



**cetep
sepes**

competências básicas do português

PRESIDENTE DA REPÚBLICA
Ernesto Geisel

MINISTRO DA EDUCAÇÃO E CULTURA
Ney Braga

PRESIDENTE DO MOBIL
Arlindo Lopes Corrêa

SECRETÁRIO EXECUTIVO DO MOBIL
Sérgio Marinho Barbosa

SECRETÁRIO EXECUTIVO ADJUNTO DO MOBIL
Odalêa Cleide Alves Ramos

COMPETÊNCIAS BÁSICAS DO PORTUGUÊS

Miriam Lemle

Anthony J.Naro

Relatório final da pesquisa apresentado
às instituições patrocinadoras Fundação
Movimento Brasileiro de Alfabetização
(Mobral) e Fundação Ford

Rio de Janeiro, maio de 1977

FICHA CATALOGRÁFICA

(Preparada pela Fundação Movimento Brasileiro de
Alfabetização - CETEP/SEDOC)

L554 Lemle, Miriam

Competências básicas do português por Mi
riam Lemle e Anthony J. Naro. Rio de Já
neiro, MOBREAL, Fundação FORD, 1977.

151 p. tab. 27 cm.

Bibliografia: p. 149-51

1. Língua portuguesa - gramática. Sinta-
xe. I. Naro, Anthony J. II. Fundação
Movimento Brasileiro de Alfabetização.
III. Fundação Ford. IV. Título.

77-95

cdd: 469.798

415.071

cdu: 469.415.6(022)

APRESENTAÇÃO

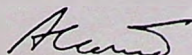
Entidade que procura assumir criticamente o seu trabalho, precisar seus objetivos e validar sua metodologia, o Movimento Brasileiro de Alfabetização está investindo recursos humanos e financeiros em pesquisa de campo, objetivando aumentar a eficácia de suas estratégias de trabalho e de todo o processo de ensino-aprendizagem.

Competências Básicas do Português, na medida em que oferece alternativas metodológicas e sugestões de elaboração de material didático que permitam acelerar o processo de leitura e a aquisição da escrita na norma culta, vem ao encontro dos objetivos do MOBRAL.

Partindo do pressuposto de que a língua nacional, homogênea, com uma gramaticalidade legítima, sobrepõem-se os falares regionais, e de que é indispensável a qualquer projeto de alfabetização partir da realidade lingüística, heterogênea e multiforme, da sua clientela, a pesquisa Competências Básicas do Português tentou configurar em alguns aspectos, que pareceram relevantes aos desígnios do MOBRAL e da equipe do projeto, a fala dos mobralenses do Rio de Janeiro.

À coordenadora Miriam Lemle e ao coordenador-associado Anthony Julius Naro foram confiados o projeto e a condução dos trabalhos, que atenderam perfeitamente à expectativa - e que esperamos possam ser replicados em outras cidades e em outras Unidades da Federação.

Ao MOBRAL, além do suporte teórico e prático que representa Competências Básicas do Português para os efeitos de seus projetos, cabe o orgulho de ter participado na tarefa de acumular estudos descritivos e comparativos dos diversos dialetos em que se manifesta a língua portuguesa.



ARLINDO LOPES CORRÊA
Presidente do MOBRAL

I N D I C E

Equipe de trabalho

Agradecimentos

LISTA DAS TABELAS

1 - CARACTERIZAÇÃO DA PESQUISA	1
1.1 - Objetivos	1
1.2 - Caracterização da amostra de língua falada	5
1.2.1 - Seleção dos informantes	5
1.2.2 - Esquema das entrevistas	5
1.2.3 - Caracterização dos informantes	7
2 - A REGRA DE CONCORDÂNCIA	17
2.1 - Fatores condicionantes	17
2.1.1 - Conceito de regra variável	17
2.1.2 - Descrição das variáveis	18
2.2 - Método de avaliação da interação dos fatores	25
2.3 - Programas utilizados	28
2.4 - Resultados	28
2.5 - Conclusões teóricas	40
2.6 - Conclusões práticas	50
3 - COMPLEXIDADE SINTÁTICA	51
3.1 - O problema da avaliação	51
3.1.1 - Fundamentação teórica	51
3.1.2 - A construção de uma medida de complexidade sintática	53
3.1.2.1 - Modelo de desempenho da integração sintática	53
3.1.2.2 - O método de quantificação	56
3.1.2.3 - Comprovação do método de quantificação	59
3.1.3 - Outras medidas de complexidade	79
3.2 - Caracterização da amostra	83
3.3 - Método de arborização	86
3.3.1 - O sintagma nominal	86
3.3.2 - O sintagma verbal	90
3.3.3 - O sintagma adjetivo	97
3.3.4 - O sintagma adverbial	100

3.3.5 - O sintagma preposicional	101
3.3.6 - Sentença	103
3.3.7 - A coordenação	108
3.4 - Operacionalização da quantificação: descrição do programa TREEBUILD (por Gregory R. Guy)	114
3.4.1 - Introdução	114
3.4.2 - Estrutura geral	115
3.4.3 - Sub-rotinas principais	117
3.4.4 - Uso	118
3.4.5 - Fluxograma geral	119
3.5 - Etapas analíticas	121
3.6 - Resultados	123
3.6.1 - Descrição dos cálculos	123
3.6.2 - Medidas obtidas	129
3.6.2.1 - Os sintagmas menores	138
3.6.2.2 - As sentenças principais	140
3.6.2.3 - Relação entre complexidade dos sintagmas e seu nível de encaixe	144
3.6.3 - Conclusões práticas	145
3.6.4 - Seguintos desejáveis da pesquisa	146
ABSTRACT	148
REFERÊNCIAS	149

EQUIPE DE TRABALHO

COORDENADOR: Miriam Lemle

COORDENADOR-ASSOCIADO: Anthony Julius Naro

ASSESSOR DE COMPUTAÇÃO: Gregory Guy

EQUIPE DE

ANÁLISE SINTÁTICA: Maria Angela Botelho Pereira
Maria Aparecida Botelho Pereira
Margarida Maria Martins Salomão
Maria Cecília de Jesus Gomes

EQUIPE DE

TABELAMENTO

DA CONCORDÂNCIA: Renata Gerard Bondim de Menezes
Maria Cândida Drummond Mendes de Barros
Ora Meisel
Alcides Cordeiro Lopes
Mauro Baltazar
Maria Marta Dias Pereira Scherre
Nelise Pires de Omena
Vera Maria de Baêre Naegeli

EQUIPE DE ENTREVISTAS: Renata Gerard Bondim de Menezes
Maria Cândida Drummond Mendes de Barros
Ora Meisel
Alcides Cordeiro Lopes
Mauro Baltazar
Maria Marta Dias Pereira Scherre
Nelise Pires de Omena
Sheila Maria Feital
Adauto de Souza Santos
Vania Lins
Maysa Ferreira
Irley Franco
Helena Duque Estrada Lopes
Vera Naegeli
Marcia Aguiar Malta
Marco Antonio de Assis Beja
Maria da Glória Esperança Martagão Gesteira
Nancy P. Naro

EQUIPE DE

PARENTIZAÇÃO

DE SENTENÇAS: Eliane dos Santos Silva
Ora Meisel
Arlene Herman

AGRADECIMENTOS

Quero testemunhar gratidão a todos aqueles que, unindo forças, tornaram possível o presente estudo: ao Mobral, que lhe deu motivo; à Fundação Ford, que se aliou ao esforço; ao Museu Nacional, que me permitiu escolher caminhos e deu-me o tempo e o espaço para trilhá-los; a Anthony Naro, aportador do know-how que deu forma à pesquisa; a todos os colegas da equipe, que se dedicaram com seu trabalho; aos alunos da Faculdade de Letras da UFRJ e colegas do Departamento de Antropologia, bons amigos e colaboradores na reflexão; aos mobralenses entrevistados, e os outros por eles representados, a dívida maior.

Miriam Lemle

Por minha parte, quero me associar aos agradecimentos já expressos pela Professora Miriam Lemle. Devo gratidão especial ao Departamento de Letras da Pontifícia Universidade Católica do Rio de Janeiro, que, além de me deixar livre para participar desta pesquisa, me facilitou o uso das instalações do Rio Datacentro, e ao Departamento de Lingüística e Filologia da Universidade Federal do Rio de Janeiro, que sempre tem me estimulado a trilhar novos caminhos na pesquisa humanística. Além de tudo, quero deixar expressa a minha gratidão à Professora Miriam Lemle, sem cujos agudos "insights" e perspicazes percepções este trabalho não poderia ter sido realizado.

Anthony J. Naro

LISTA DAS TABELAS

Tabela 1	- Categorização "old"; probabilidades e frequências	30
Tabela 2	- Categorização "neamal"; grupo morfológico; probabilidades e frequências	31
Tabela 3	- Categorização "neamal"; grupos sintático e semântico; probabilidades e frequências	33
Tabela 4	- Categorização "neamal"; número da entrevista; probabilidades e frequências	34
Tabela 5	- Categorização "neocode"; grupo morfológico; probabilidades e frequências	35
Tabela 6	- Categorização "neocode"; grupos semântico e posicional; input global	36
Tabela 7	- Categorização "neocode"; número da entrevista; probabilidades e frequências	37
Tabela 8	- Categorização "neamal"; grupo morfológico; modelo aditivo	38
Tabela 9	- Categorização "neamal"; grupos semântico e posicional; modelo aditivo	39
Tabela 10	- Grupo morfológico; amálgama da codificação "neamal"	41
Tabela 11	- Efeito de homogeneização paradigmática	43
Tabela 12	- Grupo posicional; amálgama da categorização "neamal"	44
Tabela 13	- Grupo semântico; amálgama da categorização "neamal"	46
Tabela 14	- Medidas de complexidade do sintagma nominal	130
Tabela 15	- Medidas de complexidade do sintagma verbal	131

Tabela 16 - Medidas de complexidade de sentença encaixada	132
Tabela 17 - Medidas de complexidade da sentença principal	133
Tabela 18 - Medidas de profundidade de encaixe	134
Tabela 19 - Relação entre complexidade e grau de encaixe para sintagmas nominais	135
Tabela 20 - Relação entre complexidade e grau de encaixe para sintagmas verbais	136
Tabela 21 - Relação entre complexidade e grau de encaixe para sentenças encaixadas	137

1- CARACTERIZAÇÃO DA PESQUISA

1.1 - Objetivos

A força motriz do presente estudo foi a convicção de que é responsabilidade do sistema educacional nacional tornar disponíveis a todos os membros da comunidade todos os instrumentos de acesso ao sistema cultural e produtivo da nação. Um desses instrumentos é a plena capacidade de uso do sistema de comunicação nacional, a língua portuguesa.

