidos via regras. Na LFG, a codificação ocorre quase que totalmente no léxico. Apenas algumas partes da informação se associam às regras em forma de esquemas, restrições e condições. Já na GPSG a codificação da informação ocorre nas regras, pela associação de seus termos às *feature*s.

3.11. Formato das regras

A DCG e a LFG utilizam o formato e definição das regras das gramáticas livres de contexto, enquanto que a GPSG utiliza o formato ID-LP, ou seja, regras de dominância imediata e regras de precedência linear.

3.12. Meta-regras

Enquanto as meta-regras da LFG são aplicadas em nível lexical, as meta-regras da GPSG se aplicam às próprias regras gramaticais, produzindo assim, um novo conjunto de regras. Já a DCG não faz uso de meta-regras.

3.13. Níveis de estruturação sintática

Enquanto a DCG e a GPSG apresentam somente um nível de descrição sintática, a LFG apresenta dois níveis de representação, ou seja, a estrutura de constituintes e a estrutura funcional.

3.14. Controle de formação de estruturas

O controle na formação de estruturas de uma sentença é feito por argumentos na DCG, por esquemas na LFG e por *feature*s, Restrições de Co-ocorrência das *Features* (FCRs) e Especificações de *Defaults* das *Features* (FSDs) na GPSG.

3.15. Garantia de boa formação de sentenças

A boa formação de sentenças na DCG é regulamentada pelos argumentos, enquanto que na LFG ela é regulamentada pela Condição de Unicidade e pelos princípios de completeza e coerência, os quais garantem a correta estruturação das sentenças e a abrangência da correspondência entre a informação lingüística das regras gramaticais e as estruturas sintáticas produzidas pela LFG. Já na GPSG, os princípios e convenções HFC (*Head Feature Convention*), FFP(*Foot Feature Principle*) e CAP(*Control Agreement Principle*) garantem a boa formação de sentenças.

3.16. Dependências de longa distância

Os três formalismos admitem a especificação de dependências de longa distância. Entretanto, na DCG elas são representadas por argumentos, na LFG pelas metavariáveis de dominância limitada (indicadas pelos símbolos $\Uparrow$ e $\Downarrow$) e na GPSG pela utilização da *feature* SLASH.

3.17. Poder matemático

A LFG é o formalismo com maior poder matemático por permitir a correspondência direta com funções matemáticas e seus valores. Já a DCG, por admitir desestruturação das regras livres de contexto, apresenta um poder matemático médio. A GPSG, por sua vez, possui poder matemático baixo devido ao alto nível de abstração.

3.18. Nível de abstração

Um formalismo com alta abstração é um formalismo que não explicita completamente o controle sobre seu funcionamento, ou seja, permite que estruturas lingüísticas sejam resolvidas (tanto no processo de análise quanto no de síntese) de maneira não determinística. Assim sendo, a GPSG é um formalismo com alta abstração, pois, apesar de garantir a estruturação correta de sentenças através de seus princípios, não tem controle sobre quais *feature*s podem ser instanciadas durante o processamento de uma sentença.

Na DCG e LFG, o oposto ocorre, ou seja, tudo que é derivado/interpretado está explícito na definição do formalismo através de argumentos na DCG e de esquemas na LFG.

3.19. Semântica

Enquanto a DCG expressa suas relações semânticas através das formas lógicas do Cálculo de Predicados e a LFG através das formas semânticas de seu léxico, a GPSG o faz através do modelo de denotações de Montague (1974).

4. A Redução entre DCG, LFG e GPSG

A redução entre DCG e LFG já foi estudada anteriormente (Pardo, 1999), sendo apresentada neste relatório, portanto, somente a redução entre DCG e GPSG e entre LFG e GPSG. A redução no caminho inverso, isto é, da GPSG para a DCG ou GPSG para a LFG também parece possível, para construções simples da língua, pois os três formalismos são restritivos e permitem, portanto, a correspondência entre suas representações. Porém, um pouco de expressividade é perdida quando a redução envolve a GPSG, devido ao seu alto nível de abstração. Casos mais complexos que envolvem, por exemplo, dependências de longa distância, não foram averiguados, necessitando, portanto, de estudos adicionais.

Os critérios de redução entre os formalismos se baseiam tanto em sua dependência notacional quanto em sua dependência conceitual. O grau de dependência notacional reflete a correspondência (parcial ou completa) entre os símbolos de cada formalismo envolvido, assim como a formação de suas regras gramaticais e a especificação do léxico. Por exemplo, a DCG e a LFG apresentam claramente a notação de uma GLC enquanto que a GPSG utiliza o formato ID-LP (*Immediate Dominance-Linear Precedence*). O grau de dependência conceitual reflete a correspondência contextual (parcial ou completa) entre as construções gramaticais permitidas pelos formalismos envolvidos.

Para a dependência notacional, são considerados os seguintes aspectos: a) a ocorrência de regras de formato único na DCG, que indicam tanto a dominância quanto a precedência entre os termos gramaticais, além de incluírem também as especificações lexicais; b) regras de formato único também na LFG, a qual, no entanto, representa independentemente as informações de ordem lexical, sendo estas responsáveis por expressar o nível de detalhe lingüístico em foco; c) as regras ID-LP, i.e., regras de dominância imediata e regras de

precedência linear na GPSG. Além desses aspectos, vale notar ainda que a DCG e LFG adotam o formato padrão da GLC em suas regras gramaticais, i.e., A → β, para 'A' não terminal e 'β' uma forma sentencial qualquer, enquanto a GPSG utiliza a forma normal de Greibach, i.e., A → aβ, para 'A' não terminal, 'a' terminal e 'β' uma cadeia de símbolos não terminais.

Para a dependência conceitual, os seguintes aspectos são relevantes: a) a construção argumental na DCG, sendo este seu único recurso para especificações dessa natureza; b) a utilização de esquemas, restrições e condições na LFG, respectivamente usados para representar a estruturação superficial, as delimitações de significado ou forma e as condições de determinação de escolhas superficiais em função do contexto; c) o uso de *features* na GPSG, associadas tanto aos termos das regras quanto aos próprios itens lexicais, os quais são representados separadamente das regras.

A gramática DCG da Figura 12, por exemplo, pode ser reduzida para a gramática LFG da Figura 13 e para a gramática GPSG da Figura 14 e vice-versa.

sentença → sn(Número, Gênero), sv(Número).
sn(Número, Gênero) → determinante(Número, Gênero), substantivo(Número, Gênero).
sn(singular, masculino) → nomepróprio.
sv(Número) → verbo(Número), sn(_, _).
sv(Número) → verbo(Número), sn(_, _), sp.
sp → preposição, sn(_, _).
determinante(singular, masculino) → [o].
determinante(plural, feminino) → [as].
substantivo(singular, masculino) → [menino].
substantivo(plural, feminino) → [flores].
nomepróprio → [maria].
verbo(singular) → [deu].
preposição → [a].

Figura 12 – Gramática DCG

REGRAS:

sentença → sn sv
(↑Suj)=↓ ↑=↓

sn → { determinante substantivo
nomepróprio }

sv → verbo sn (sp)
(↑ObjD)=↓ (↑(↓prep))=↓

sp → preposição sn
(↑ObjI)=↓

LÉXICO:

o: determinante, (↑Número)=singular
(↑Gênero)=masculino
(↑Det)=o
as: determinante, (↑Número)=plural
(↑Gênero)=feminino
(↑Det)=as
flores: substantivo, (↑Número)=plural
(↑Gênero)=feminino
(↑Pred)=’flores’
menino: substantivo, (↑Número)=singular
(↑Gênero)=masculino
(↑Pred)=’menino’
maria: nomepróprio, (↑Pred)=’maria’
deu: verbo, (↑Número)=singular
(↑Pred)=’deu<(↑Suj)(↑ObjD)(↑Prep ObjI)>’
a: preposição, (↑caso)=a

Figura 13 – Gramática LFG

REGRAS ID:

sentença → sn, H^2[-SUBJ]
sn → determinante, H^1
sn → nomepróprio
N^1 → H[1]
sv → H[2], sn
sv → H[3], sn, sp
sp → H[4], sn

REGRAS LP:

[SUBCAT] < ~[SUBCAT]
$N^2 < P^2 < V^2$

LÉXICO:

< *o*, determinante, {}, **o'** >
< *as*, determinante, {}, **as'** >
< *flores*, [[+N][-V][BAR 0][SUBCAT 1][+PLU]], {}, **flores'** >
< *menino*, [[+N][-V][BAR 0][SUBCAT 1][-PLU]], {}, **menino'** >
< *maria*, nomepróprio, {}, **maria'** >
< *deu*, [[-N][+V][BAR 0][SUBCAT 3][-PLU][PAST]], {}, **deu'** >
< *a*, [[-N][-V][BAR 0][SUBCAT 4][PFORM *a*]], {}, **a'** >

Figura 14 – Gramática GPSG

As gramáticas acima permitem reconhecer como válida, por exemplo, a sentença "o menino deu as flores a Maria".

A seguir discute-se os princípios de redução DCG-GPSG e LFG-GPSG.

4.1. Redução DCG-GPSG

Por serem formalismos restritivos, a redução de uma gramática DCG para uma equivalente em GPSG é possível e pode ser feita de forma completa, ou seja, sem perda de expressividade lingüística, devido à correspondência direta dos recursos da DCG com recursos da GPSG. Já a redução inversa, isto é, uma gramática GPSG para sua equivalente em DCG, pode, às vezes, ocorrer de forma parcial, devido ao alto nível de abstração da GPSG. Além disso, a GPSG apresenta recursos que não apresentam correspondência em DCG, como o uso de X-Bar e das meta-regras, implicando, assim, a perda de expressividade lingüística durante tal redução.

