

# PUC

---

Série: Monografias em Ciência da Computação,  
No. 15/90

O COMPONENTE DE REALIZAÇÃO NA GERAÇÃO AUTOMÁTICA DE TEXTOS:  
EFEITOS DA CONCILIAÇÃO

Clarisse S. Souza  
Maria das Graças V. Nunes

Departamento de Informática

---

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO  
RUA MARQUÊS DE SÃO VICENTE, 225 - CEP-22453  
RIO DE JANEIRO - BRASIL

PUC/RJ - DEPARTAMENTO DE INFORMÁTICA

Series: Monografias em Ciência da Computação, 15 /90

Editor: Paulo Augusto Silva Veloso

Outubro, 1990

O COMPONENTE DE REALIZAÇÃO NA GERAÇÃO AUTOMÁTICA DE  
TEXTOS: EFEITOS DA CONCILIAÇÃO

Clarisse S. Souza

Maria das Graças V. Nunes

Trabalho parcialmente financiado pela FINEP

**Responsável por publicações:**

Rosane Teles Lins Castilho  
Assessoria de Biblioteca, Documentação e Informação  
PUC RIO, Departamento de Informática  
Rua Marquês de São Vicente, 225 - Gávea  
22453 - Rio de Janeiro, RJ  
BRASIL

Tel.: (021) 529-9386

TELEX: 31078

FAX: (021) 274-4546

BITNET: [rserrtlc@lncc.bitnet](mailto:rserrtlc@lncc.bitnet)

# O Componente de Realização na Geração Automática de Textos:

## Efeitos da Conciliação.

*Maria das Graças Volpe Nunes e Clarisse Sieckenius de Souza*

Departamento de Informática - PUC/RJ

Rua Marquês de São Vicente 225

22453 - Rio de Janeiro/RJ

### *ABSTRACT*

A geração automática de textos de boa qualidade exige que, durante a fase de planejamento da estrutura dos mesmos, se levem em consideração fatores cognitivos ligados a estruturas linguísticas utilizadas na redação final. O *planejamento conciliatório* é uma proposta de modelo computacional para geradores de textos descritivos, permitindo que o componente de *realização* de tais geradores atinja melhores resultados em termos de seu funcionamento e produto, quando comparado aos de outras propostas. O presente trabalho delinea e discute as principais características de tal componente.

**Palavras-chave:** Geração de Linguagem Natural, Realização de Texto

## 1. INTRODUÇÃO.

O processo tradicional de geração automática de texto envolve duas fases conceitualmente distintas: o planejamento e a realização. Enquanto que ao planejamento cabe selecionar a mensagem a ser veiculada, à realização cabe escolher a maneira adequada de apresentá-la. Grande parte dos sistemas existentes (v. [1] e [2]) utiliza o modelo sequencial de geração, no qual a realização é feita após o planejamento total do texto. A estrutura estritamente sequencial, embora computacionalmente eficiente, resente-se das consequências de um planejamento que não leva em conta aspectos linguísticos em pontos cruciais da estruturação do texto como, por exemplo, na determinação da ordem de apresentação das informações selecionadas. Com isto, a qualidade do texto final é sensivelmente prejudicada. Quando, porém, o texto tem possibilidades de ser melhor escrito, através do uso de modelos não-sequenciais como o integrado [3] e o intercalado [4], ou se tem uma sobrecarga da interface planejador-realizador, ou uma considerável superposição de critérios e imbricação de estruturas de conhecimento. Ambas alternativas acarretam desvantagens computacionais.

A tentativa de manter a estrutura sequencial, eliminando-se as desvantagens acima, resultou num modelo conciliatório [5], onde planejador e realizador, embora sequenciais, compartilham conhecimento cognitivo e linguístico. São cognitivos os critérios relativos a seleção, a ordem de apresentação e ao foco da informação contida no texto. São linguísticos os critérios relativos a composição de cada parágrafo, em termos da estrutura final sugerida para as várias sentenças veiculadoras do conteúdo informacional. Como se pode ver, este mecanismo promove considerável penetração da atividade de planejamento no âmbito linguístico, constituindo um valioso look-ahead. O presente trabalho delinea a filosofia geral de operação do realizador para o planejamento conciliatório, respaldado em resultados atuais do sistema GEMA, um gerador de descrições de Diagramas de Fluxos de Dados [6].