A língua portuguesa, código comunicativo de uma sociedade complexa e heterogênea, é ela mesma um sistema heterogêneo, uma rede de subsistemas, de variedades dialetais, decorrentes dos diferentes graus de coesão e contato dos diversos subgrupos da comunidade total. Assim sendo, ter plena capacidade de uso da língua portuguesa implica em ter a capacidade de traquejo adequado com mais de uma variedade dialetal.

Para que o sistema de ensino possa tomar a si a tarefa de promover esse quase que pluralismo sócio-dialetal, é necessário que se acumulem estudos descritivos e comparativos desses diversos dialetos em que a língua se manifesta.

É dentro dessa visão que o presente estudo se enquadra. A pesquisa, patrocinada pelas fundações Mobral e Ford, tomou como amostra a clientela do Mobral da cidade do Rio de Janeiro, e partiu da proposição de que a primeira variedade dialetal que essa camada populacional visa adquirir para capacitar-se a uma participação mais plena na vivência nacional é a varie-

dade da língua escrita de nível jornalístico e da literatura contemporânea mais acessível.

O objetivo do nosso estudo foi o de verificar pontos de diferenciação entre a variedade de língua portuguesa falada por esse grupo social e as variedades de língua escrita nas quais almejam eles adquirir competência. O conhecimento dos pontos de discrepância entre a competência lingüística em uso pelos falantes e aquela utilizada nas variedades de língua escrita acima mencionadas possibilita a previsão de pontos de dificuldade no processo de aprendizado, e conseqüentemente fornece informações relevantes para uma gradação adequada do material didático. Essa é a meta prática para a qual nos foi solicitado este trabalho.

Os aspectos da divergência interdialetoal comparativamente avaliados foram dois, o primeiro visando avaliar a capacidade de manejo da estruturação sintática em si, e o segundo visando estudar o uso de uma regra sintática em particular - a regra que estabelece a concordância entre sujeito e verbo.

O estudo sintático decorre da observação intuitiva de que as sentenças produzidas na linguagem falada de um modo geral tendem a ser bastante mais "simples" (aceite-se provisoriamente a indefinição do termo) do que as de certas variedades, inclusive jornalísticas, da linguagem escrita. Uma gradação de complexidade sintática também parece existir na linguagem escrita, se compararmos, por exemplo, as sentenças encontradas em histórias em quadrinhos com aquelas dos noticiários jornalísticos ou, mais ainda, com as dos editoriais. A serem objetivamente

comprováveis tais intuições, ter-se-ia, do ponto de vista teórico, um dado interessante sobre o grau de exigência imposto à mente para o processamento sintático nas diversas variedades da língua, e do ponto de vista prático, supondo que haja alguma relação entre as capacidades lingüísticas produtivas e as de leitura de um indivíduo, um certo embasamento para sistematizar uma didática de expansão gradual da capacidade de leitura. O nosso intuito nessa parte do trabalho foi o de estabelecer um método de quantificação da complexidade sintática, e, através dele, comparar as medidas obtidas para as seguintes variedades do português: linguagem falada por mobralsenses, linguagem escrita de histórias em quadrinhos, fotonovelas, jornais classe A (O Globo, Jornal do Brasil, Diário de Notícias), jornais classe B (Jornal dos Sports, O Dia, A Notícia), literatura nacional.

O segundo ponto focalizado na pesquisa foi a regra de concordância do verbo com o sujeito na variedade de linguagem oral empregada pelos mobralsenses. Enquanto que na linguagem escrita e na fala das pessoas cultas essa regra é categórica (i.e., o verbo sempre concorda com o sujeito em pessoa e número), na fala do grupo social ao qual pertencem os mobralsenses a concordância do verbo com o sujeito é uma regra variável (i.e. às vezes eles usa sujeito plural com verbo singular). Procuramos determinar, dentro da metodologia da moderna sociolingüística, como os fatores lingüísticos (morfológicos, sintáticos e semânticos) e sociais (nível de formalidade) interagem para favorecer ou desfavorecer o uso da regra. É claro que o enten-

dimento dos fatores que regem o uso da regra, além de interessante do ponto de vista teórico, nos armará para a elaboração de uma estratégia de ensino da norma culta da língua, passo necessário se tivermos em mente o quão estigmatizante é o uso falho da regra.

Tanto o estudo da complexidade sintática quanto o da regra de concordância foram feitos com base em amostras de dados especialmente coletados para esse fim.

O corpus da língua falada por mobraleses foi coletado da seguinte maneira. Em postos do Mobra da área urbana do Rio de Janeiro foram selecionados vinte informantes naturais do Rio de Janeiro ou adjacências (Nova Iguaçu, Caxias), homens e mulheres numa faixa etária dos 17 aos 50 anos, caracterizados pelas alfabetizadoras como sendo pessoas "desembaraçadas". Com cada um deles foram feitas sete entrevistas, em locais e circunstâncias variadas, a fim de que pudéssemos captar uma gama de comportamentos lingüísticos. Esse corpus de 140 entrevistas, gravadas eletromagneticamente e depois transcritas numa ortografia semi-fonética em alguns aspectos, serviu-nos de base seja para o estudo da regra de concordância seja para o estudo da complexidade sintática. Descrevê-lo-emos em maior detalhe na próxima seção.

O corpus que nos serviu de base para o estudo da complexidade sintática consistiu de sentenças aleatoriamente extraídas de histórias em quadrinhos (60 sentenças), fotonovelas (140 sentenças), jornais (400 sentenças) e literatura nacional (400 sentenças). A amostra de sentenças da língua falada por mobra-

lenses não foi escolhida aleatoriamente. Em cada informante, selecionamos no mínimo 30 sentenças que nos pareciam das mais complexas (i. e., com grandes encaixes à esquerda). No capítulo 2 será dada uma descrição detalhada dessa amostra.

1.2 - Caracterização da amostra de língua falada

1.2.1 - Seleção dos informantes - O critério básico que adotamos na seleção dos informantes foi o de que fossem naturais do Rio de Janeiro ou adjacências (i.e., Grande Rio), ou pelo menos que tivessem passado a parte mais significativa da infância nessa região. Essa opção decorreu do nosso intuito de eliminar o fator adicional de variação dialetal geográfica. O número de informantes entrevistados foi de 20 (vinte). Procuramos equilibrar o número de homens e mulheres (temos 9 mulheres e 11 homens), como também a gama de idades (temos 6 informantes com mais de 40 anos e 14 com menos de 30).

1.2.2 - Esquema das entrevistas - Sendo que um dos fatores de variação que desejávamos testar era o fator estilístico de grau de formalidade do encontro, planejamos um esquema de 7 entrevistas de uma hora de duração para cada informante, nas quais as circunstâncias seriam controladamente variadas de maneira a que pudéssemos vir a obter uma gama de comportamentos linguísticos no entrevistando. O esquema planejado foi o seguinte:

1^a entrevista - local das aulas.

- 2^a entrevista - local das aulas.
- 3^a entrevista - residência do entrevistador.
- 4^a entrevista - residência de outro entrevistador.
- 5^a entrevista - local das aulas.
- 6^a entrevista - local das aulas ou residência do entrevistando - um trecho de gravação feito à revelia.
- 7^a entrevista - local das aulas ou residência do entrevistando, em companhia de algum amigo deste convidado a participar.

O tipo de expectativa que tínhamos com este esquema seria a de que, supondo que os falantes cuidassem de usar "corretamente" a concordância em circunstâncias percebidas como mais tensas e formais, viesse a verificar-se uma razão direta entre o índice de concordância e o nível de formalidade social percebido pelo informante na entrevista.

Quanto à temática das entrevistas, deixamos os entrevistadores em liberdade, uma vez que o nosso interesse lingüístico precípuo não dependia diretamente de qualquer área vocabular específica. Recomendamos-lhes apenas que nas primeiras duas entrevistas levantassem os principais dados biográficos dos informantes e de seus pais (ou pessoas que os criaram), incluindo história de residência e tipo de trabalho. Oferecemos-lhes uma lista de tópicos facultativa, de possível interesse antropológico, e que foi a seguinte (sugerida por Roberto da Matta):

- a. Festividades sociais: Carnaval, Natal, 7 de setembro, etc.

- b. Festividades individuais: aniversários, casamentos, batizados.
- c. Opiniões sobre relações pais-filhos: autoridade e liberdade.
- d. Opiniões sobre relações homem-mulher: planos matrimoniais, relação de igualdade ou desigualdade entre homem e mulher, liberdade sexual.
- e. Tipos sociais: o malandro, o tofário, o caxias.
- f. Opiniões sobre a relação patrão-empregado.
- g. Sorte é azar.
- h. Eventos que resultam em "Você não sabe quem está falando".
- i. Planos de vida e projetos de trabalho.