Os princípios de redução de uma gramática DCG para sua gramática GPSG equivalente são apresentados a seguir, não sendo considerados casos mais complexos como, por exemplo, a extraposição segundo Pereira (1981):

- Primeiramente, deve-se percorrer as regras da DCG formatando-as de acordo com a forma normal de Greibach. Por exemplo:

  DCG: sn(N,G) → det(N,G), **substantivo**(N,G).

  GPSG: sn → det, **Termo_Intermediário**
  Termo_Intermediário → substantivo

- Os termos das regras não lexicais (i.e., que não introduzem itens lexicais) que não são núcleos (*heads*) nem símbolos pré-terminais são mapeados diretamente na representação em GPSG. Caso o termo possua algum argumento que corresponda a alguma *feature* da GPSG, essa *feature* deve ser associada ao termo na GPSG. Por exemplo, um argumento associado a um termo na DCG indicando número como plural poderia ser mapeado para a *feature* [+PLU] em GPSG;
- Os termos das regras não lexicais que são núcleos (*heads*) e não são símbolos pré-terminais, são mapeados na GPSG adotando-se a representação de núcleo (a letra H). Caso o termo possua algum argumento que corresponda a alguma *feature* da GPSG, essa *feature* deve ser associada ao termo na GPSG. Por exemplo:

  DCG: sentença → sn(N,G), **sv**(N).

  GPSG: sentença → sn, **H**

- Os termos das regras não lexicais que são símbolos pré-terminais (podendo também ser núcleos), são mapeados na GPSG pela letra H, juntamente com a *feature* SUBCAT associada a um índice que indica o tipo do item lexical (substantivo, verbo transitivo direto, verbo transitivo indireto, verbo bitransitivo, verbo intransitivo, verbo de ligação, adjetivo e preposição). Cada tipo deve ter um único índice associado. Assim, caso existam dois itens lexicais de categoria substantivo, ambos devem ser associados ao mesmo tipo e, portanto, a um único índice. Caso o termo possua algum argumento que corresponda a alguma *feature* da GPSG, essa *feature* deve ser associada ao termo na GPSG. Por exemplo:

  DCG: sv(N) → **verbo**(N), sn(_,_).

  GPSG: sv → **H[1]**, sn

- As regras que introduzem itens lexicais devem dar origem a modificações no léxico da GPSG pela inclusão de suas respectivas informações. O léxico deve conter também a classe de palavras à qual o item pertence, as formas amorfas (caso sejam especificadas na DCG) e a informação semântica. Caso a classe à qual o item lexical pertença seja substantivo, verbo, adjetivo ou preposição, a classe do léxico deve ser substituída pelas *feature*s que a representam, e, além disso, caso o termo que introduza o item lexical possua algum argumento que corresponda a alguma *feature* da GPSG, essa *feature* deve ser associada ao item lexical da GPSG. Por exemplo:

  DCG: substantivo(sing,masc) → [menino].

  GPSG: < *menino*, [[+N][-V][BAR 0][-PLU][SUBCAT ...]], {}, **menino'** >

- Por último, deve-se percorrer as regras da DCG selecionando-se relações de precedência que sejam válidas, isto é, que não sejam contrariadas posteriormente por outras regras. Com isso, monta-se as regras de precedência linear da GPSG. Por exemplo:

  DCG: sentença → sn(N,G), sv(N).

  GPSG: sn < sv

Para a redução de uma gramática GPSG a uma gramática DCG, os princípios acima podem ser aplicados de forma reversa. Por exemplo:

- Os termos das regras GPSG que não são núcleos nem símbolos pré-terminais são mapeados diretamente na representação em DCG. Caso o termo possua alguma *feature* que indique

dependência contextual, essa *feature* deve ser mapeada para o argumento correspondente em DCG.

- Os termos das regras que são núcleos e/ou itens pré-terminais são mapeados em DCG na forma ordinária, isto é, em forma de categorias gramaticais. Caso o termo possua alguma *feature* que indique dependência contextual, essa *feature* deve ser mapeada para o argumento correspondente em DCG. Por exemplo:

  GPSG: sn → determinante, $\mathbf{H^1}$

  DCG: sn → determinante, **substantivo**.

- Cada entrada do léxico da GPSG deve ser mapeada em uma regra lexical na DCG, sendo o termo do lado esquerda da regra da DCG correspondente às *features* que o indicam, caso estas estejam presentes no léxico da GPSG. Por exemplo:

  GPSG: < *menino*, [[+N][-V][BAR 0][-PLU][SUBCAT ...]], {}, **menino'** >

  DCG: substantivo(singular) → [menino].

onde:
- as *features* [+N] e [-V] correspondem à classe gramatical substantivo;
- a *feature* [-PLU] da GPSG correspondente ao argumento de valor singular na DCG.

- Percorre-se as regras de precedência da GPSG e, segundo estas, formata-se as regras geradas em DCG. Por exemplo, considerando-se as regras em GPSG:

  S → SN, SV, ADJ_ADV

  SN < SV

as seguintes regras em DCG seriam geradas:

S → SN, SV, ADJ_ADV.
S → ADJ_ADV, SN, SV.
S → SN, ADJ_ADV, SV.

Como a regra de precedência linear da GSPG especifica somente que SN deve preceder SV ignorando o adjunto adverbial (ADJ_ADV), tal termo pode estar em qualquer posição na estrutura da sentença. Logo, devem existir regras na DCG que contemplem cada uma das possíveis posições do adjunto adverbial na sentença e, ao mesmo tempo, satisfaçam a restrição SN < SV.

4.2. Redução LFG-GPSG

A redução de uma gramática LFG para uma equivalente em GPSG é possível, por serem formalismos restritivos, e apesar de poder ser completa, há perda de expressividade lingüística, pois não há correspondência entre a informação representada pelos esquemas funcionais da LFG na GPSG. Já a redução inversa, isto é, uma gramática GPSG para uma gramática LFG, pode, às vezes, ocorrer de forma parcial, devido ao alto nível de abstração da GPSG. Além disso, a GPSG apresenta recursos que não apresentam correspondência em LFG, como o uso de X-Bar e das meta-regras gerais, implicando, assim, a perda de expressividade lingüística durante tal redução.

Os princípios de redução de uma gramática LFG para sua gramática GPSG equivalente são apresentados a seguir, não sendo considerados casos mais complexos como, por exemplo, a dependência de longa distância ou referência anafórica:

- Primeiramente, as regras duplicadas, ou seja, aquelas que introduzem mais de uma derivação/interpretação possível através do uso de chaves {} e/ou (), devem ser desmembradas, retirando-se assim, a ambigüidade. Por exemplo, a regra da LFG

$$sn \rightarrow \left\{ \begin{array}{l} \text{determinante substantivo} \\ \text{nomeproprio} \end{array} \right\}$$

deve ser desmembrada em:

sn → determinante substantivo
sn → nomeproprio

- Deve-se percorrer as regras da LFG formatando-as de acordo com a forma normal de Greibach;
- Os termos das regras que não são núcleos (*heads*) nem símbolos pré-terminais são mapeados diretamente em GPSG. Caso o termo possua algum esquema, restrição e/ou condição que corresponda a alguma *feature* da GPSG, essa *feature* deve ser associada ao termo na GPSG. Por exemplo, um esquema associado a um termo na LFG indicando número como plural (↓Num=plural) poderia ser mapeado para a *feature* [+PLU] em GPSG;
- Os termos das regras que são núcleos (*heads*) e não são símbolos pré-terminais são mapeados no núcleo H, na GPSG. Caso o termo possua algum esquema, restrição e/ou condição que corresponda a alguma *feature* da GPSG, essa *feature* deve ser associada a H na GPSG;
- Os termos das regras que são símbolos pré-terminais (podendo também ser núcleos) devem ser mapeados na letra H juntamente com a *feature* SUBCAT e um índice indicando o tipo do item lexical (substantivo, verbo transitivo direto, verbo transitivo indireto, verbo bitransitivo, verbo intransitivo, verbo de ligação, ..., adjetivo e preposição). Cada tipo deve ter um único índice associado, por exemplo, caso existam dois itens lexicais do tipo substantivo, ambos devem ser associados ao mesmo índice. Caso o termo possua algum esquema, restrição e/ou condição que corresponda a alguma *feature* da GPSG, essa *feature* deve ser associada ao termo na GPSG;
- Deve-se percorrer as regras da LFG procurando relações de precedência que sejam válidas, ou seja, que não sejam contrariadas posteriormente. Com isso, monta-se as regras de precedência da GPSG;
- Por último, as regras que introduzem itens lexicais devem dar origem a modificações no léxico da GPSG, pela inclusão de suas respectivas informações. O léxico deve conter também a classe de palavras à qual o item pertence, as formas amorfas (caso sejam especificadas no léxico da LFG) e a informação semântica. Caso a classe à qual o item lexical pertença seja substantivo, verbo, adjetivo ou preposição, a classe do léxico deve ser substituída pelas *feature*s que a representam, e, além disso, caso o símbolo pré-terminal e/ou o próprio item lexical da LFG possua algum esquema, restrição e/ou condição que corresponda a alguma *feature* da GPSG, essa *feature* deve ser associada ao item lexical da GPSG. Por exemplo:

LFG: deu: verbo, (↑Número)=singular
(↑Pred)='deu<(↑Suj)(↑ObjD)(↑Prep ObjI)>'

GPSG: < deu, [[-N][+V][BAR 0][-PLU][SUBCAT ...]], {}, **deu'** >

Para a redução de uma gramática GPSG a uma gramática LFG equivalente, os princípios acima podem ser aplicados de forma reversa. Por exemplo:

- As regras disjuntas são mapeadas em uma única regra através do uso de parênteses e/ou chaves. Por exemplo:

GPSG: sn → determinante, substantivo
sn → nomepróprio

LFG: sn → { determinante substantivo
nomepróprio }

- Os termos das regras GPSG que não são núcleos nem símbolos pré-terminais são mapeados diretamente na representação em LFG. Caso o termo possua alguma *feature* que indique dependência contextual, essa *feature* deve ser mapeada para o esquema correspondente em LFG. Por exemplo:

  GPSG: sentença → **sn**, sv

  LFG: sentença → **sn** sv
  (↑Suj)=↓

- Os termos das regras que são núcleos e/ou itens pré-terminais são mapeados em DCG na forma ordinária, isto é, em forma de categorias gramaticais. Caso o termo possua alguma feature que indique dependência contextual, essa feature deve ser mapeada para o esquema correspondente em LFG. Por exemplo:

  GPSG: sn → determinante, $\mathbf{H^1}$

  LFG: sn → determinante **substantivo**

- Cada entrada do léxico da GPSG é mapeada para sua correspondente em LFG, sendo as categorias dos itens lexicais e os esquemas associados a cada entrada do léxico da LFG correspondentes às *features* que os indicam, caso elas estejam presentes no léxico da GPSG. Por exemplo:

  GPSG: < deu, [[-N][+V][BAR 0][-PLU][SUBCAT ...]], {}, **deu'** >

  LFG: deu: verbo, (↑Número)=singular
  (↑Pred)='deu<(↑Suj)(↑ObjD)(↑Prep ObjI)>'

onde:
- as *features* [-N] e [+V] correspondem à classe gramatical verbo;
- a *feature* [-PLU] da GPSG corresponde ao esquema (↑Número)=singular na LFG;
- a *feature* SUBCAT indica o tipo do item lexical, neste caso, verbo bitransitivo, através do qual se obtém a forma semântica correspondente na LFG, ou seja, (↑Pred)='deu<(↑Suj)(↑ObjD)(↑Prep ObjI)>'.

- Percorre-se as regras de precedência da GPSG e, segundo estas, formata-se as regras geradas em LFG, de modo similar ao realizado no caso da DCG.

5. *Um Sistema de Redução DCG-LFG-GPSG*

A fim de ilustrar a possibilidade de redução entre os formalismos sob investigação, foi construído um sistema de redução DCG-LFG e DCG-GPSG. Na verdade, a primeira etapa de construção da interface, já apresentada em (Pardo, 1999), foi estendida para permitir também a redução DCG-GPSG. O funcionamento básico da interface não foi alterado nessa extensão, logo, detalhes podem ser encontrados em (Pardo, 1999).

Aproveita-se assim, uma construção gramatical simples em DCG como entrada do sistema de redução para gerar suas formas correspondentes ora em LFG, ora em GPSG, sendo que ambientes separados podem ser vistos simultaneamente durante a interação.

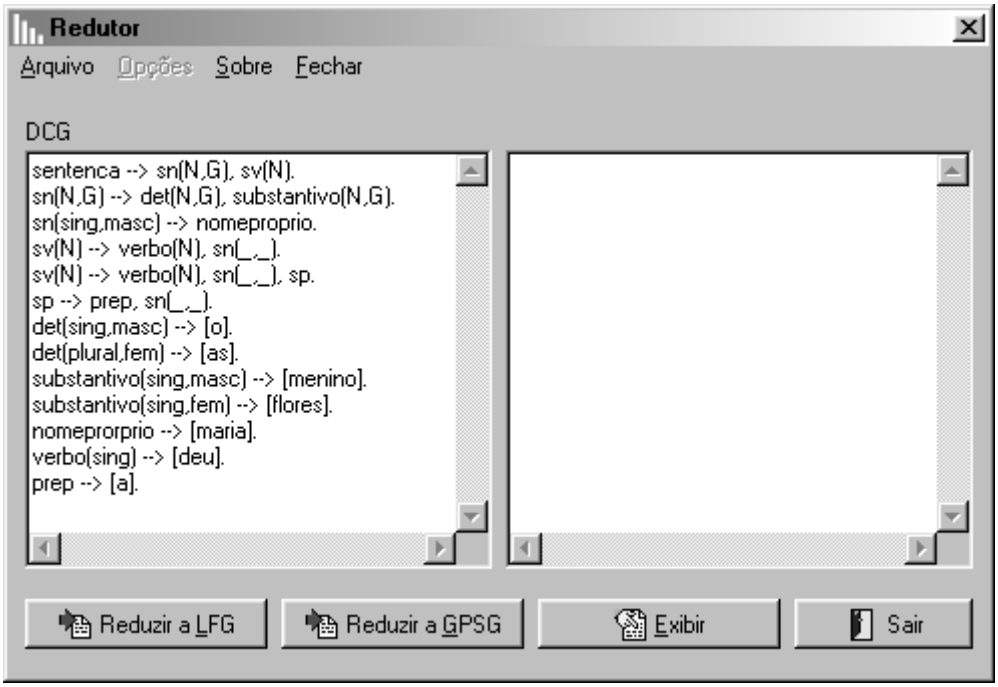
Reduções que envolvem um grau de complexidade maior, por exemplo, dependências de longa distância, não foram contempladas pelo sistema. Para tanto, seria necessário investigar a correspondência das gramáticas de extraposição (Pereira, 1981) com a LFG e a GPSG.

O sistema de redução apresentado nesta seção se destina ao usuário especialista ou leigo, tendo o especialista a função de detalhar especificações DCG, caso queira uma redução completa em LFG ou GPSG. Caso não haja detalhamento suficiente, o sistema produzirá somente reduções parciais, tomando por base o máximo número de informações coerentes que ele consiga extrair da DCG. Informações incoerentes consistem, p.ex., da ocorrência única de um argumento na DCG (i.e., somente de um lado de uma regra de produção), que será considerada incoerente para a geração de regras na LFG ou GPSG e, portanto, não implicará qualquer alteração nas especificações formais em construção.

As subseções seguintes apresentam um exemplo detalhado da execução do sistema, mediante *dumpings* de tela.

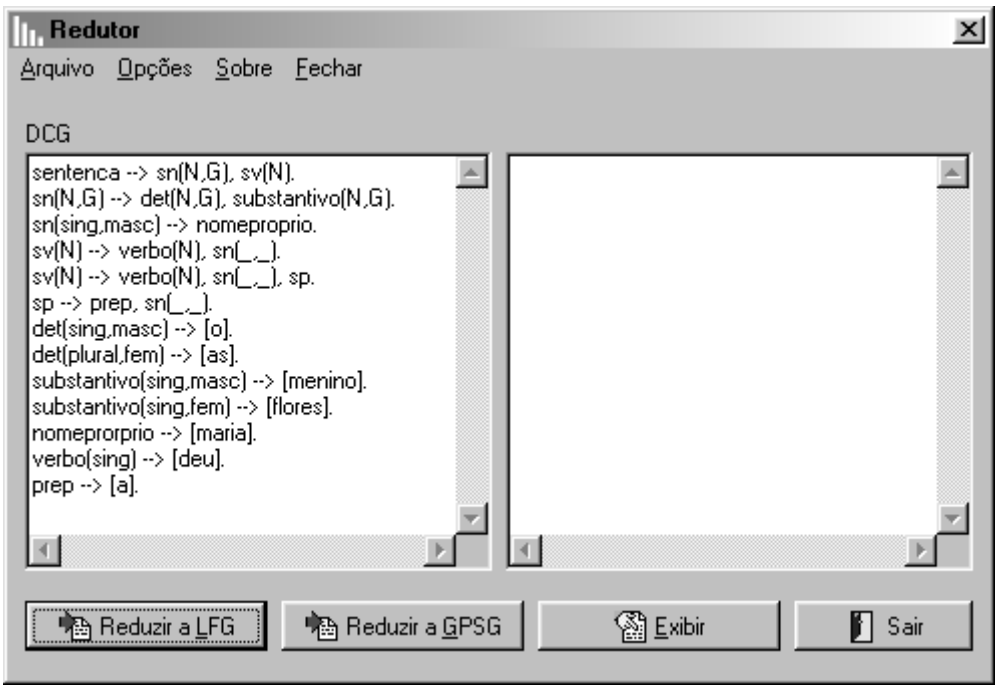
5.1. Procedimentos para redução DCG-LFG

<u>Primeiro passo:</u> especificação da DCG, podendo ser tipada pelo usuário ou carregada de um arquivo.



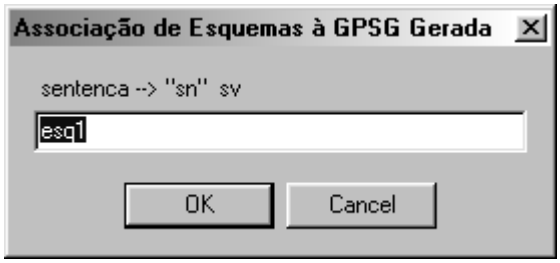
Janela 1 – Especificação da DCG

Segundo passo: ao se selecionar o botão “Reduzir a LFG”, se a gramática em DCG estiver especificada de forma correta, o botão “Exibir” e o menu suspenso “Opções” são habilitados.



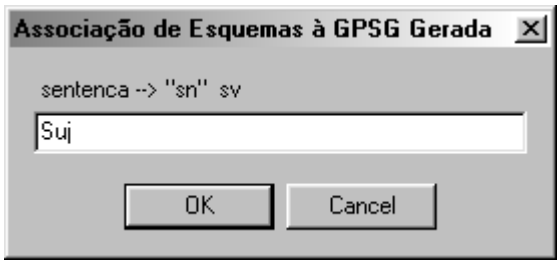
Janela 2 – Verificação da DCG e Redução à LFG

Terceiro passo: o programa associa automaticamente esquemas funcionais padrões a seus respectivos termos durante a montagem da LFG, porém, caso se queira alterar esses esquemas, basta selecionar o menu suspenso “Opções”. A partir daí, o sistema exibe um esquema por vez indicando a regra e o termo correspondente da LFG aos quais o esquema está associado. Por exemplo, para a regra indicada na Janela 3 abaixo, o programa indica que esq1 é a estrutura padrão associada ao termo entre aspas (sn) da regra introduzida por sentença.