O realizador opera em duas etapas : a da geração de sentenças básicas a partir dos predicados da representação do objeto a ser descrito e a da geração de sentenças transformadas. Elas refletem a estratégia de "refinamentos sucessivos" no processo de geração, preservando o sentido inicial planejado, através de transformações mantenedoras de significado. As sentenças básicas assemelham-se as

sentenças nucleares das primeiras propostas chomskyanas para uma gramática gerativa, isto é: tem uma sintaxe elementar do tipo sujeito-verbo-complemento. As sentenças transformadas resultam, como o diz o nome, da aplicação de regras transformacionais sobre as primeiras. Apesar da clara influência chomskiana [7,8] no modelo, o realizador não deve ser tomado como uma implementação do modelo gerativo-transformacional.

Todo texto descritivo é presumidamente composto de um parágrafo introdutório que enumera os componentes essenciais do objeto da descrição e de parágrafos sucessivos que detalham cada componente e seus respectivos sub-componentes. Para o GEMA, cada processo do DFD em questão dá origem a um parágrafo e o corpo descritivo é precedido por uma introdução padronizada. A descrição obviamente não pode contar com semântica de domínio (do sistema modelado), sob pena de impor ao gerador a condição de incorporar um "modelo de mundo". Isto significa que, para o fragmento de DFD da Figura 1, o sistema jamais será capaz de inferir que se trata de um sistema destinado a gerenciar um hotel (o que é evidente para seres humanos). A semântica disponível é essencialmente aquela contida na metodologia de modelagem, acrescida de "primitivas" julgadas indispensáveis para possibilitar um mínimo de penetração nas atividades dos processos do sistema. Exemplos de tais primitivas serão encontrados no decorrer deste trabalho.

## **2. O REALIZADOR.**

As estruturas linguísticas utilizadas pelo planejador objetivam sinalizar o foco, ordenar as informações no parágrafo e obter uma boa qualidade estilística do texto. No primeiro caso, para marcar o foco de elementos, são utilizadas estruturas linguísticas onde aparecem deslocamentos. Essencialmente, trata-se do deslocamento para frente (evidenciação) de um complemento verbal não-preposicionado, através da voz passiva, ou preposicionado, através da topicalização. Quando o foco e o sujeito de uma sentença, claramente não há necessidade de deslocamentos. A contrapartida do foco - ou seja, o fato de um elemento não estar em evidência - tende a ser sinalizada por estruturas encaixadas, através do emprego de orações relativas, onde figuram tais elementos. No segundo caso, para dar conta da ordenação das informações no parágrafo, são empregadas a justaposição e a coordenação de sintagmas ou sentenças. As regras de coordenação controlam a escolha entre períodos justapostos

ou coordenados, tanto quanto as regras de encaixe controlam a escolha da relativização. Finalmente, no terceiro caso, motivações essencialmente estilísticas levam a inclusão de estruturas onde se apresentam anáforas, elípses e nominalizações. Isto contribui consideravelmente para a qualidade do texto. Vale dizer que tais estruturas linguísticas não são as únicas utilizáveis para marcar foco, ordenação e estilo. Porém, são estruturas que certamente podem desempenhar tal papel.

A estrutura do parágrafo é, de fato, o território de conciliação entre a informação que se quer veicular e a linguagem que se deve usar. Para tratar da questão do foco, em sua contrapartida linguística, há princípios como os exemplificados a seguir.

- Cada informação só é realizada uma vez no parágrafo.
- O volume de uma informação, se alto, (predicado com alta cardinalidade) pode gerar sub-focos de informação, cujos predicados são realizados em bloco no início do parágrafo.
- A informação que está em foco deve preceder aquela que não está em foco, em um mesmo nível hierárquico de sentença.
- Uma informação que não está em foco deve ser apresentada preferencialmente adjunta (através de encaixe) ao elemento essencial a que ela se reporta.
- Objetos ou processos cuja descrição se faça através de períodos compostos devem observar as seguintes heurísticas:
  - (a) o encaixe de orações dentro de outras orações só pode ocorrer se a estrutura resultante não apresentar encaixes recursivos.
  - (b) o número máximo de encaixes por período é três.