1.2.3 - Caracterização dos informantes

1) Maria Luzia F. 23 anos. Região Administrativa do Mobral: Portuária. A família tem sede em Caxias, moradores de casario rural. O pai é de Campos (R.J.) e a mãe de S. Vicente de Paula (R.J.) Chegou a ingressar na escola primária aos 8 anos de idade, mas cessou a frequência por motivo de doença mental do pai. Aos 10 anos foi entregue pela mãe, que se empregou como doméstica, para uma irmã da igreja evangélica, para a qual fazia pequenos serviços, numa vida aparentemente bastante reclusa. Aos 11 anos começou a trabalhar em casa de família. Frequenta a igreja evangélica ("crente"). Diz ter parentes

"mais ou menos de vida", i. e. classe média baixa. Trabalha atualmente como bedel numa escola de freiras.

2) Elvira P. da C. 45 anos. Região Administrativa do Mobral: Copacabana. Nasceu na roça perto de Petrópolis. Perdeu o pai aos 2 anos. Foi criada pelo irmão mais velho. Não freqüentou escola quando criança. Começou a trabalhar como doméstica em Petrópolis aos 21 anos. Aprendeu corte e costura em Petrópolis, e aos 29 anos transferiu-se para o Rio, onde começou com um emprego de costureira. Tem uma casinha própria às margens da Av. Brasil. Fala fluente e ininterruptamente, mas parece não ter contatos humanos importantes em sua vida.

3) Marlene da C. 26 anos. Região Administrativa do Mobral: Copacabana. Nasceu em Caxias (R.J.). Pais da roça do Espírito Santo. Infância em favela de Parada de Lucas. Começou a trabalhar como doméstica aos 17 ou 18 anos. O atual patrão, fazendeiro, está quase sempre ausente, e entrega-lhe, além da incumbência de tomar conta da casa, pequenas responsabilidades de pagamentos e administração de despesas.

4) Sidnei B. dos S. 16 anos. Região Administrativa do Mobral: Centro. Nasceu no Rio de Janeiro. Pai baiano, era estivador. Mãe carioca, dona de casa. Mora na Ladeira do Barroso. Freqüentou escola em pequeno, durante alguns meses, e interrompeu porque "apanhava dos garotos maiores". Trabalha em biscates de pintor e carregador. Luta box, amadoristicamente. Gosta de cinema e vê

muita televisão. Parece bastante sociável e ligado a companheiros seus coetâneos.

5) Ubirajara T. 42 anos. Região Administrativa do Mobral: Tijuca. Nasceu no Rio, bairro do Estácio, pais de Friburgo (R.J.). Perdeu o pai aos 8 anos. Infância de rua, foi carregador de feira, vendedor de bala em porta de cinema, menino de pensão em Botafogo. Seu curriculum vitae profissional é bem variegado: trabalhou em fábrica de calçados, na Limpeza Urbana, em construção civil (azulejos), como contínuo e mecânico da Sursan e na Esag. Atualmente trabalha na portaria de um depósito da Esag. É casado e separado da mulher. Gosta de samba, já frequentou escolas de samba, gosta de ver futebol no Aterro do Flamengo, quando mais jovem gostava de jogar. É um carioca típico, de contato humano generoso.

6) Ilda F. S. 43 anos. Região Administrativa do Mobral: Copacabana. Nasceu em Nova Iguaçu. Perdeu a mãe aos 5 anos, foi criada por uma tia e depois pela madrasta. Chegou a ir à escola quando pequena, mas foi expulsa por mau comportamento e falta de firmeza em casa. Começou a trabalhar em casa de família aos 16 anos. Trabalha há 24 anos na mesma casa. Tem uma filha de 18 anos, criada na casa dos patrões, com os quais parece haver uma relação de grande paternalismo e dependência. Diz gostar da igreja católica e frequenta também centros espíritas.

7) Madalena M. de S. 45 anos. Região Administrativa do Mobral: Copacabana. Nasceu em Valão do Barro, R.J.

Pai e mãe residiam em Raiz da Serra, R.J., e eram lavradores. Veio trabalhar em emprego doméstico no Rio aos 8 anos de idade. Frequenta candomblé. Gostava de samba quando mais jovem, já desfilou no Salgueiro. Parece ligar-se bastante às pessoas em cuja casa trabalha.

8) Henrique Q. N. 17 anos. Região Administrativa do Mobra: Lagoa. Mãe de Campos, R. J. Pai filho de italiano com espanhola, feirante. Passou a primeira infância em São João de Meriti. Ajudava o pai na feira, e trabalhou também em fábrica de sandálias. Nessa época chegou a ir à escola, mas sem regularidade por falta de direção em casa. Perdeu a mãe há poucos meses, e mora agora com a avó na Cruzada São Sebastião. Já teve um emprego de boy numa boutique de Ipanema, e atualmente trabalha numa lavanderia no Leblon. Gosta de futebol e de praia, dança "brown", gosta de música "black" e mantém amizades com jovens tanto da Cruzada quanto de São João de Meriti.

9) Jorge M. 16 anos. Região Administrativa do Mobra: Tijuca
Pais ambos cariocas. Profissão do pai: bombeiro hidráulico. Profissão da mãe: dona de casa. Reside no Beco dos Oicós, Tijuca. Não trabalha nem parece interessar-se seriamente por adquirir capacitação profissional. Teve frequência ocasional à escola primária, sem explicação do motivo por que não o fez regularmente. Gosta de praia e de futebol, frequenta bares com os amigos, tem namoradas e

relações de amizade com outros rapazes. Tem o hábito de assistir televisão.

10) Sonia M. S. 22 anos. Região Administrativa: Lagoa. Nasceu em Laranjeiras. Pai de Petrópolis, R.J., mãe de Minas Gerais. A família mora agora em Engenheiro Pedreira, R.J. Trabalha como doméstica desde os 14 anos, em família rica. Aos 5 e 6 anos freqüentou a escola, mas saiu "porque não gostava". Tem elos de amizade com as colegas de trabalho e visita regularmente a família. Envolve-se também bastante com as pessoas para quem trabalha. Gosta muito de ver televisão e segue com regularidade várias novelas.

11) Elisabeth E. da S. 19 anos. Região Administrativa: Lagoa. Mora com a família (mãe e 4 irmãos) num conjunto residencial da Rua Marquês de São Vicente. Nunca trabalhou, por ter sido diagnosticada como retardada ou doente dos nervos. A família vive de uma soma de salários dos irmãos e irmãs, operários de fábrica ou pequenos funcionários de escritório. A informante realiza a contra-gosto tarefas domésticas, conversa com vizinhos, assiste televisão, freqüenta festinhas. Está ligada a um grupo de jovens da paróquia local. É também devota de Yemanjá e já passou por fenômenos de transe.

12) Izeni M. A. (sexo feminino). 23 anos. Região Administrativa. Lagoa. Pai lavrador em Paraíba do Sul, R.J., mãe dona de casa. Só aos 22 anos saiu de casa para trabalhar como empregada doméstica. Visita raramente a família, com a qual parece ter atritos. Não transpare-

ceu que tivesse grande contato com outras pessoas como ela. Tem um namorado, vai ocasionalmente a bailes num clube. Vê um pouco de televisão. Tem metas existenciais típicas de classe média, como um apartamento em Ipanema, montar a cavalo com roupas apropriadas.

13) Cantidio S. 54 anos. Região Administrativa: Botafogo. Filho de lavradores, do interior de Campos. Veio ao Rio com 18 anos. Há 35 anos é roupeiro do time de basquetebol amador de um dos grandes clubes esportivos do Rio e nessa ocupação teve oportunidade de fazer várias viagens pelo Brasil e mesmo ao exterior. Mora num apartamento em Botafogo com a esposa. Quando jovem, gostava de jogar futebol e já tomou parte em escola de samba. Gosta de cinema, de televisão, e de sair com amigos para bater papo e tomar cerveja em bares. Cultiva o contato com familiares. Gaba-se de ter amigos importantes que já foram à sua casa.

14) Valdecio R. S. 18 anos. Região Administrativa do Mobral: Botafogo. A família é de Niterói. Está servindo exército no quartel do Leblon, e morando em casa de um irmão em Santa Tereza. Antes do serviço militar, era pedreiro. Nas horas de folga, gosta de jogar futebol e ir à praia. É torcedor do Flamengo. Gosta de ouvir a Rádio Globo e vê um pouco de televisão nos fins de semana. Vai aos bailes do Clube Carioca. Tem uma namorada fixa. Deve ter tido relativamente pouco contato com a classe média.

15) Paulo R. S. - 19 anos. Posto do Mobral: Copacabana. Não tem conhecimento do pai e a mãe é falecida. Morou em Nova Iguaçu até os 5 anos, com os avós. Depois esteve até os 14 anos num colégio interno em Nova Iguaçu, do qual fugia sempre que possível. Mora atualmente com os avós no morro do Leme. Trabalha numa drogaria, como entregador. Gosta de jogar futebol e de ir à praia, onde às vezes joga vôlei. Visita parentes em Realengo, e sai às vezes com outros rapazes, indo a bailes. Não aprecia carnaval, não vê televisão nem vai a cinema. Seu contato com classe média deve ser reduzidíssimo, limitando-se a contatos eventuais na drogaria.

16) Vanilda - 15 anos. Posto do Mobral: Tijuca. Filha de uma empregada doméstica. Nunca viveu com a mãe e conheceu pouco o pai. Foi criada pela avó em Coelho da Rocha até os 9 anos. Foi um pouco à escola mas não gostava e apanhava da avó que queria forçá-la a ir. Com a morte da avó, foi morar com a mãe, no emprego desta. Depois preferiu empregar-se também em casa de família. Empregou-se pela primeira vez aos 12 anos. Gosta de samba e bailes. Frequenta centros de macumba com a mãe e pretende integrar-se cada vez mais nessa atividade. Teve vários namorados, com os quais vai a bailes e ensaios de escola de samba. Ocasionalmente, assiste televisão. Deseja ser médica.