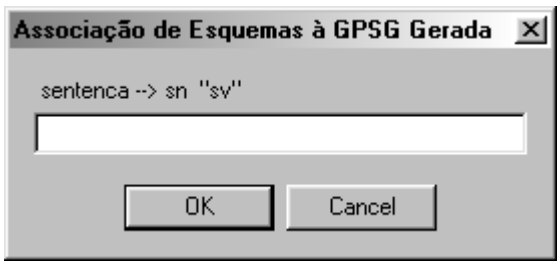


Janela3 – Especificação de Esquemas Padrões

Para alterar tal esquema, basta introduzir o novo esquema, como indica a Janela 4, e assim, o esquema asssociado a sn (indicado pelas aspas) passa a ser Suj. A seguir deseja-se que o sv seja inserido diretamente na estrutura como um todo. Neste caso, não é associado um esquema a ele, como indica a Janela 5. Dessa forma, o usuário pode especificar todos os esquemas conforme deseje.



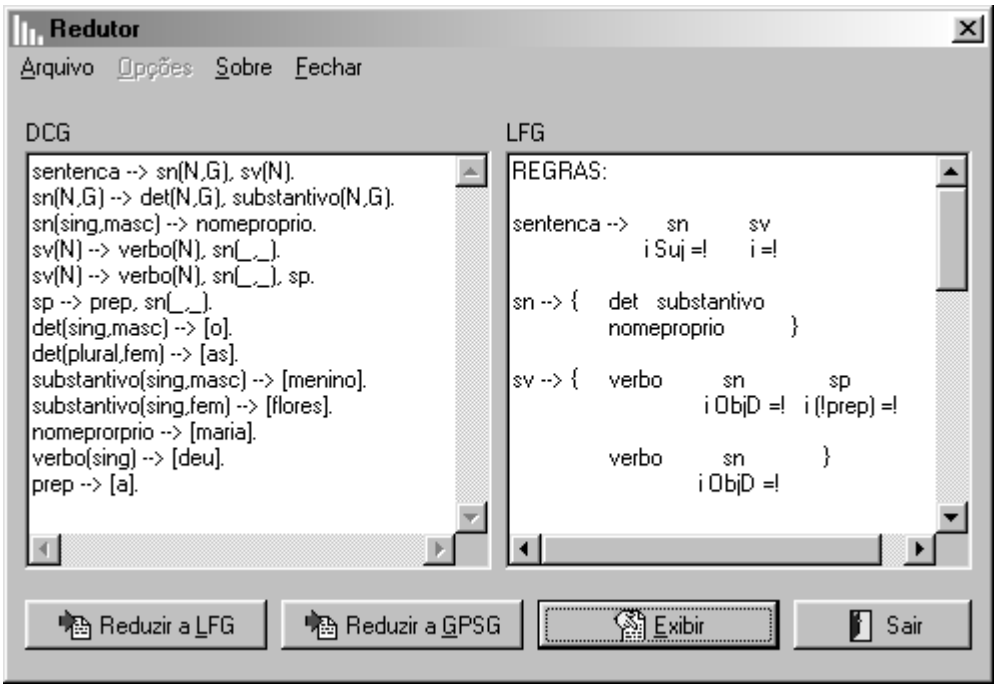
Janela 4 – Especificação do Esquema de sn



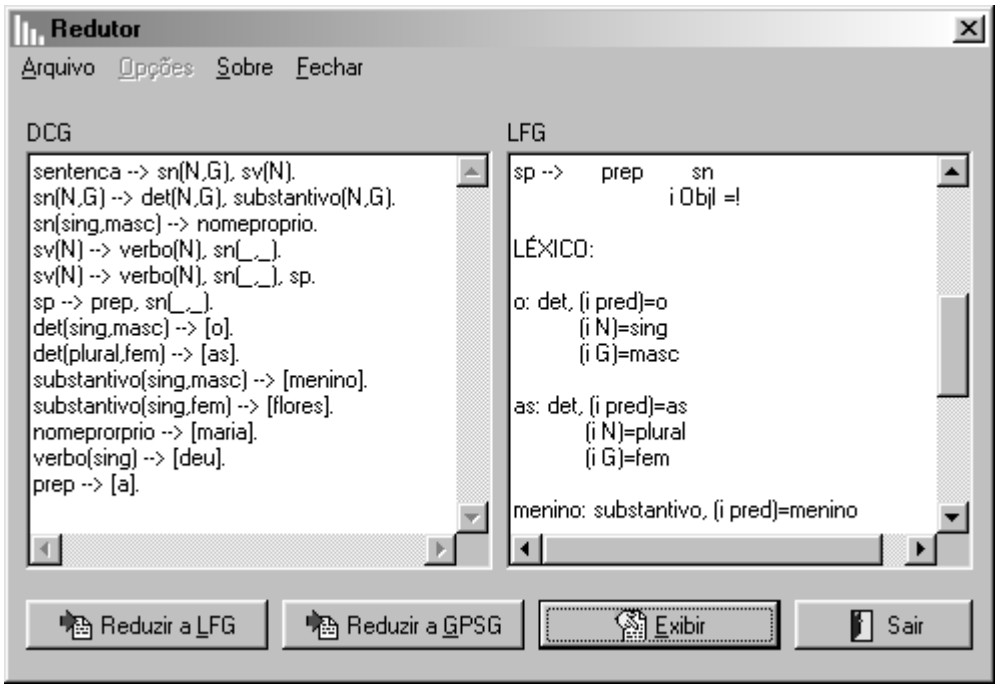
Janela 5 – Especificação do Esquema de sv

Esta etapa opcional destina-se ao usuário especialista que deseja uma redução completa. A geração da gramática em LFG correspondente à DCG inicial (Janela 1) é, então, realizada, procedendo-se ao passo 4.

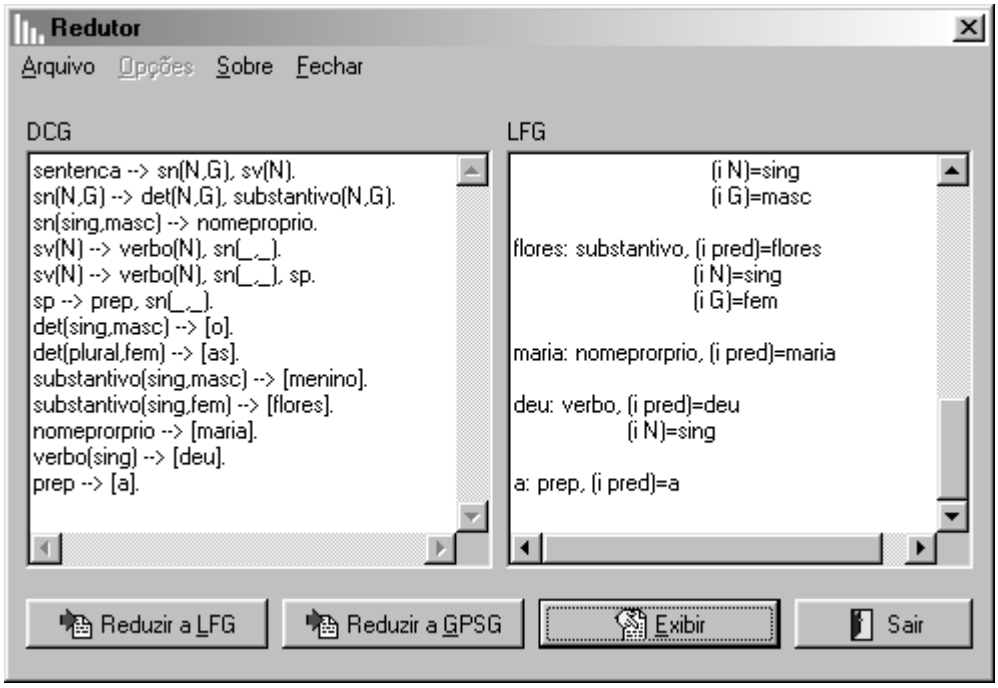
Quarto passo: ao se selecionar o botão "Exibir", a LFG gerada é exibida na tela. O conteúdo desta tela pode ser maior do que o exibido e, portanto, a tela pode ser rolada para baixo, como mostram as Janelas 6, 7 e 8. Vale notar que os símbolos "↑" e "↓", nessas ilustrações, foram substituídos por "i" e "!", respectivamente, pois a linguagem utilzada (Delphi 3.0 Client/Server) não provê os símbolos originais dos formalismos.