O realizador deve operar linguisticamente respeitando uma ordem de aplicação de regras. A necessidade desta ordenação surge de fatos óbvios como a impossibilidade de se encaixar um predicado em outro, por exemplo, depois de o candidato ao encaixe já ter sido coordenado a outros de mesma natureza. A consequência disto seria carregar para a posição de encaixe mais do que a informação pertinente a operação. Por isto, identificam-se operações sensíveis a foco e operações não-sensíveis a foco.

A geração de sentenças básicas é feita através de uma Gramática de Unificação de Base Lexical em que a cada atividade corresponde um verbo, cuja entrada no dicionário possui os seguintes componentes: morfologia, semântica, sintaxe e restrições. O campo sintaxe expressa uma regra gerativa para o predicado contido no campo semântica, através de um processo de unificação. Na Figura 2, vê-se exemplo de uma entrada lexical do GEMA.

A filosofia desta etapa de geração é tomar cada predicado da representação de conhecimento proposta pelo planejador e transformá-lo em uma sentença básica, com sintaxe fixa do tipo: SUJEITO, VERBO, OBJETO-DIRETO, COMPLEMENTO-PREPOSICIONADO. Este padrão pode em princípio ser alterado, para servir com maior conveniência a variadas aplicações. Entretanto, do ponto de vista teórico, para efeitos de um modelo computacional de geração de texto, não parece haver fortes indicações de que tais alterações venham a ser necessárias.

O Dicionário tem um valor crucial para o processo de geração. Ele é composto de duas partes distintas: uma parte estática e uma parte dinâmica. A parte estática do Dicionário inclui todas as lexicalizações dos verbos correspondentes a primitivas semânticas, de palavras relativas a metodologia de modelagem dos sistemas em questão e de palavras marcadoras de estruturas sintáticas da língua. A parte dinâmica inclui os nomes dos componentes do objeto de entrada e mais as nominalizações referentes aos processos, obtidas por um processador morfológico acoplado a interface.

A geração de sentenças transformadas - de fato, períodos transformados - tem como entrada as sentenças básicas. Estruturado o parágrafo, o gerador de sentenças transformadas aplica operações de coordenação, relativização, topicalização, passivização, anaforização, elipse e nominalização. A aplicação das regras transformacionais tem como ponto crítico o escopo e a ordem. Para controlar o escopo, o planejador é incumbido de retratar na estrutura de parágrafo os blocos de sentenças básicas que devem ser submetidas simultaneamente ao gerador transformacional. Este nível de escopo é o primeiro e mais genérico. O segundo restringe o anterior, visando atingir o alvo exato para as transformações. Ele resulta da avaliação fina de parâmetros compartilhados entre sentenças candidatas, controlada por um conjunto de matrizes. Estas matrizes especificam padrões gerais de compartilhamento de traços sintáticos e/ou semânticos entre sentenças candidatas. O controle da ordem, em contrapartida,



resulta da interação de dois fatores: a precedência das regras transformacionais umas em relação às outras e o percurso de alto a baixo, da esquerda para a direita, da própria estrutura de parágrafo.

A coordenação é um caso especialmente interessante para mostrar o efeito do planejamento conciliatório e o tipo de controles necessários. Uma observação mais detalhada de seu tratamento ilustra os principais pontos abordados neste trabalho. Um exemplo extraído do GEMA apoia a apresentação.

A decisão final sobre que sentenças básicas coordenar é atingida pelo exame das sentenças candidatas (i.e. dentro de um mesmo bloco), a luz de matrizes que refletem o seguinte princípio:

- As coordenações operam simultaneamente sobre categorias semânticas e sintáticas. A identificação de uma oportunidade sintática para a coordenação é acompanhada da leitura do campo RESTRIÇÕES das matrizes, onde se observam semelhanças e dissemelhanças semânticas entre os referentes dos termos.