17) George Claudio S. - 37 anos. Não é mobralense, mas é analfabeto. Nasceu em Barra do Piraí (R.J.)

e veio para o Rio com 1 mês. Os pais são de Queimados (R.J.), estivador e doméstica, respectivamente. Mora em Nova Iguaçu. É faxineiro de um edifício em Ipanema. É casado e tem quatro filhas. Sua história de trabalho inclui vendedor de cocadas, camelô, entregador de farmácia, operário em fábrica de vidro, entregador de jornal, baterista de circo, baterista de boite. Participa intensamente de uma escola de samba em Nova Iguaçu, da qual é diretor artístico. Tem muita vida de família, visitando os pais, os sogros. É uma personalidade extremamente criativa e extrovertida, com traços nítidos de liderança: organiza almoços e festas, nas quais canta, toca violão, bateria, bate papo até tarde. Não vê muita televisão, mas tem o aparelho em casa.

18) Lidio P. O. R. - 16 anos. Posto do Morro: Copacabana. Mora no morro do Pavão com a mãe e quatro irmãos. A mãe é doméstica diarista. Durante a infância, residiu no Pavão, em Bonsucesso e em Caxias, tendo frequentado escola primária nos três lugares, até o segundo primário. Trabalha como auxiliar numa oficina de protético, com tarefas de compra de matéria-prima e entrega de dentaduras. Está bastante integrado na comunidade do morro onde reside, participando de peladas e de ensaios de um bloco carnavalesco que desfila na avenida. Ocasionalmente, visita parentes em Bonsucesso e Caxias. Foi poucas vezes ao cinema e quase não tem oportunidades de assistir televisão.

19) J. Carlos, 19 anos. Nasceu em Nova Friburgo (R.J.). Não conheceu o pai, mas sabe que era boiadeiro. Morava com a mãe e dois irmãos mais velhos, trabalhando como engraxate e outros biscates. Perdeu a mãe aos 11 anos, decidiu vir ao Rio, morou nas ruas, depois fez amizade com um indivíduo que lhe deu trabalho numa banca de jornal e uma vaga num quarto de sua casa (uma vila no Flamengo). Na época das entrevistas, trabalhava para uma firma de jardinagem e limpeza das ruas. Nas horas de lazer, gosta de jogar futebol no Aterro do Flamengo. Frequenta bailes "brown", vai à praia, ouve um pouco de rádio e quase não vê televisão nem vai ao cinema. É um indivíduo sem vida familiar nem comunal. de contatos humanos de natureza efêmera. O seu grau de contato com a classe média é nulo. Tem metas de vida humildes e limitadas à solução dos problemas imediatos de sobrevivência material.

20) Jorge F. Santos. Posto do Mobral: Vila Isabel. Nasceu em Minas Gerais e foi trazido para o Rio com poucos meses de idade. Morou na favela da Baixa do Sapateiro. Foi à escola, mas sem aproveitamento porque "fazia bagunça". Aos dez anos começou a trabalhar juntando latas e ferro velho para vender. Foi operário da Quartzolit. Atualmente é porteiro de um clube do Grajaú. É filho de uma família de doze, e ainda mora com os pais e numerosos

irmãos. Visita outros ramos de família, e parece ligado à casa. Gosta de sair com namoradas. Não é de carnaval nem futebol, e não parece estar dentro de grupo algum de amizades. Frequentou ocasionalmente centros de macumba, e atualmente está sendo cativado por um grupo jovem da igreja católica. Cinema e televisão entram só ocasionalmente em sua vida.

2 - A REGRA DE CONCORDÂNCIA

2.1 - Fatores condicionantes

2.1.1 - Conceito de regra variável

Dá-se, nos estudos socio-lingüísticos, o nome técnico de regra variável a uma regra lingüística facultativa, i. e., que ora se aplica ora não se aplica. O conceito foi introduzido por William Labov, o qual mostrou a importância de se estabelecerem, para uma regra facultativa, os fatores lingüísticos e extralingüísticos que favorecem ou refreiam a aplicação da regra. Veja-se William Labov: *Contraction, Deletion and Inherent Variability of the English Copula*, Language in the Inner City, pp. 65-129. Philadelphia: University of Pennsylvania Press, 1972. Originalmente em Language 45 (1969).

A regra de concordância do verbo com o sujeito na fala da camada social mais baixa da nossa sociedade é um caso exemplar desse tipo de variação lingüística. Um mesmo falante pode empregar, na mesma sentença, o mesmo verbo ora com ora sem a concordância. Por exemplo, "um mês eles pagavava no dia certo, outro mês não pagavaam". Temos numerosos exemplos desse tipo em nossas entrevistas. No entanto, a variação não é aleatória. Fatores diversos, lingüísticos e extralingüísticos, podem influir no desempenho da regra, favorecendo ou refreando a sua aplicação. Em cada caso particular em que a regra se aplicou ou

deixou de aplicar-se, podemos supor que o resultado final foi determinado pela resultante das várias pressões em interação. Cabe ao lingüista depreender os fatores condicionantes da aplicação de uma regra variável, e avaliar a importância relativa de cada um deles. Esses fatores podem ser de vários tipos. Podem ser sociais, quando os vários sub-grupos de uma comunidade empregam em grau diverso uma dada regra. No caso da concordância verbo-sujeito, por exemplo, a classe alta aplica a regra em quase 100% dos casos, ao passo que na classe social estudada encontramos uma gama que vai dos 30% aos 80% de freqüência de aplicação da regra, com uma grã-média de 50%. O fator estilístico costuma também ser um fator de peso: uma dada regra varia em freqüência de aplicação conforme o nível de formalidade percebido pelos falantes no seu encontro. Finalmente, temos os fatores lingüísticos propriamente, i. e., os contextos semânticos, sintáticos ou morfológicos em que ocorre uma dada possibilidade de aplicação da regra.

2.1.2 - Descrição das variáveis

Passaremos a descrever as variáveis por nós hipotetizadas como fatores de influência relevante na aplicação da regra de concordância verbo-sujeito pela população de mobaralenses da nossa amostra.

i. Variável estilística.

Conforme exposto em 1.2.2 (página 5) a seqüência de sete entrevistas que fizemos com cada informante foi planejada de

forma a que variassem os locais e os participantes do encontro. Essa variação nas circunstâncias extra-lingüísticas, supostamente, poderia acarretar reações lingüísticas diversas, as entrevistas mais formais provocando um mais alto nível de atenção e conseqüentemente um índice mais alto de aplicação da concordância verbo-sujeito.

ii. Variável Semântica.

Essa variável consiste no traço semântico de definição ou indefinição do sujeito. Por exemplo, se o falante está falando da sua família, e diz, referindo-se aos seus irmãos: "Eles tava estudando e por isso sô chegava em casa de noite", consideramos o sujeito de tava e chegava como definido, porque a sua referência é óbvia. Ao contrário, nos exemplos abaixo a referência exata do sujeito de cada verbo não é óbvia, e os casos desse tipo foram classificados com o valor de indefinido: "O trem estava atrasando porque estavam consertando uns trilhos lá em Madureira".

"Dizem que o lobishomem aparece na lua nova."

"Aquele dor que eu sentia nunca que passava, então eles resolveram operar".

O valor indefinido da variável foi por nós rotulado de I, e o valor definido de P (plural).

iii. Variável posicional.

Essa variável diz respeito à posição do sujeito relativamente ao verbo. Consideramos os seguintes valores:

A - Sujeito imediatamente antes do verbo.

Ex.: Aqueles tempo foram difícil pra mim.

S - Sujeito precedendo o verbo, mas dele separado por mais de quatro palavras átonas, ou por um conjunto de palavras contendo acento.

Ex.: Os vizinhos de lá todo dia de noite aparecem para pedir coisa emprestada.

Eles também já não querem saber disso.

O - Sujeito oculto.

Ex.: Eles disseram que não vinha, mas vieram

D - Sujeito depois do verbo.

Ex.: Sumiu todos os meus discos.

Veio muitos parentes de fora para a festa.

iv. Variável morfológica.

Essa variável diz respeito ao tipo de diferença fônica entre a forma singular e a forma plural do verbo, diferença essa que podemos colocar numa escala ordinal, hipotetizando que o grau de saliência fônica da diferença morfológica entre a forma singular e a forma plural esteja diretamente relacionado com a probabilidade de aplicação da regra de concordância.

Dentro dessa hipótese, experimentamos com três modalidades de sub-divisão da escala de saliência fônica, a segunda mais fina do que a primeira e a terceira agrupando certas sub-classes da segunda. Apresentaremos as três categorizações.

1^a categorização (rotulada como "old").

R - Verbo regular. Diferença entre singular e plural reside em desinência átona.

Ex.: fala-falam ; come-comem

V - diferença entre singular e plural reside numa vogal nasal

final átona.

Ex.: faz-fazem; quer-querem;

L - sobreposição de raiz e desinência, com acento

Ex.: dê-dão; está-estão; vai-vão

E - acento na desinência, diferença fonológica maior do que simples nasalização. Pretérito perfeito dos verbos regulares.

Ex.: falou-falaram; comeu-comeram

F - pretêritos perfeitos irregulares

Ex.: foi-foram; teve-tiveram; trouxe-trouxeram; fez-fizeram

W - formas inteiramente distintas para o singular e o plural.
Caso único: é - são.

Para essa categorização a hipótese que aventamos é a de que a hierarquização das probabilidades de concordância se apresenta na seguinte ordem:

R	V	L	E	F	W
sem acento		com acento			

3a. categorização (rotulada como "neamal").

B - verbos regulares nos quais a diferença entre a vogal das desinências singular e plural consiste apenas na nasalização.

Ex.: come-comem; constrói-controem; ve-vêem.