Janela 6 – Exibição da LFG Correspondente



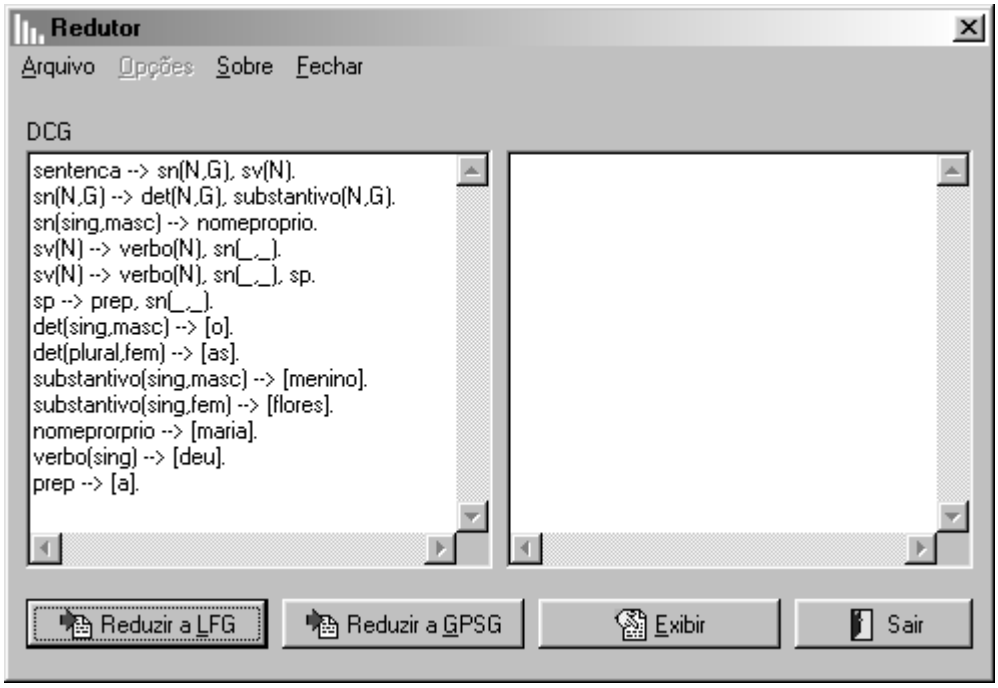
Janela 7 - Exibição da LFG Correspondente



Janela 8 - Exibição da LFG Correspondente

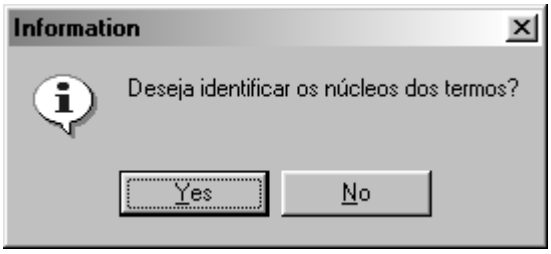
5.2. Procedimentos para redução DCG-GPSG

Primeiro passo: especificação da DCG, podendo ser tipada pelo usuário ou carregada de um arquivo.



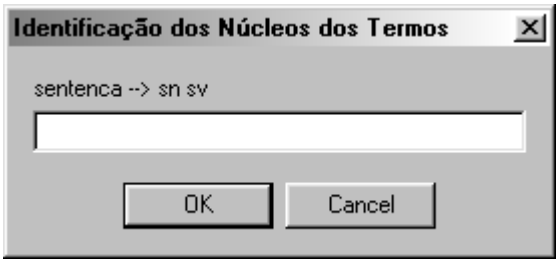
Janela 9 – Especificação da DCG

Segundo passo: ao se selecionar o botão "Reduzir a GPSG", se a gramática em DCG estiver especificada de forma correta, o botão "Exibir" será habilitado. A seguir, telas para identificação dos núcles (*heads*) das regras serão exibidas. A Janela 10 exibe a tela de consulta ao usuário.



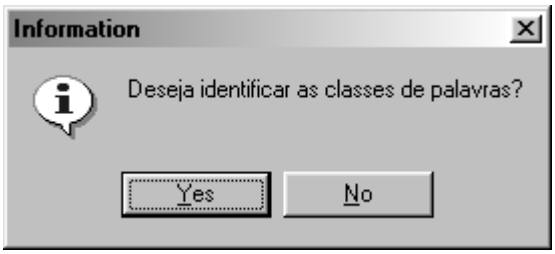
Janela 10 – Identificação dos Núcleos dos Termos

Caso o usuário seja leigo, ele pode selecionar o botão "No" e cancelar a opção. Caso contrário, o sistema começará a perguntar os núcles de cada regra, janela a janela. A Janela 11 mostra um exemplo de uma tela de identificação de núcleo, onde o usuário respondeu com sv.



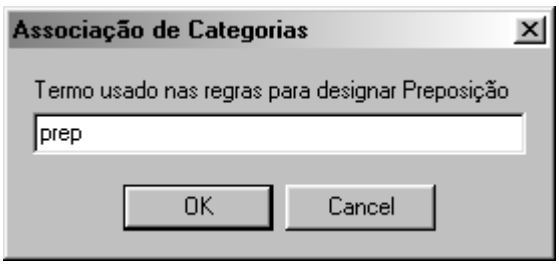
Janela 11 – Especificação do Núcleo de Sentença

Terceiro passo: o sistema perguntará ao usuário se ele deseja identificar as classes de palavras que ele utilizou na gramática DCG. A Janela 12 mostra a tela de consulta.



Janela 12 – Identificação das Classes de Palavras

Caso o usuário seja leigo, ele pode selecionar o botão "No" e cancelar a opção. Caso contrário, o sistema começará a perguntar os classes das palavras, janela a janela. A Janela 13 mostra um exemplo de uma tela de identificação das classes, onde o usuário respondeu que a palavra utilizada por ele para representar a classe das preposições foi prep.

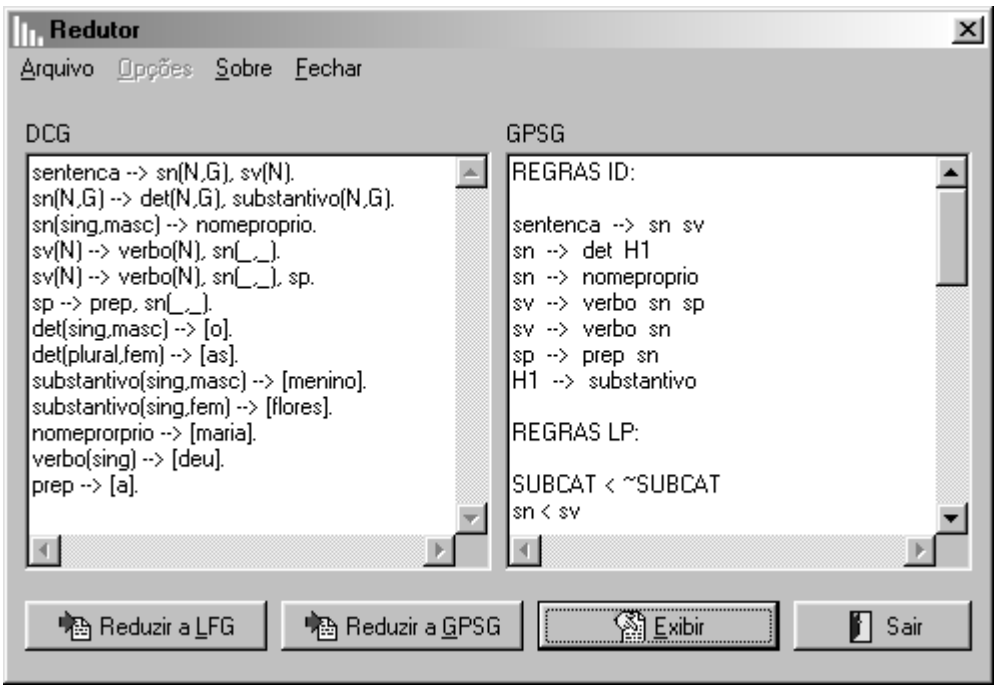


Janela 13 – Identificação da Classe das Preposições

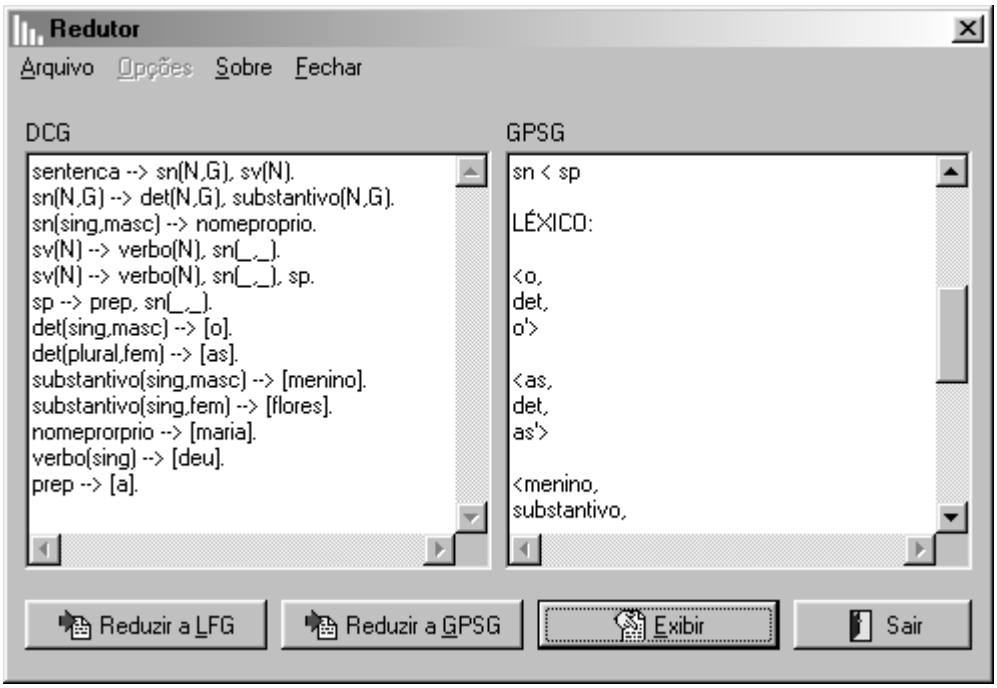
A geração da gramática em GPSG correspondente à DCG inicial (Janela 9) pode ser realizada, procedendo-se ao passo 4.

Quarto passo: ao se selecionar o botão "Exibir", a GPSG gerada é exibida na tela.