- As coordenações entre elementos sintáticos apresentam uma hierarquia de complexidade que é: coord-SN < coord-SPrep < coord-SV < coord-S. Ou seja, quanto menor a complexidade dos constituintes a serem coordenados, tanto menor o grau de complexidade da própria coordenação, tanto mais leve o texto resultante da coordenação.

- As coordenações do tipo coord-SV e coord-S são sensíveis a foco, na medida em que operações de deslocamento (sinalizadoras de foco) que envolvam SNs ou SPreps dominados por SVs ou Ss coordenados não podem ser aplicadas. Seus operandos encontram-se fixados dentro de uma estrutura da qual não podem mais ser extraídos.

As matrizes presumem que sua aplicação se dá a nível de cada parágrafo. Com isto, o sujeito das sentenças básicas é sempre idêntico (i.e. o processo) e o que resta avaliar são semelhanças e dissemelhanças dentro do sintagma verbal (envolvendo verbo, objeto direto e complemento preposicionado). Note-se que, pela filosofia de geração de sentenças básicas, para cada -act, SNs e SPreps compartilham o mesmo tipo semântico (e.g. entrada, saída, destino, origem), salvo exceção assinalada no campo RESTRIÇÃO. A configuração das matrizes é a seguinte:

VERBO    SN    SPrep

-----  
IGUAIS

A    C    E

-----  
DIFERENTES

B    D    F

-----  
RESTRIÇÃO

-    -    -

-----  
Cada matriz cria uma série de padrões para a avaliação de sentenças básicas candidatas a coordenação. São eles:

ACE - este caso não causa coordenação porque trata-se de duplicatas

ACF - este caso aponta coordenação de SPreps

ADE - este caso aponta coordenação de SNs

ADF - este caso aponta coordenação de SNs e SPreps

BCE - este caso aponta coordenação dos verbos

BCF - este caso aponta coordenação em torno do SN, envolvendo o verbo e o SPrep

BDE - este caso aponta a coordenação em torno de SPrep, envolvendo verbo e SN

BDF - este caso aponta uma coordenação de Ss

De imediato, os padrões ACF e ADE são identificados como livres de foco, podendo e devendo ser submetidos a transformações antes das regras de foco se aplicarem. Da mesma forma, o padrão ACE é identificado como inoperante e o padrão BDF é identificado como sensível a foco, por definição. Adicionalmente, os padrões BDE, BCF e BCE são sensíveis a foco, do momento em que envolvem a coordenação a nível de constituintes de SV (tipo coord-SV). Por fim, o padrão ADF não mantém independentes SN e SPrep, com o que é sensível a foco. Isto mostra como as matrizes são ferramentas poderosas para o controle de escopo no caso da coordenação.

No que tange a ordem de aplicação das regras, por motivos de manutenção das estruturas mínimas de entrada para a série de transformações, ela é a seguinte:

- coordenação livre de foco
- topicalização e passivização
- relativização
- coordenação sensível a foco
- elipse
- anáfora e nominalização

Para a coordenação sensível a foco, é necessário que já se tenham feito deslocamentos e encaixes e que tais operações estruturais estejam sinalizadas na nova observação da matriz. Desta vez, o campo RESTRIÇÕES estará assinalando as estruturas correntes, as quais poderão influenciar a aplicação da regra de coordenação. Adicionalmente, para o padrão BCE - que mantém invariantes SN e SPrep e coordena verbos - um teste adicional de compatibilidade da preposição deve ser realizado, para filtrar estruturas resultantes estranhas ao português.

### **3. UM REALIZADOR EM FUNCIONAMENTO.**

Para exemplificar a atuação do realizador, no âmbito do GEMA, vale esclarecer a terminologia usada. P-acts são primitivas semânticas relativas as atividades principais do processo, ou seja, o que é feito de uma entrada para resultar numa saída. R-acts e S-acts são primitivas relativas a tráfego de entrada e saída, respectivamente.

A Figura 3 mostra a estrutura de um parágrafo gerada pelo planejador, relativo a um certo processo P1.

**P-acts:**

- (1) P-act1(P1, Registro, Nome, Endereço)
- (2) P-act2(P1, Número de Associados por Bairro, Nome, Endereço)

**R-acts:**

- (1) R-act1(P1, Nome, Cliente)
- (2) R-act1(P1, Endereço, Cliente)

Os R-acts1 (1) e (2) acima correspondem a P-act1 e P-act2.