N - forma singular em a átono e plural em am (na representação ortográfica)

Ex.: fala-falam

- V - diferença entre singular e plural consiste numa vogal final átona, possivelmente nasalizada

Ex.: faz-fazem; quer-querem

- G - sobreposição de raiz e desinência, com acento

Ex.: está-estão; dá-dão; vai-vão

- W - formas totalmente distintas para singular e plural

Ex.: é - são

- K - pretêritos perfeitos regulares

Ex.: falou-falaram; comeu-comeram

- Y - pretêritos perfeitos irregulares, com mudança de acento

Ex.: trouxe- trouxeram; disse- disseram; fez- fizeram

- Z - Caso único: foi-foram

- U - Caso único: veio-vieram

- Q - Caso único: era-eram

2a. categorização (rotulada como "neo-code")

Nesta categorização as seguintes classes de "neamal" foram desdobradas:

"neamal" "neocode"

B $\left\{ \begin{array}{l} B \text{ polissílabos} \\ T \text{ monossílabos} \end{array} \right.$

G $\left\{ \begin{array}{l} G \text{ está-estão; dá-dão} \\ H \text{ vai-vão} \end{array} \right.$

K $\left\{ \begin{array}{l} J \text{ comeu-comeram; partiu-partiram} \\ K \text{ falou-falaram} \end{array} \right.$

Y $\left\{ \begin{array}{l} Y \text{ trouxe- trouxeram; disse- disseram} \\ X \text{ fez- fizeram; teve- tiveram} \end{array} \right.$

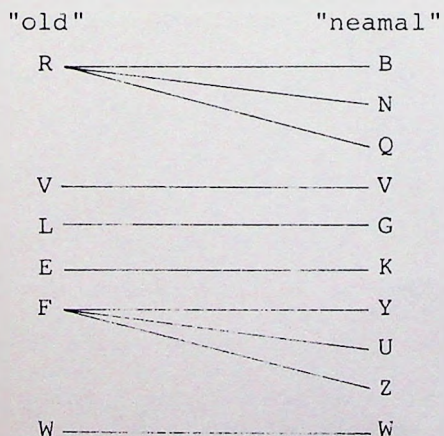
A necessidade de se elaborar a categorização "neo-code" decorreu do fato dos resultados "old" não mostrarem uma correspondência satisfatória ao modelo hipotetizado. Mantendo a hipótese original, impomos uma subdivisão mais fina, a fim de localizar com precisão os pontos de discrepância. A "neo-code" foi em seguida reduzida à "neamal", já que certas das suas categorias se revelaram sem importância. As correspondências entre "old" e "neamal" são:

- A) À categoria R da categorização "old" correspondem as categorias B e N de "neamal". A subdivisão foi feita em vista do fato de que na categoria B a diferença fonética entre singular e plural consiste apenas na nasalização da vogal final do plural, na pronúncia real dos falantes da nossa amostra. (Isto é, eles frequentemente não dizem [elis komẽy], mas sim [elis komĩy]). Na categoria N, esses falantes na maioria dos casos têm para a forma do plural a pronúncia de [ũ] ou [u] (onde a pronúncia da norma culta é [ãw]). Assim, a saliência é maior na categoria N ([a / u]) do que em B ([i / ãy]).
- B) Da categoria F original foram retirados os pares veio - vieram e foi-foram, e computados separadamente, nas categorias U e Z, respectivamente. Isto porque cada um desses verbos tem na verdade características formais idiossincráticas na oposição singular - plural.
- C) A classe Q, contendo apenas a forma era - eram, foi estabelecida com o intuito de testar uma hipótese adicional: a pressão para uniformização paradigmática.

Compare-se o paradigma do verbo falar com o do verbo ser, observando-se as posições de suas formas do presente e pretérito imperfeito na escala de saliência fônica da oposição singular-plural. Para falar, o presente e o pretérito imperfeito estão na mesma posição, ambos R: fala-falam; falava-falavam. Para ser, o pretérito imperfeito é R e o presente do indicativo é W: era-eram; é - são.

A hipótese aventada é que, além do princípio da saliência fônica, como cruzando forças com ele, esteja em jogo também um princípio de uniformidade paradigmática, que tenderia a homogeneizar as probabilidades de concordância de um mesmo verbo em todas as suas formas. Evidência favorável à hipótese será um índice de concordância mais alto para era-eram do que para os demais representantes da categoria formal R.

D) As demais categorias foram mantidas, e a correspondência global é a seguinte:



Para a categorização "neamal", a hipótese que aventamos é a de que o modelo de hierarquização das probabilidades de concordância seja o seguinte:

<u>B</u>	<u>N</u>	<u>V</u>	<u>G</u>	<u>Z</u>	<u>W</u>	<u>K</u>	<u>Y</u>
contraste na			contraste na sobre-			contraste na de-	
desinência,			posição raiz-desi-			sinência, com	
sem acento			nência, com acento			acentos.	

As diferenças entre as duas classificações serão comentadas na seção 5 deste capítulo.

2.2 - Método de avaliação da interação dos fatores

Os fatores que influem na probabilidade de se realizar a aplicação de uma regra gramatical, sejam lingüísticos ou sociais, podem ser colocados em grupos tais que os fatores de cada grupo são mutuamente exclusivos, i. e., nunca co-ocorrem em um contexto. Assim, cada contexto será caracterizado pela presença nele de um determinado conjunto de fatores. No caso da regra da concordância verbal temos quatro grupos de fatores - classe morfológica, posição, natureza do sujeito, e situação social - e em cada contexto verifica-se a presença de um fator de cada grupo.

Utilizamos três modelos para avaliar a interação destes fatores. No primeiro, consideramos insignificante tal interação e, isolando todos os casos em que ocorre dado fator independentemente de quaisquer outros fatores por ventura existentes, calculamos a frequência relativa de aplicação da regra para esse fator. Até recentemente foi esse o modelo mais usado em pesquisas lingüísticas. O segundo modelo é aditivo - postula que os efeitos provocados por cada fator presente em dado contexto se somam aritmeticamente onde "efeito" se mede através do desvio da grã-média. É esse o modelo mais usado em ciências sociais e subjaz à técnica conhecida como "análise multidimensional da variância". O terceiro modelo, criado especialmente para pesquisas lingüísticas pelo Prof. D. Sankoff, da Université de Montréal, diz respeito à probabilidade, e não à frequência, de aplicação de uma regra.

Neste modelo, pressupõe-se que a probabilidade da aplicação de uma regra em dado contexto contendo $\underline{n}$ fatores é dado por:

$$(1) \quad \frac{p}{1-p} = \frac{p_0}{1-p_0} \times \frac{p_1}{1-p_1} \times \dots \times \frac{p_j}{1-p_j}$$

onde os p_i ($1 < i < n$) são quantidades no intervalo entre zero e um ($0 < p_i < 1$) associados aos $\underline{n}$ fatores presentes no contexto e p_0 é outra quantidade ($0 < p_0 < 1$) associada à regra em si. É praxe geral chamar os p_i de "probabilidade de aplicação da regra na presença do fator f_i ", embora a equação (1) não esteja de acordo com os modelos probabilísticos convencionais. A quantidade p_0 é chamada de "input" e pode ser considerada a

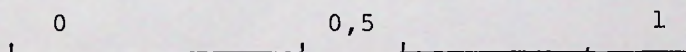
probabilidade de aplicação quando todos os outros fatores são neutros. Para ver isto um pouco melhor, chamemos a quantidade à direita da equação (1) de $\hat{p}$ e resolvamo-la para p .

$$p = \frac{\hat{p}}{1+\hat{p}}$$

Vemos então que a probabilidade global de aplicação da regra tende a zero quando $\hat{p}$ tende a zero e que tende a 1 quando $\hat{p}$ tende a infinidade. Por outro lado, $\hat{p}$ aumenta quando o valor do termo

$$\frac{p_i}{1-p_i} \quad (1 < i < j)$$

for maior que 1 e diminui quando o valor deste termo for menor do que 1, por causa de sua definição multiplicativa. Cada um destes termos, por sua vez, será menor que 1 se p_i for menor que 0,5, igual a 1 se p_i for igual a 0,5, e maior que 1 se p_i for maior que 0,5. Concluimos que um fator com probabilidade menor que 0,5 inibe a aplicação da regra, enquanto que uma probabilidade maior que 0,5 favorece a sua aplicação:



Voltando agora ao p_0 , vê-se que quando os fatores são neutros, isto é, cada $p_i = 0,5$ a probabilidade da aplicação da regra é dada por p_0 .

Comparando os três modelos, sentimos que o da frequência deve ser inapropriado, dado que não leva em conta qualquer in-

terrelacionamento dos fatores. No entanto, só cálculos estatísticos sofisticados poderiam resolver entre os outros dois modelos. Felizmente, este problema não se põe para nós uma vez que os três modelos deram resultados paralelos.

2.3 - Programas Utilizados

Para o cálculo das frequências relativas de aplicação da regra de concordância para cada fator foi utilizado o programa SWAMINC3 (V. "The Swaminc Program, Anthony J. Naro, mimeografado, Departamento de Letras, PUC/RJ). Para a análise da variância, utilizamos o subprograma ANOVA (modalidade "classic experimental") de SPSS 6.02. Aproveitando uma opção deste subprograma, provocamos o cálculo de uma tabela de "multiple classification analysis", o que mede o efeito de cada fator quando ajustado pelos efeitos dos outros fatores de acordo com o modelo aditivo. (V. 'Analysis of variance and covariance', Statistical Package for the Social Sciences, Norman H. Nie et al, New York: McGraw-Hill, pp. 398-433). No caso do modelo de Sankoff, utilizamos o programa VARBRUL2 (V. VARBRUL2, David Sankoff, mimeografado, Centre de Recherches Mathématiques, Université de Montreal).

2.4 - Resultados

A seguir apresentamos, para todos os 20 informantes, os resultados na forma de frequências e probabilidades, para a codificação "old"; para 13 informantes, apresentamos sob essa mesma forma, resultados para "neo-code" e "neamal". A seguir re-

latamos resultados do mesmo tipo referentes ao grupo posicional, ao tipo de sujeito, e ao número de entrevista. Esta última tabela foi baseada na codificação "neamal" para os 13 informantes mencionados acima, e na codificação "old" para os restantes.