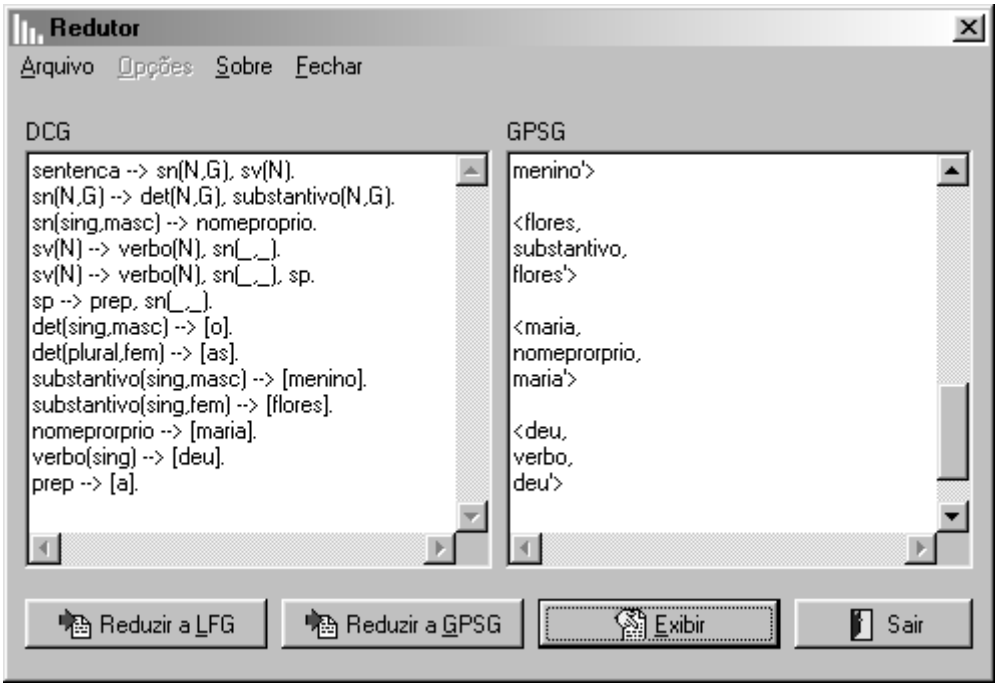
As Janelas 14, 15, 16 e 17 exibem a GPSG gerada por um usuário leigo, ou seja, que não especificou os núcleos e as classes de palavras.



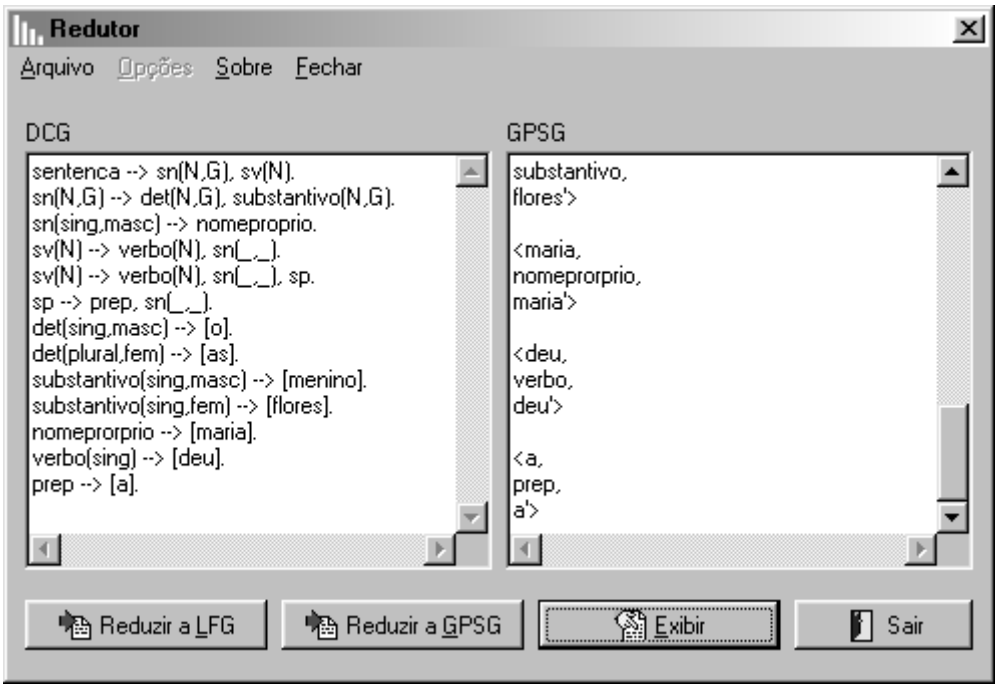
Janela 14 – Especificação da GPSG Correspondente sem Definições de Núcleo e Classe



Janela 15 - Especificação da GPSG Correspondente sem Definições de Núcleo e Classe

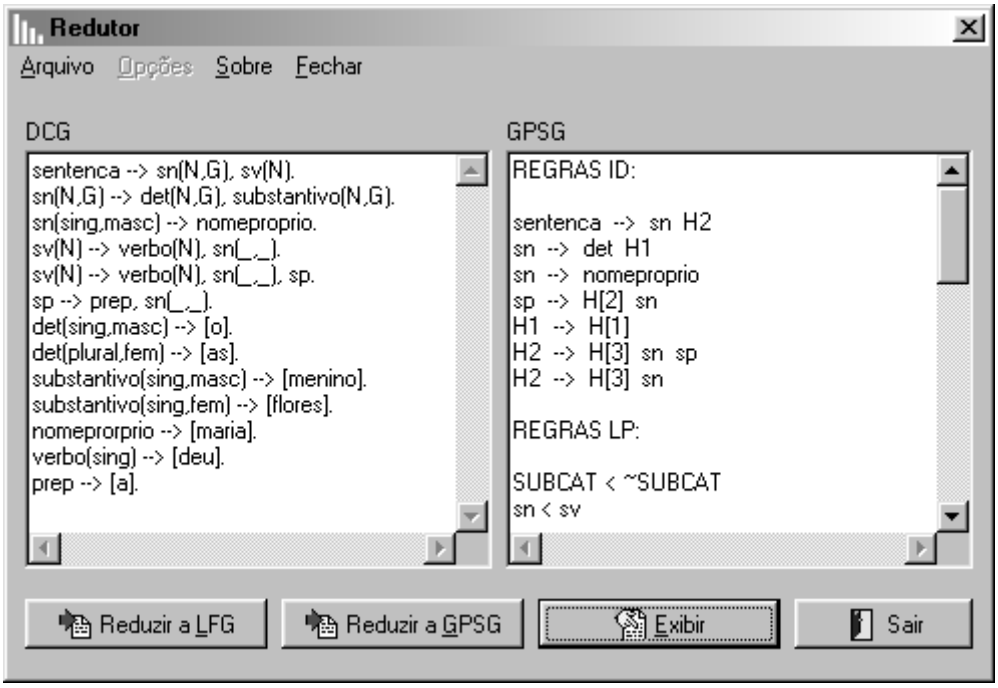


Janela 16 - Especificação da GPSG Correspondente sem Definições de Núcleo e Classe

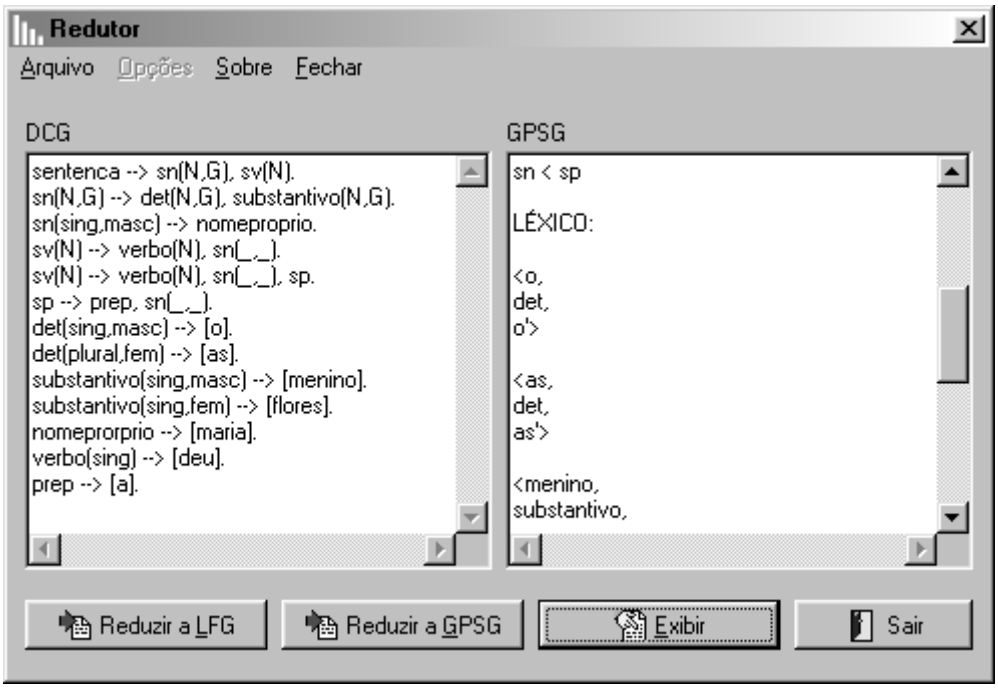


Janela 17 - Especificação da GPSG Correspondente sem Definições de Núcleo e Classe