S-acts:

- (1) S-act1(P1, Registro, Sócios)
- (2) S-act1(P1, Registro, Convidados)
- (3) S-act2(P1, Número de Associados por Bairro, Gerente)

O S-acts1 (1) e (2) acima correspondem a P-act1 e o S-act2 (3) a P-act2.

Para gerar as sentenças básicas, na primeira fase da realização, o sistema consulta o léxico via o índice de cada -act. No exemplo, R-act1 leva a entrada do verbo "receber", P-act1, ao verbo "compor", P-act2, "calcular", e assim por diante. As sentenças básicas geradas a partir dessas consultas, aqui retratadas não na forma anotada em que efetivamente se representam, mas de forma estilizada, são:

R-acts:

- 1) P1 recebe Nome de Cliente.
- 2) P1 recebe Endereço de Cliente.

S-acts:

- 1) P1 acrescenta Registro em Sócios.
- 2) P1 acrescenta Registro em Convidados.
- 3) P1 manda Número de Associados por Bairro para Gerente.

P-acts:

- 1) P1 compõe Registro com Nome e Endereço.
- 2) P1 calcula o Número de Associados por Bairro com Nome e Endereço.

Consultando cada uma das matrizes, dado um -act, obtemos os seguintes padrões:

Matriz P-act : não é consultada pois cada bloco só tem um p-act. (v. Figura 3).

Matriz R-act : ADE - coordenação de SNs, ou seja, entre Nome e Endereço.

Matriz S-act : ACF - coordenação de SPreps, ou seja, entre Sócios e Convidados.

Aplicadas as operações de coordenação, o realizador produz, em cada bloco, as operações de topicalização, resultando em estruturas intermediárias equivalentes a:

(a) no primeiro bloco:

Com Nome e Endereço, P1 compõe Registro. (para P-act)

P1 acrescenta Registro em Sócios e Convidados. (para S-act)

P1 recebe Nome e Endereço de Cliente. (para R-act)

(b) e no segundo:

Com Nome e Endereço, P1 calcula Número de Associados por Bairro. (para P-act)

P1 envia Número de Associados por Bairro para Gerente. (para S-act)

P1 recebe Nome e Endereço de Cliente. (para R-act)

Aplicando as operações de passivização e relativização, o realizador produz estruturas intermediárias equivalentes a:

(a) no primeiro bloco:

Com Nome e Endereço, que são recebidos de Clientes por P1, P1 compõe Registro, que é acrescentados em Sócios e Convidados por P1.

(b) e no segundo:

Com Nome e Endereço, que são recebidos de Clientes por P1, P1 calcula Número de Associados por Bairro, que é enviado para Gerente por P1.

É interessante observar, no funcionamento do realizador, o porquê das coordenações se dividirem em livres-de-foco e sensíveis-a-foco. É clara a vantagem da coordenação já efetuada por ocasião da topicalização e da relativização que a ela se seguem.

Uma etapa final de operação efetua as transformações de anáfora (pronominal e não-pronominal) e elipse, resultando em um texto do seguinte tipo:

- Com Nome e Endereço, recebidos de Clientes, P1 compõe Registro, que é acrescentado em Sócios e Convidados. Com a mesma entrada, ele calcula o Número de Associados por Bairro, que é enviado para Gerente.

Nesta etapa são considerados elementos estilísticos finos para decidir que estruturas de superfície empregar. Por exemplo, uma regra estilística relativa a elipse determinou que a primeira oração relativa do texto se tenha atualizado em sua forma reduzida. Outra regra estilística relativa a anáforas determinou que o pronome "eles" não tenha sido empregado para remeter a "Nome e Endereço", face a potencial ambigüidade com o sintagma nominal mais próximo ("Sócios e Convidados"). Ao invés disto, ela determinou o emprego de outro tipo de anáfora ("a mesma entrada"). Outras regras - aqui não exemplificadas - introduzem refinamentos como a expressão "arquivo" antes da atualização textual de depósitos de dados e artigos definidos antes da atualização textual de entidades externas. O resultado final é um texto como:

- Com Nome e Endereço, recebidos dos Clientes, P1 compõe Registro, que é acrescentado nos arquivos Sócios e Convidados. Com a mesma entrada, ele calcula o Número de Associados por Bairro, que é enviado para Gerente.