Os resultados do modelo ANOVA são apresentados unicamente para 13 informantes na categorização "neamal". Neste modelo temos testes da significância de cada um dos efeitos principais, bem como das suas interações, onde por "interação" se entende agora interações não previstas pelo modelo aditivo. De fato, se estas interações não previstas forem grandes, o modelo fica incapaz de separar o efeito de cada fator considerado individualmente. Para todos os informantes, com a exceção de Cantídio, todos os efeitos, tanto principais quanto interações não previstas, se mostraram estatisticamente significativos ao nível de 0,001. No entanto, consideramos justificável desprezar as interações não previstas uma vez que, mesmo jogando-as no termo de erro, conseguimos dar conta de mais de 85% da variância em todos os informantes menos um.

Para as categorizações "neo-code" e "neamal" damos probabilidades e freqüências para amálgama dos 13 informantes, e um adicional, Natanael, amigo do informante Paulo. Os efeitos ajustados, calculados por ANOVA, são apresentados unicamente para "neamal".

TABELA 1 - CATEGORIZAÇÃO "OLD"; PROBABILIDADES E FREQUÊNCIAS

INFORMANTES		R Fala-Falam Escreve-Escroevem Constrói-Constroem	V Diz-Dizem Quer-Querem	L Está-Estão Vai-Vão	E Falou-Falaram Escreveu-Escreveram	F Fez-Fizeram Trouxe-Trouxeram Foi-Foram	W É - São
ELVIRA 844	PROB. FREQ.	0,12 $\frac{132}{440} = 30,0\%$	0,34 $\frac{31}{51} = 60,8\%$	0,39 $\frac{69}{107} = 64,5\%$	0,74 $\frac{117}{137} = 85,4\%$	0,82 $\frac{60}{65} = 92,3\%$	0,62 $\frac{35}{44} = 79,5\%$
M.L.FAUSTINA 841	PROB. FREQ.	0,08 $\frac{51}{444} = 11,5\%$	0,42 $\frac{27}{55} = 49,1\%$	0,49 $\frac{49}{92} = 53,3\%$	0,74 $\frac{94}{121} = 77,7\%$	0,63 $\frac{27}{39} = 69,2\%$	0,76 $\frac{69}{90} = 76,7\%$
HILDA 743	PROB. FREQ.	0,29 $\frac{111}{352} = 31,5\%$	0,10 $\frac{20}{118} = 19,9\%$	0,60 $\frac{59}{105} = 56,2\%$	0,68 $\frac{74}{100} = 74,0\%$	0,45 $\frac{7}{19} = 36,8\%$	0,90 $\frac{42}{49} = 85,7\%$
SIDNEI 689	PROB. FREQ.	0,19 $\frac{175}{360} = 48,6\%$	0,37 $\frac{29}{38} = 76,3\%$	0,57 $\frac{84}{104} = 80,8\%$	0,73 $\frac{107}{119} = 89,9\%$	0,69 $\frac{24}{29} = 82,8\%$	0,48 $\frac{28}{39} = 71,8\%$
UBIRAJARA 665	PROB. FREQ.	0,14 $\frac{107}{351} = 30,5\%$	0,10 $\frac{11}{44} = 25,0\%$	0,64 $\frac{91}{113} = 80,5\%$	0,62 $\frac{42}{55} = 76,4\%$	0,84 $\frac{31}{34} = 91,2\%$	0,79 $\frac{60}{68} = 88,2\%$
SONIA 602	PROB. FREQ.	0,34 $\frac{225}{303} = 74,3\%$	0,07 $\frac{2}{6} = 33,3\%$	0,57 $\frac{60}{67} = 89,6\%$	0,69 $\frac{98}{106} = 92,5\%$	0,77 $\frac{34}{36} = 94,4\%$	0,72 $\frac{78}{84} = 92,3\%$
ILZENI 439(431)	PROB. FREQ.	0,04 $\frac{9}{205} = 4,4\%$	- $\frac{0}{8} = 0,0\%$	0,60 $\frac{36}{60} = 60,0\%$	0,72 $\frac{65}{95} = 68,4\%$	0,59 $\frac{15}{28} = 53,6\%$	0,81 $\frac{35}{43} = 81,4\%$
MADALENA 396(394)	PROB. FREQ.	0,14 $\frac{41}{132} = 31,1\%$	0,16 $\frac{14}{46} = 30,4\%$	0,63 $\frac{32}{44} = 72,7\%$	0,71 $\frac{86}{96} = 89,6\%$	0,81 $\frac{39}{43} = 90,7\%$	0,65 $\frac{28}{35} = 80,0\%$
GEORGE M. 394(388)	PROB. FREQ.	0,08 $\frac{15}{187} = 8,0\%$	0,15 $\frac{14}{30} = 46,7\%$	0,67 $\frac{41}{68} = 60,3\%$	0,83 $\frac{31}{39} = 79,5\%$	+	0,87 $\frac{52}{64} = 81,3\%$
MARLENE 376	PROB. FREQ.	0,08 $\frac{12}{182} = 6,6\%$	0,15 $\frac{2}{6} = 33,3\%$	0,70 $\frac{23}{40} = 57,5\%$	0,80 $\frac{70}{94} = 76,5\%$	0,64 $\frac{8}{13} = 61,5\%$	0,79 $\frac{26}{41} = 63,4\%$
HENRIQUE 372	PROB. FREQ.	0,16 $\frac{41}{193} = 21,2\%$	0,16 $\frac{4}{19} = 21,1\%$	0,42 $\frac{29}{59} = 49,2\%$	0,87 $\frac{57}{64} = 89,1\%$	0,52 $\frac{6}{10} = 60,0\%$	0,83 $\frac{23}{27} = 85,2\%$
JORGE F. 312	PROB. FREQ.	0,14 $\frac{10}{125} = 8,0\%$	0,05 $\frac{1}{22} = 4,5\%$	0,74 $\frac{21}{36} = 58,3\%$	0,80 $\frac{40}{51} = 78,4\%$	0,73 $\frac{10}{16} = 62,5\%$	0,79 $\frac{40}{62} = 64,5\%$
VANILDA 305	PROB. FREQ.	0,25 $\frac{100}{165} = 60,6\%$	0,23 $\frac{18}{25} = 72,0\%$	0,61 $\frac{37}{44} = 84,1\%$	0,79 $\frac{39}{42} = 92,9\%$	0,55 $\frac{5}{7} = 71,4\%$	0,59 $\frac{18}{22} = 81,8\%$
GEORGE C. 228	PROB. FREQ.	0,06 $\frac{9}{97} = 9,3\%$	0,18 $\frac{5}{14} = 35,7\%$	0,58 $\frac{23}{44} = 52,3\%$	0,74 $\frac{33}{41} = 80,5\%$	0,83 $\frac{9}{12} = 75,0\%$	0,80 $\frac{16}{20} = 80,0\%$
LIDIO 217	PROB. FREQ.	0,09 $\frac{9}{110} = 8,2\%$	0,16 $\frac{1}{5} = 20,0\%$	0,77 $\frac{23}{36} = 63,9\%$	0,83 $\frac{42}{51} = 82,4\%$	0,84 $\frac{7}{11} = 63,6\%$	0,39 $\frac{2}{4} = 50,0\%$
WALDECIO 166(152)	PROB. FREQ.	0,01 $\frac{3}{57} = 5,3\%$	0,05 $\frac{2}{9} = 22,2\%$	0,47 $\frac{22}{29} = 75,9\%$	0,83 $\frac{50}{55} = 90,9\%$	0,99 $\frac{2}{5} = 40,0\%$	0,76 $\frac{9}{11} = 81,8\%$
ELISABETH 127(120)	PROB. FREQ.	0,29 $\frac{5}{56} = 8,9\%$	- $\frac{0}{2} = 0,0\%$	0,12 $\frac{1}{30} = 3,3\%$	0,98 $\frac{22}{24} = 91,7\%$	0,93 $\frac{9}{11} = 81,8\%$	0,71 $\frac{2}{6} = 33,3\%$
J. CARLOS 112(109)	PROB. FREQ.	0,01 $\frac{6}{39} = 15,4\%$	0,20 $\frac{7}{17} = 41,2\%$	0,95 $\frac{16}{19} = 84,2\%$	0,95 $\frac{23}{24} = 95,8\%$	+	0,51 $\frac{7}{14} = 50,0\%$
CANTIDIO 151(105)	PROB. FREQ.	0,18 $\frac{24}{61} = 39,3\%$	0,40 $\frac{8}{18} = 44,4\%$	0,56 $\frac{14}{18} = 77,8\%$	+	0,85 $\frac{7}{8} = 87,5\%$	+
PAULO 98	PROB. FREQ.	0,18 $\frac{2}{45} = 4,4\%$	0,00 $\frac{0}{3} = 0,0\%$	0,33 $\frac{4}{12} = 33,3\%$	0,81 $\frac{17}{21} = 81,0\%$	0,42 $\frac{3}{7} = 42,9\%$	0,70 $\frac{7}{10} = 70,0\%$