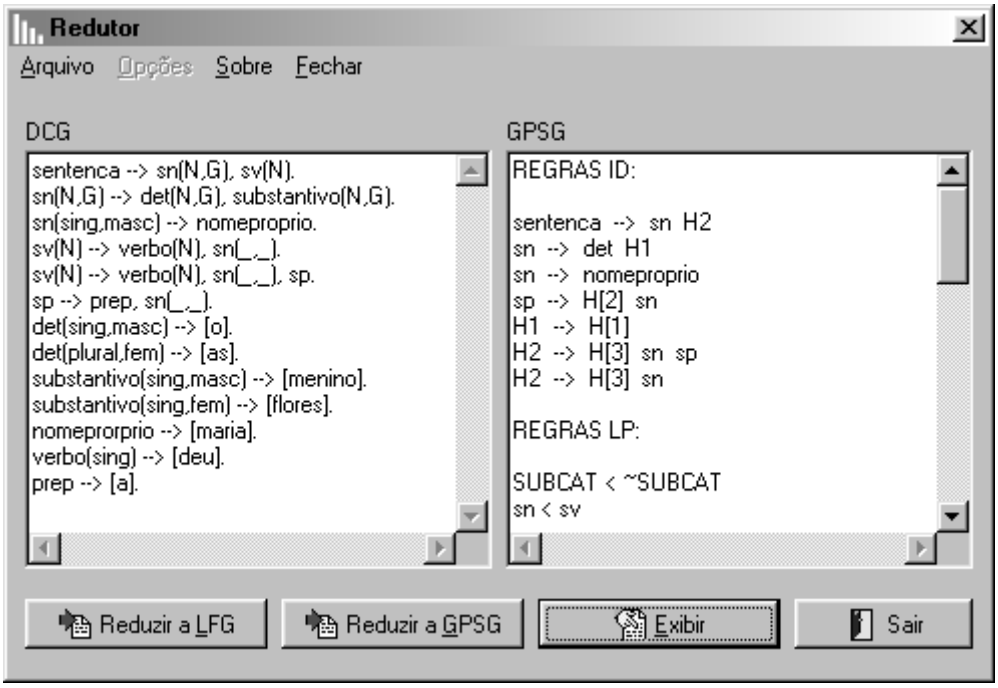
As Janelas 18, 19, 20 e 21 exibem a GPSG gerada por um usuário que especificou somente os núcleos das regras, como pode ser visto nas Janelas 10 e 11.



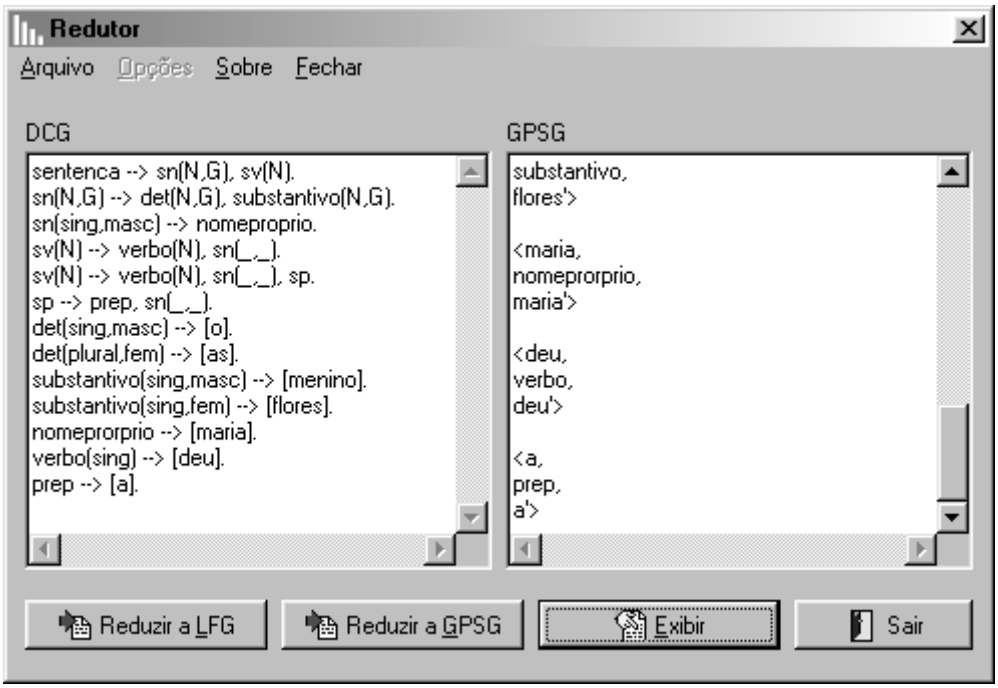
Janela 18 - Especificação da GPSG Correspondente Definições de Classe