#### 4. DISCUSSÃO.

O GEMA é o primeiro de uma série de protótipos de geração de linguagem [9,10] em que a influência do planejamento conciliatório é clara. Ele se distingue dos demais pela precariedade da informação semântica disponível. O texto por ele gerado revela, com isto, os limites da "semântica contida na sintaxe", uma vez que que a interligação das duas categorias é amplamente explorada no processo global de produção. Assim, a saída do GEMA revela um caso evidente de tratamento de linguagem natural como código, e não como comportamento (a sua real natureza).

A proposta de realização apresentada neste trabalho tem significativas vantagens do ponto de vista do texto gerado e do ponto de vista computacional. Em primeiro lugar, a abordagem conciliatória pos-

sibilita ao realizador ser um componente de "decisão", e não mero executor de uma estrutura totalmente especificada, sem ser demasiadamente complexo. Comparado a realizadores de outras abordagens, o realizador conciliatório apresenta duas grandes vantagens: não precisa determinar o escopo de sentenças, grave problema para alguns geradores [4], e opera com transformações cognitivamente motivadas, o que assegura a legibilidade final do texto [11].

A realização em dois passos distintos resulta em vantagens computacionais interessantes. Primeiramente, contribui para maior flexibilidade na especificação do gerador, visto que na etapa de geração de sentenças básicas podem-se empregar estruturas e procedimentos totalmente distintos daqueles empregados na etapa de transformações. Em segundo lugar, como a etapa de geração de sentenças básicas e a que preserva a manutenção do significado pretendido na mensagem, qualquer processo de backtracking necessário retorna a um ponto posterior a ela. Isto diminui o custo de tais retrocessos, tendo adicionalmente sólida motivação teórica. Em terceiro lugar, o controle da aplicação das regras fica significativamente mais simples do que seria caso contrário. Esta característica pode ser mais claramente avaliada quando se percebe que em nossa proposta há um patamar linguístico comum a todas as estruturas (as sentenças básicas) antes de elas começarem a diferenciar-se através de estruturas encaixadas ou coordenadas, que as podem tornar extremamente complexas. Começar de um patamar não-linguístico e efetuar toda a realização em uma só etapa implica prever todos os passos de possíveis operações linguísticas aplicáveis (esgotar todas as possibilidades) sem dispor de um mínimo de estrutura sintática que possa reduzir o elenco de opções a seguir. O uso de matrizes, exemplificado no caso da coordenação, mostra o efeito desta redução.

Dentre os desafios que esta abordagem precisa enfrentar, contudo, figuram dois de maior relevância: o tratamento de sinônimos e a flexibilidade de gerar mais de um texto para descrever o mesmo objeto. Ambos remetem a pontos de rigidez da abordagem conciliatória, tal como implementada no GEMA.

A disponibilidade de sinônimos é extremamente importante em qualquer sistema de geração de texto. Os motivos para empregá-los são variados. Dentre eles, há o de substituição de uma palavra por seu sinônimo devido a ela estar sendo empregada com muita frequência. Num processamento automá-

tico, decidir o que é "muita frequência" já é por si só um problema complexo. Entretanto, resolvida a questão, quando o gerador chega ao nível de escolha lexical, um grande número de decisões sintáticas já foram tomadas e a estrutura global onde a palavra vai ser empregada está fortemente definida. A sinonímia semântica, como se sabe, não implica as mesmas características sintáticas: o exemplo mais simples é o de verbos sinônimos que tem transitividade diferente. Ora, trocar um destes verbos pelo outro significa refazer a estrutura sintática onde ele ocorre (que pode extrapolar o nível de sentença). O controle deste retrocesso é consideravelmente complexo e a solução temporária do GEMA foi a de somente adotar sinônimos com restrições semânticas e sintáticas equivalentes na aplicação. Isto, porém, claramente requer revisão.