TABELA 2 - CATEGORIZAÇÃO 'NEAMAL'
GRUPO MORFOLÓGICO
PROBABILIDADES E FREQUÊNCIAS

INFORMANTES		B	N	V	G (= L)	Z	W	K (= E)	Y	U	Q	INPUT
		Come-Comem Vê-Vêem Constrói-Constroem	Fala-Falam	Faz-Fazem Quer-Querem	Está-Estão Dá-Dão Vai-Vão	Foi-Foram	E - São	Falou-Falaram Comeu-Comeram	Trouxe-Trouxeram Fez-Fizeram Disse-Disseram	Veio-Vieram	Era-Eram	
AMALGAMA DOS 14 INFORMAN- TES 5589	PROB.	0,06	0,17	0,27*	0,58	0,69	0,81	0,80	0,83	0,74	0,22	0,51
	FREQ.	$\frac{85}{626} = 13,6\%$	$\frac{622}{2099} = 29,6\%$	$\frac{124}{286} = 43,4\%$	$\frac{503}{768} = 65,5\%$	$\frac{95}{131} = 72,5\%$	$\frac{457}{555} = 82,4\%$	$\frac{730}{897} = 81,4\%$	$\frac{98}{111} = 88,3\%$	$\frac{25}{34} = 73,5\%$	$\frac{29}{82} = 35,4\%$	$\frac{2768}{5589} = 49,5\%$
ELVIRA 844(838)	PROB.	0,06	0,19	0,40	0,46	0,84	0,67	0,80	0,86	+	0,30	0,51
	FREQ.	$\frac{13}{99} = 13,1\%$	$\frac{114}{330} = 34,5\%$	$\frac{31}{51} = 60,8\%$	$\frac{69}{107} = 64,5\%$	$\frac{14}{16} = 87,5\%$	$\frac{35}{44} = 79,5\%$	$\frac{117}{137} = 85,4\%$	$\frac{40}{43} = 93,0\%$	$\frac{6}{6} = 100,0\%$	$\frac{5}{11} = 45,5\%$	$\frac{444}{844} = 52,6\%$
M.L. FAUSTINA 841	PROB.	0,02	0,12	0,52	0,60	0,57	0,83	0,82	0,90	0,41	0,59	0,47
	FREQ.	$\frac{1}{62} = 1,6\%$	$\frac{44}{371} = 11,9\%$	$\frac{27}{55} = 49,1\%$	$\frac{49}{92} = 53,3\%$	$\frac{10}{19} = 52,6\%$	$\frac{69}{90} = 76,7\%$	$\frac{94}{121} = 77,7\%$	$\frac{16}{18} = 88,9\%$	$\frac{1}{2} = 50,0\%$	$\frac{6}{11} = 54,5\%$	$\frac{317}{841} = 37,7\%$
UBIRAJARA 665(649)	PROB.	0,05	0,24	0,12	0,70	+	0,84	0,68	0,59	0,88	0,67	0,51
	FREQ.	$\frac{12}{119} = 10,1\%$	$\frac{91}{227} = 40,1\%$	$\frac{11}{44} = 25,0\%$	$\frac{91}{113} = 8,5\%$	$\frac{16}{16} = 100,0\%$	$\frac{60}{68} = 88,2\%$	$\frac{42}{55} = 76,4\%$	$\frac{6}{8} = 75,0\%$	$\frac{9}{10} = 90,0\%$	$\frac{4}{5} = 80,0\%$	$\frac{342}{665} = 51,4\%$
SONIA 602(599)	PROB.	0,13	0,48	0,08	0,60	0,81	0,76	0,73	0,61	+	0,48	0,83
	FREQ.	$\frac{29}{58} = 50,0\%$	$\frac{189}{236} = 80,1\%$	$\frac{2}{6} = 33,3\%$	$\frac{60}{67} = 89,6\%$	$\frac{19}{20} = 95,0\%$	$\frac{78}{84} = 92,9\%$	$\frac{98}{106} = 92,5\%$	$\frac{12}{13} = 92,3\%$	$\frac{3}{3} = 100,0\%$	$\frac{7}{9} = 77,8\%$	$\frac{497}{602} = 82,5\%$
ILZENI 439 (374)	PROB.	-	0,06	-	0,61	0,55	0,81	0,72	+	0,44	-	0,62
	FREQ.	$\frac{0}{51} = 0,0\%$	$\frac{9}{152} = 5,9\%$	$\frac{0}{8} = 0,0\%$	$\frac{36}{60} = 60,0\%$	$\frac{10}{21} = 47,6\%$	$\frac{35}{43} = 81,4\%$	$\frac{65}{95} = 68,4\%$	$\frac{4}{4} = 100,0\%$	$\frac{1}{3} = 33,3\%$	$\frac{0}{2} = 0,0\%$	$\frac{160}{439} = 36,4\%$
GEORGE M. 394 (388)	PROB.	0,06	0,16	0,25	0,79	+	0,93	0,90	+	+	0,36	0,54
	FREQ.	$\frac{2}{51} = 3,9\%$	$\frac{12}{133} = 9,0\%$	$\frac{14}{30} = 46,7\%$	$\frac{41}{68} = 60,3\%$	$\frac{1}{1} = 100,0\%$	$\frac{52}{64} = 81,3\%$	$\frac{31}{39} = 79,5\%$	$\frac{4}{4} = 100,0\%$	$\frac{1}{1} = 100,0\%$	$\frac{1}{3} = 33,3\%$	$\frac{159}{394} = 40,4\%$

INFORMANTES		B	N	V	G (= L)	Z	W	K (= E)	Y	U	Q	INPUT
		Conte-Comem Vê-Võem Constrô-Constroem	Fala-Falam	Faz-Fazem Quer-Querem	Está-Estão Dá-Dão Vai-Vão	Foi-Foram	E - São	Falou-Falarum Comeu-Comeram	Trouxe-Trouxeram Fez-Fizeram Disse-Disseram	Veio-Vieram	Era-Eram	
MARLENE 376(375)	PROB. FREQ.	0,12 $\frac{3}{38} = 7,9\%$	0,10 $\frac{7}{129} = 5,4\%$	0,20 $\frac{2}{6} = 33,3\%$	0,78 $\frac{23}{40} = 57,5\%$	0,48 $\frac{1}{5} = 20,0\%$	0,85 $\frac{26}{41} = 63,4\%$	0,86 $\frac{70}{94} = 74,5\%$	0,87 $\frac{6}{7} = 85,7\%$	+	0,26 $\frac{2}{15} = 13,3\%$	0,57 $\frac{141}{376} = 37,5\%$
HENRIQUE 372(371)	PROB. FREQ.	0,18 $\frac{5}{36} = 13,9\%$	0,23 $\frac{35}{154} = 22,7\%$	0,23 $\frac{4}{19} = 21,1\%$	0,53 $\frac{29}{59} = 49,2\%$	0,75 $\frac{4}{6} = 66,7\%$	0,89 $\frac{23}{27} = 85,2\%$	0,92 $\frac{57}{64} = 89,1\%$	0,18 $\frac{1}{3} = 33,3\%$	+	0,46 $\frac{1}{3} = 33,3\%$	0,54 $\frac{160}{372} = 43,0\%$
VANILDA 305(295)	PROB. FREQ.	0,12 $\frac{18}{39} = 46,2\%$	0,41 $\frac{82}{117} = 70,1\%$	0,28 $\frac{18}{25} = 72,0\%$	0,68 $\frac{37}{44} = 84,1\%$	0,59 $\frac{4}{6} = 66,7\%$	0,61 $\frac{18}{22} = 81,8\%$	0,85 $\frac{39}{42} = 92,9\%$	+	-	-	0,56 $\frac{217}{305} = 71,1\%$
GEORGE C. 228(209)	PROB. FREQ.	- $\frac{0}{18} = 0,0\%$	0,50 $\frac{7}{74} = 9,5\%$	0,17 $\frac{5}{14} = 35,7\%$	0,54 $\frac{23}{44} = 52,3\%$	0,82 $\frac{6}{7} = 85,7\%$	0,77 $\frac{16}{20} = 80,0\%$	0,70 $\frac{33}{41} = 80,5\%$	0,83 $\frac{3}{4} = 75,0\%$	-	0,30 $\frac{2}{5} = 40,0\%$	0,67 $\frac{95}{228} = 41,7\%$
LIDIO 217(180)	PROB. FREQ.	- $\frac{0}{23} = 0,0\%$	0,17 $\frac{9}{85} = 10,6\%$	0,23 $\frac{1}{5} = 20,0\%$	0,82 $\frac{23}{36} = 63,9\%$	+	0,32 $\frac{6}{6} = 100,0\%$	0,88 $\frac{2}{4} = 50,0\%$	+	-	-	0,46 $\frac{42}{51} = 82,4\%$
CANTIDIO 151(100)	PROB. FREQ.	0,24 $\frac{2}{11} = 18,2\%$	0,51 $\frac{21}{43} = 48,8\%$	0,78 $\frac{8}{18} = 44,4\%$	0,81 $\frac{14}{18} = 77,8\%$	0,67 $\frac{2}{3} = 66,7\%$	+	+	+	+	0,10 $\frac{12}{12} = 100,0\%$	0,12 $\frac{3}{3} = 100,0\%$
PAULO 98 (66)	PROB. FREQ.	0,18 $\frac{2}{27} = 7,4\%$	0,27 $\frac{0}{3} = 0,0\%$	0,3 $\frac{4}{12} = 33,3\%$	0,4 $\frac{2}{5} = 40,0\%$	0,7 $\frac{7}{10} = 70,0\%$	0,81 $\frac{17}{21} = 81,0\%$	0,5 $\frac{1}{2} = 50,0\%$				0,33 $\frac{33}{98} = 33,7\%$
NATANAEL 57(28)	PROB. FREQ.	0,3 $\frac{0}{3} = 0,0\%$	0,21 $\frac{0}{21} = 0,0\%$	0,2 $\frac{1}{2} = 50,0\%$	0,4 $\frac{4}{8} = 50,0\%$	0,5 $\frac{2}{4} = 50,0\%$	0,68 $\frac{13}{19} = 68,4\%$					0,35 $\frac{20}{57} = 35,1\%$