Janela 19 - Especificação da GPSG Correspondente Definições de Classe

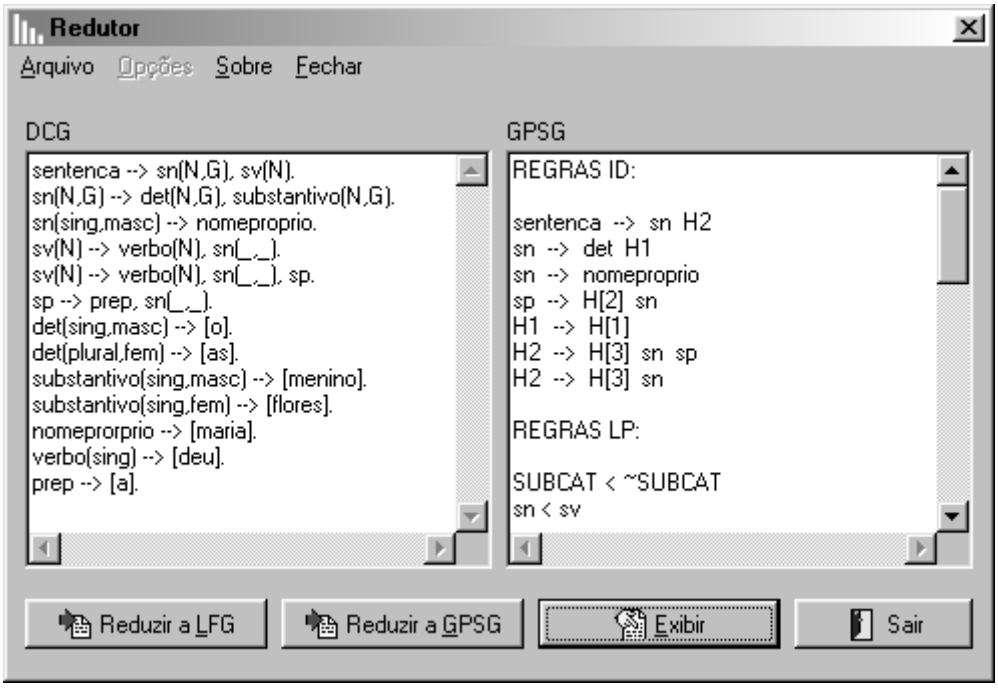


Janela 20 - Especificação da GPSG Correspondente Definições de Classe

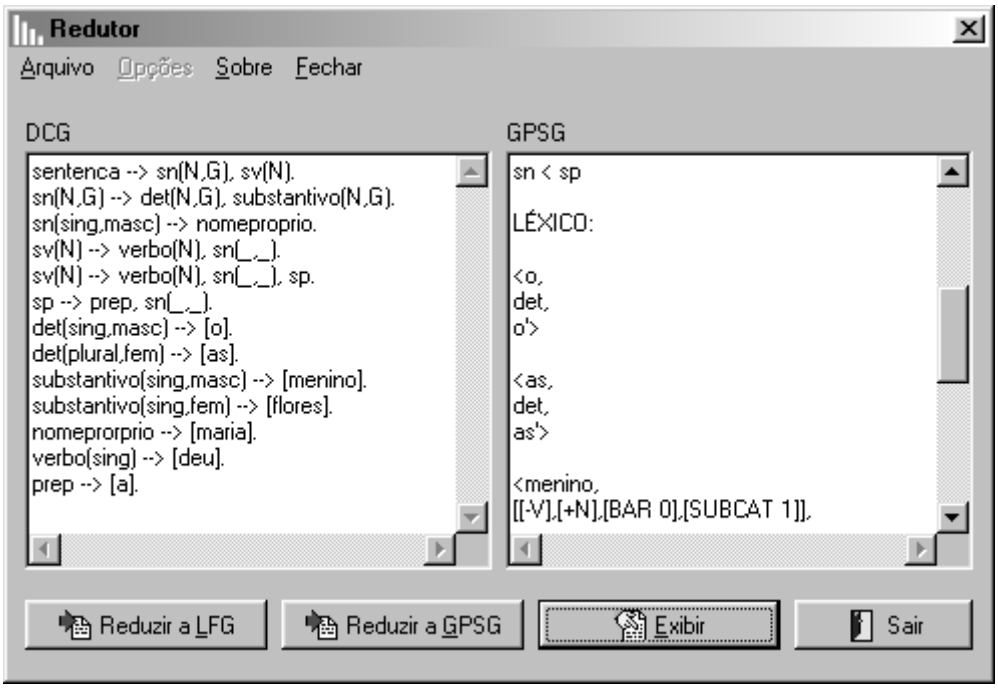


Janela 21 - Especificação da GPSG Correspondente Definições de Classe

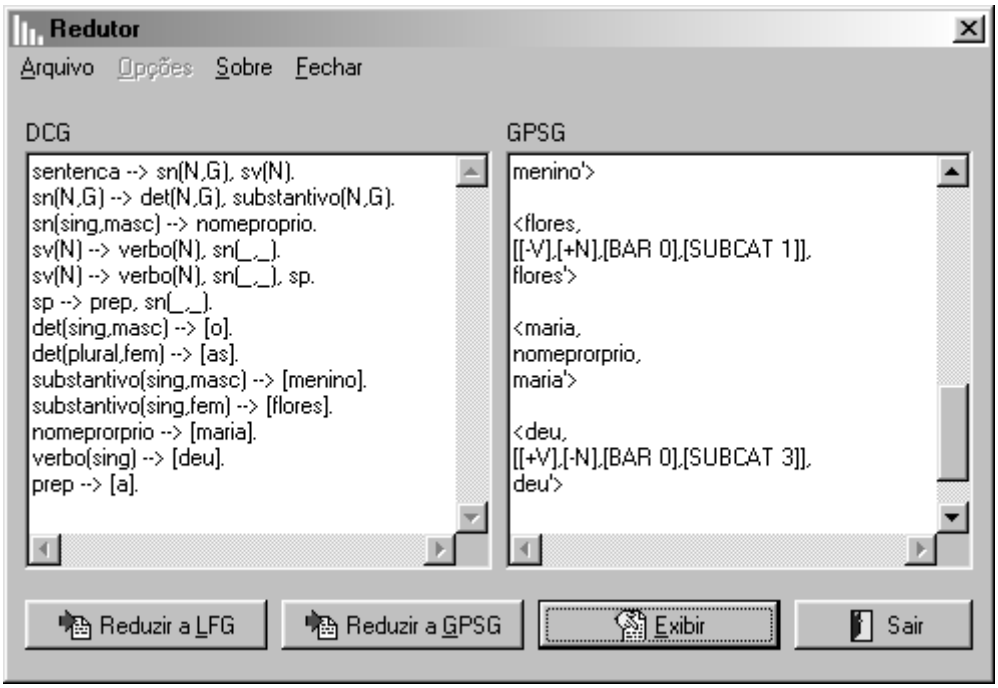
As Janelas 22, 23, 24 e 25 exibem a GPSG gerada por um usuário que especificou tanto os núcleos das regras como as classes das palavras, como pode ser visto nas Janelas 12 e 13.



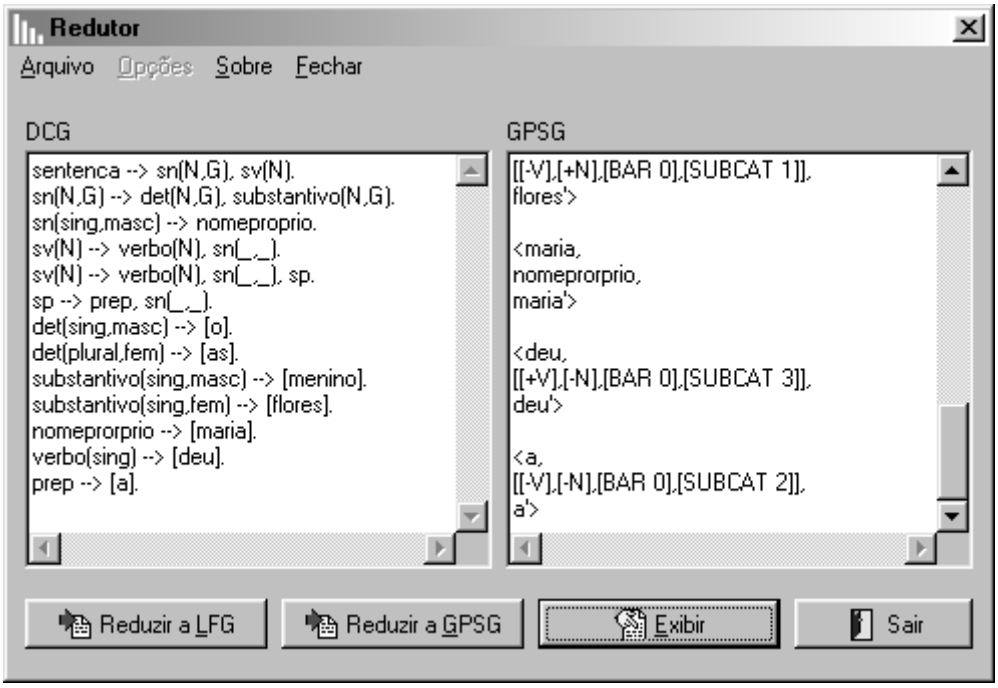
Janela 22 - Especificação da GPSG Correspondente com Definições de Núcleo e Classe



Janela 23 - Especificação da GPSG Correspondente com Definições de Núcleo e Classe



Janela 24 - Especificação da GPSG Correspondente com Definições de Núcleo e Classe



Janela 25 - Especificação da GPSG Correspondente com Definições de Núcleo e Classe

Devido à alta abstração da GPSG, a ferramenta adapta a gramática GPSG reduzida a partir da DCG ou da LFG da seguinte forma:

- A Teoria X-Bar é desconsiderada para que a gramática, caso necessário, possa ser interpretada por um *parser* (analisador);
- Pela falta de um método para identificação do tipo do verbo sem que o usuário tenha que interceder, todos os verbos são associados à mesma subcategorização;
- A representação de morfologias irregulares no léxico é desconsiderada, devido aos problemas de reconhecimento da morfologia nos argumentos da DCG.

6. *Conclusão*

Foi apresentada neste relatório uma descrição das principais características do formalismo gramatical GPSG, com o objetivo de comparar seu potencial descritivo/gerativo com o potencial dos formalismos DCG e LFG. Além disso, foi apresentada uma ferramenta para redução automática entre DCG e GPSG e entre DCG e LFG, sendo que esta última redução já havia sido amplamente explorada em (Pardo, 1999).

A redução entre os formalismos DCG, LFG e GPSG mostra-se possível e automatizável para construções simples como as ilustradas neste relatório. A completeza das reduções, principalmente a redução DCG-GPSG, é determinada, em parte, pelas informações adicionais fornecidas pelo usuário especialista, que tem a função de suprir o grau de refinamento necessário na GPSG, ausente na DCG, para construir a especificação formal correspondente. Caso o especialista não intervenha, a interface fornecerá, na maioria das vezes, uma redução DCG-GPSG incompleta. Já as reduções DCG-LFG são, em sua maioria, completas, dado que os esquemas para a LFG são produzidos seguindo um padrão predefinido, caso não haja informações suficientes na DCG correspondente.

Reduções que envolvem um grau de complexidade maior, por exemplo, dependências de longa distância, ou referências anafóricas, não foram contempladas neste trabalho. Para tanto, se for considerado o próprio formalismo DCG, seria necessário, por exemplo, investigar a correspondência das gramáticas de extraposição (Pereira, 1981) com a LFG e a GPSG, no que diz respeito à sua dependência notacional e conceitual.

A ferramenta de redução apresentada, além de servir ao propósito de verificação da possibilidade de redução DCG-LFG ou DCG-GPSG, pode ainda ser explorada para outros tipos de correspondência. Em particular, para a redução no sentido inverso ao apresentado neste relatório, isto é, LFG-DCG ou GPSG-DCG ou para a redução GPSG-LFG (ou vice-versa). A expansão de tal interface para considerar ainda outros formalismos também apresenta uma perspectiva interessante, por exemplo, considerando-se FUGs (*Functional Unification Grammars*) (Kay, 1983) ou PATR-II (Shieber, 1984), conforme discutido em (Shieber, 1987), assim como a utilização da interface para fins didáticos, como a introdução de lingüistas e cientistas da computação aos aspectos formais da Engenharia das Línguas Naturais.

Referências Bibliográficas

Chomsky, N. (1970). Remarks on nominalization. In Rodrerick A. Jacobs and Peter S. Rosenbaum (eds), *Readings in English Transformational Grammar*, pp. 184-221. Walthan, Massachusetts: Ginn and Company.

Gazdar, G.; Klein, E.; Pullum, G.; Sag, I. (1985). *Generalized Phrase Structure Grammar*. Harvard University Press. Cambridge, Massachusetts.

Harman, G. (1963). Generative grammars without transformational rules: a defense of phrase structure. *Language*, pp. 597-616.

Jackobson, Roman, C.; Fant, G. M.; Halle, M.(1951). *Preliminaries to Speech Analysis*. Cambridge, Massachusetts: MIT Press.

Kaplan, R. M.; Bresnan, J. (1982). Lexical-Functional Grammar: A Formal System for Grammatical Representation. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*. Cambridge, MS, MIT Press.

Kay, M. (1983). *Unification Grammar*. Technical Report, Xerox Palo Alto Research Center, Palo Alto CA.

Keenan, E. L. (1974). The functional principle: generalizing the notion of ‘subject of’. In Michael W. La Galy, Robert A. Fox and Anthony Bruck (eds), *Papers from the 10th Regional Meetinig of the Chicago Linguistics Society*, pp. 298-309.

Montague, R. (1974). *Formal Philosophy*: New Haven: Yale University Press.

Pardo, T.A.S. (1999). *Estudo comparativo dos formalismos gramaticais DCG e LFG*. Relatório Técnico RT-DC 001/99, Departamento de Computação, Universidade Federal de São Carlos.

Pereira, F.C.N. (1981). Extraposition Grammars. *AJCL*, Vol. 7, N. 4, Oct-Dec.

Pereira, F.C.N.; Warren, D.H.D. (1980). Definite Clause Grammars for Language Analysis – A Survey of the Formalism and a Comparison with Augmented Transition Networks. *Artificial Intelligence*, 13, pp. 231-278.

Pereira, F.C.N. (1981). Extraposition Grammars. *AJCL*, Vol. 7, N. 4, Oct-Dec.

Shieber, S.M. (1984). The Design of a Computer Language for Linguistic Information In *Proc. of the 10th International Conference on Computational Linguistics / 22nd Annual Meeting of the Association for Computational Linguistics*, pp. 362-366. Standford, CA.

Shieber, S. M. (1986). *An Introduction to Unification-Based Approaches to Grammar*. University of Chicago Press, USA.

Shieber, S.M. (1987). Separating Linguistic Analyses from Linguistic Theories In P. Whitelock; M.M. Wood; H.L. Somers; R. Johnson and P. Bennett (eds.) (1987), *Linguistic Theory & Computer Applications*, pp. 1-36. Academic Press Inc., San Diego, CA.