Quanto a possibilidade de o gerador produzir textos diferentes com um mesmo conteúdo, trata-se de um problema mais teórico do que prático. Do ponto de vista prático, se um sistema sabe qual é o melhor texto, ele não deve produzir um pior. No entanto, do ponto de vista teórico, o planejamento conciliatório tal como no GEMA só pode apresentar variações a nível lexical, excluindo o nível sintático e o nível de discurso. Isto se deve ao fato de que a estrutura planejada funciona como um esquema [1] na etapa de realização e o sistema não pode raciocinar sobre o porque de sua escolha. Isto certamente é uma grande restrição do "modelo" de produção de texto como um todo. Entretanto, este problema foi tratado posteriormente [11] e resolvido à luz de uma "gramática de texto", que faculta a variação a nível de discurso e de sintaxe.

Neste trabalho abordaram-se apenas questões gerais, enfatizando os efeitos da filosofia conciliatória para o realizador de um sistema de geração automática de texto. As questões mais específicas, sobretudo em termos das consequências computacionais específicas da adoção destas idéias, são objeto de estudo mais aprofundado em fase de desenvolvimento [12].



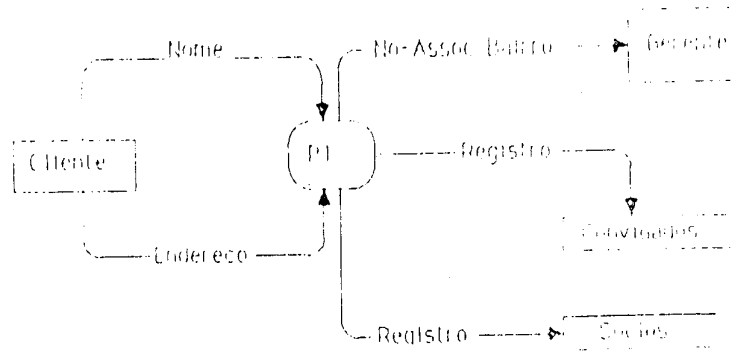


Figura 1 - Fragmento de PFD

Ent. lex - ACRESCENTAR

Restrições  
Morfologia  
Semântica

s-act1, Agente, Saída, Destino

Sintaxe

Agente	<i>acrescentar</i>	Saída	em - destino
Suj	Verbo	Obj Dir	Comp/Prep
SN	Presente	SN	SPrep.
Nun.	Indicativo	Nun.	Num.
Gen.	Singular	Gen.	
3a p.	3a p.		

Figura 2 - Exemplo de Entrada Lexical

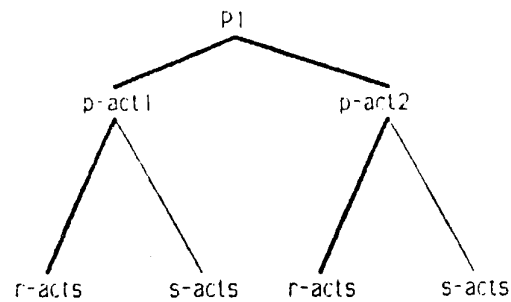


Figura 3 - Exemplo de estrutura de Parágrafo

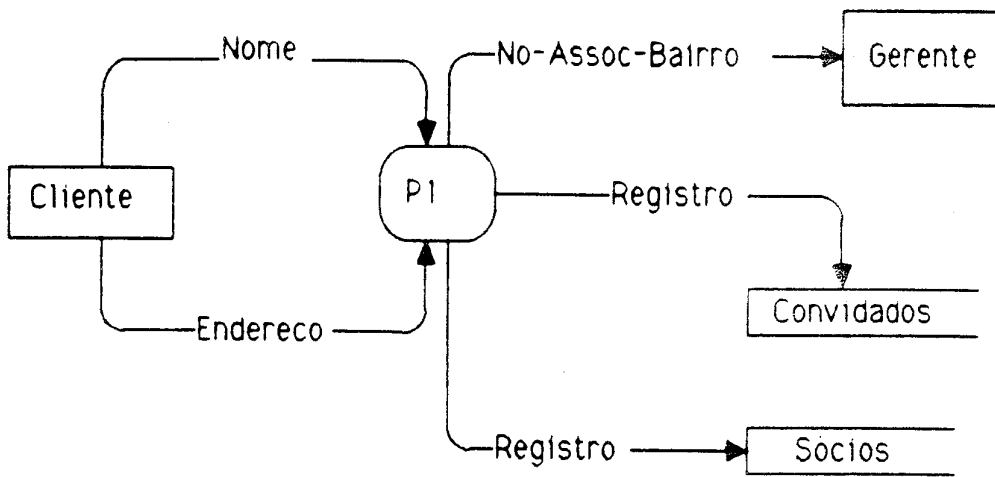


Figura 1 - Fragmento de DFD

Ent. lex. - ACRESCENTAR

Restricoes			
Morfologia			
Semantica			
s-act1, Agente, Saida, Destino			
Sintaxe			
Agente	<i>acrescentar</i>	Saida	<i>em</i> - Destino
Suj.	Verbo	Obj.Dir.	Comp.Prep.
SN	Presente	SN	SPrep.
Num.	Indicativo	Num.	Num.
Gen.	Singular	Gen.	
3a. p.	3a. p.		

Figura 2 - Exemplo de Entrada Lexical

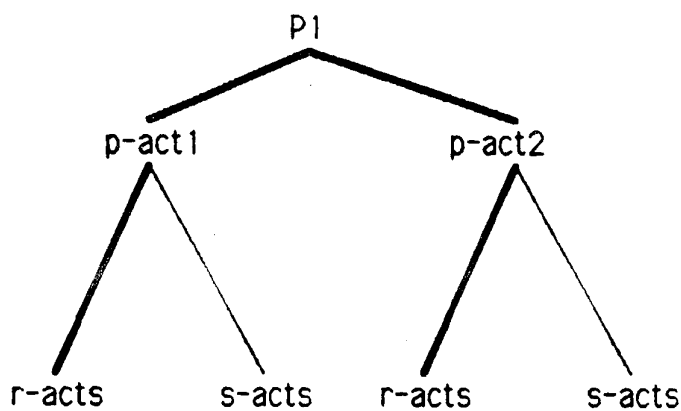


Figura 3 - Exemplo de estrutura de Paragrafo

## 5. REFERÊNCIAS

- [1] McKeown, K., *Generating Natural Language Text in Response to Question about Database Structure*. Ph.D. dissertation, University of Pennsylvania, 1982.
- [2] McDonald, D.D. & J.D. Pustejovsky. *Description- Directed Natural Language Generation*. Proceedings of the 9th. IJCAI, Vol.2, 1985.
- [3] Appelt, D.E., *Telegram : A Grammar Formalism for Language Planning*. Proceedings of the Eighth IJCAI Conference, Karlsruhe, 1983.
- [4] Hovy, E.H., *Two Types of Planning in Language Generation*. Proceedings of the 26th Meeting of the ACL, Buffalo, New York, 1988.
- [5] Scott, D.R. & C.S. Souza. *Conciliatory Planning for Extended Descriptive Texts*. Séries Monografias em Ciências da Computação, no. 13/88. Dep. Informática, PUC/RJ. 1988.
- [6] GEMA: Relatório de Pesquisa. Coletânea de Monografias a ser publicada. Dep. Informática, PUC/RJ. 1989.
- [7] Chomsky, N. *Syntactic Structures*. Mouton, 1957.
- [8] Chomsky, N. *Aspects of Theory of Syntax*. MIT Press, 1965.
- [9] Nunes, M.G.V. *Deep Generation in a Crime Knowledge- Based System*. Séries Monografias em Ciências da Computação, no. 10/88. Dep. Informática, PUC-RJ. 1988.
- [10] Nunes, M.G.V. *Um Gerador de Explicações Causais para Sistemas com Raciocinadores em Dedução Natural*. Monografia não publicada. Dep. Informática, PUC/RJ. 1989.
- [11] Scott, D.R. and Souza, C.S. *Getting the Message Across: A Computational Grammar of Text*. Technical Note in print. Philips Research Laboratories, Redhill, U.K. 1989.
- [12] Oliveira, L.L.G. *Implementação do Componente de Realização de um Gerador Automático de Textos*. Proposta de Tese. Dep. Informática, PUC/RJ. 1989.