TABELA 3 - CATEGORIZAÇÃO "NEAMAL"
GRUPOS SINTÁTICO E SEMÂNTICO - PROBABILIDADES E FREQUÊNCIAS

INFORMANTES		TIPO DE SUJEITO		POSIÇÃO DO SUJEITO			
		PLURAL	INDEFINIDO	ANTES	OCULTO	SEPARADO	DEPOIS
AMÁLGAMA DOS 14 INFORMAN- TES	PROB.	0,39	0,61	0,70	0,65	0,44	0,22
	FREQ.	$\frac{2382}{5072} = 47,0\%$	$\frac{386}{517} = 47,7\%$	$\frac{1528}{3118} = 49,0\%$	$\frac{1122}{2068} = 54,3\%$	$\frac{72}{204} = 35,3\%$	$\frac{46}{199} = 23,1\%$
ELVIRA	PROB.	0,42	0,58	0,78	0,57	0,49	0,19
	FREQ.	$\frac{353}{710} = 49,7\%$	$\frac{91}{134} = 67,9\%$	$\frac{304}{534} = 56,9\%$	$\frac{125}{257} = 48,6\%$	$\frac{11}{34} = 32,4\%$	$\frac{4}{19} = 21,1\%$
M.L.FAUSTINA	PROB.	0,34	0,66	0,71	0,61	0,34	0,34
	FREQ.	$\frac{300}{820} = 36,6\%$	$\frac{17}{21} = 81,0\%$	$\frac{190}{498} = 38,2\%$	$\frac{118}{309} = 38,2\%$	$\frac{4}{19} = 21,1\%$	$\frac{5}{15} = 33,3\%$
UBIRAJARA	PROB.	0,41	0,59	0,77	0,66	0,60	0,09
	FREQ.	$\frac{314}{622} = 50,5\%$	$\frac{28}{43} = 65,1\%$	$\frac{198}{388} = 51,0\%$	$\frac{126}{230} = 54,8\%$	$\frac{17}{35} = 48,6\%$	$\frac{1}{12} = 8,3\%$
SONIA	PROB.	0,30	0,70	0,78	0,72	0,58	0,07
	FREQ.	$\frac{423}{525} = 80,6\%$	$\frac{74}{77} = 96,1\%$	$\frac{296}{346} = 86,5\%$	$\frac{181}{212} = 85,4\%$	$\frac{14}{21} = 66,7\%$	$\frac{6}{23} = 26,1\%$
ILZENI	PROB.	0,30	0,70	0,59	0,58	0,54	0,30
	FREQ.	$\frac{146}{420} = 34,8\%$	$\frac{14}{19} = 73,7\%$	$\frac{87}{260} = 33,5\%$	$\frac{66}{156} = 42,3\%$	$\frac{4}{12} = 33,3\%$	$\frac{3}{11} = 27,3\%$
GEORGE M.	PROB.	0,20	0,80	0,68	0,60	0,34	0,38
	FREQ.	$\frac{140}{369} = 37,9\%$	$\frac{19}{25} = 76,0\%$	$\frac{81}{211} = 38,4\%$	$\frac{62}{151} = 51,2\%$	$\frac{8}{36} = 22,2\%$	$\frac{8}{26} = 30,8\%$
MARLENE	PROB.	0,19	0,81	0,44	0,71	0,79	0,13
	FREQ.	$\frac{115}{347} = 33,1\%$	$\frac{26}{29} = 89,7\%$	$\frac{44}{184} = 23,9\%$	$\frac{91}{163} = 55,8\%$	$\frac{2}{4} = 50,0\%$	$\frac{4}{25} = 16,0\%$
HENRIQUE	PROB.	0,33	0,67	0,64	0,65	0,62	0,16
	FREQ.	$\frac{128}{330} = 38,8\%$	$\frac{32}{42} = 76,2\%$	$\frac{74}{194} = 38,1\%$	$\frac{76}{148} = 52,1\%$	$\frac{4}{10} = 40,0\%$	$\frac{4}{20} = 20,0\%$
VANILDA	PROB.	0,48	0,52	0,76	0,75	0,27	0,23
	FREQ.	$\frac{180}{258} = 69,8\%$	$\frac{37}{47} = 78,7\%$	$\frac{108}{145} = 74,5\%$	$\frac{101}{134} = 75,4\%$	$\frac{5}{13} = 38,5\%$	$\frac{3}{13} = 23,1\%$
GEORGE C.	PROB.	0,14	0,86	0,79	0,71	0,22	0,28
	FREQ.	$\frac{77}{209} = 36,8\%$	$\frac{18}{19} = 94,7\%$	$\frac{46}{121} = 38,0\%$	$\frac{47}{90} = 52,2\%$	$\frac{1}{9} = 11,1\%$	$\frac{1}{8} = 12,5\%$
LÍDIO	PROB.	0,50	0,50	0,47	0,51	-	0,52
	FREQ.	$\frac{72}{188} = 38,3\%$	$\frac{12}{29} = 41,4\%$	$\frac{35}{92} = 38,0\%$	$\frac{45}{115} = 39,1\%$	$\frac{0}{1} = 0,0\%$	$\frac{4}{9} = 44,4\%$
CANTIDIO	PROB.	0,85	0,15	0,61	0,60	0,30	-
	FREQ.	$\frac{96}{141} = 68,1\%$	$\frac{3}{10} = 30,0\%$	$\frac{47}{72} = 65,3\%$	$\frac{48}{72} = 66,7\%$	$\frac{1}{2} = 50,0\%$	$\frac{3}{5} = 60,0\%$
PAULO	PROB.						
	FREQ.	$\frac{28}{91} = 30,8\%$	$\frac{5}{7} = 71,4\%$	$\frac{12}{44} = 27,3\%$	$\frac{20}{38} = 52,6\%$	$\frac{1}{5} = 20,0\%$	$\frac{0}{11} = 0,0\%$
NATANAEL	PROB.						
	FREQ.	$\frac{10}{42} = 23,8\%$	$\frac{10}{15} = 66,7\%$	$\frac{6}{29} = 20,7\%$	$\frac{14}{23} = 60,9\%$	$\frac{0}{3} = 0,0\%$	$\frac{0}{2} = 0,0\%$

TABELA 4 - CATEGORIZAÇÃO "NEAMAL"; NÚMERO DA ENTREVISTA - PROBABILIDADES E FREQUÊNCIAS

INFORMANTES		NÚMERO DA ENTREVISTA								
		1	2	3	4	5	6	7	8	9
ELVIRA	PROB. FREQ.	0,53 $\frac{79}{147} = 53,7\%$	0,55 $\frac{81}{143} = 56,6\%$	0,60 $\frac{66}{122} = 54,1\%$	0,26 $\frac{24}{76} = 31,6\%$	0,48 $\frac{51}{107} = 47,7\%$	0,53 $\frac{60}{119} = 50,4\%$	0,57 $\frac{83}{130} = 63,8\%$		
M.L. FAUSTINA	PROB. FREQ.	0,72 $\frac{46}{82} = 56,1\%$	0,61 $\frac{59}{116} = 50,9\%$	0,58 $\frac{43}{100} = 43,0\%$	0,40 $\frac{30}{117} = 25,6\%$	0,36 $\frac{48}{153} = 31,4\%$	0,41 $\frac{38}{115} = 33,0\%$	0,40 $\frac{53}{158} = 33,5\%$		
UBIRAJARA	PROB. FREQ.	0,36 $\frac{43}{96} = 44,8\%$	0,55 $\frac{40}{88} = 45,5\%$	0,44 $\frac{81}{144} = 56,3\%$	0,61 $\frac{86}{147} = 58,5\%$	0,52 $\frac{43}{98} = 43,9\%$	0,60 $\frac{23}{43} = 53,5\%$	0,53 $\frac{14}{28} = 50,0\%$	0,41 $\frac{12}{21} = 57,1\%$	
SONIA	PROB. FREQ.	0,61 $\frac{58}{99} = 84,1\%$	0,51 $\frac{94}{116} = 81,0\%$	0,32 $\frac{99}{129} = 76,7\%$	0,30 $\frac{39}{52} = 75,0\%$	0,54 $\frac{86}{98} = 87,8\%$	0,42 $\frac{47}{57} = 82,5\%$	0,78 $\frac{74}{81} = 91,4\%$		
ILZENI	PROB. FREQ.		0,58 $\frac{36}{98} = 36,7\%$		0,67 $\frac{17}{34} = 50,0\%$	0,53 $\frac{28}{63} = 44,4\%$	0,57 $\frac{14}{33} = 42,4\%$	0,53 $\frac{8}{27} = 29,6\%$	0,33 $\frac{9}{25} = 36,0\%$	0,31 $\frac{0}{64} = 0,0\%$
GEORGE M.	PROB. FREQ.		0,51 $\frac{30}{87} = 34,5\%$	0,37 $\frac{30}{82} = 36,6\%$		0,44 $\frac{60}{158} = 38,0\%$	0,67 $\frac{39}{67} = 58,2\%$			
MARLENE	PROB. FREQ.	0,67 $\frac{51}{99} = 51,5\%$	0,28 $\frac{15}{64} = 23,4\%$	0,31 $\frac{16}{72} = 22,2\%$	0,72 $\frac{19}{39} = 48,7\%$	0,64 $\frac{8}{13} = 61,5\%$	0,50 $\frac{12}{38} = 31,6\%$	0,39 $\frac{20}{51} = 39,2\%$		
HENRIQUE	PROB. FREQ.	0,58 $\frac{18}{31} = 58,1\%$	0,45 $\frac{49}{115} = 42,6\%$	0,36 $\frac{14}{62} = 22,6\%$	0,73 $\frac{20}{30} = 66,7\%$	0,31 $\frac{31}{81} = 38,3\%$	0,57 $\frac{28}{53} = 52,8\%$			
VANILDA	PROB. FREQ.		0,58 $\frac{85}{123} = 69,1\%$	0,27 $\frac{18}{38} = 47,4\%$	0,37 $\frac{23}{35} = 65,7\%$	0,79 $\frac{65}{78} = 83,3\%$	0,48 $\frac{26}{31} = 83,9\%$			
GEORGE C.	PROB. FREQ.	0,27 $\frac{12}{40} = 30,0\%$	0,76 $\frac{20}{41} = 48,8\%$	0,60 $\frac{23}{39} = 59,0\%$	0,52 $\frac{12}{36} = 33,3\%$	0,51 $\frac{15}{38} = 39,5\%$	0,10 $\frac{1}{11} = 9,1\%$	0,82 $\frac{12}{23} = 52,2\%$		
LIDIO	PROB. FREQ.	0,58 $\frac{8}{17} = 47,1\%$	0,48 $\frac{32}{51} = 62,7\%$	0,44 $\frac{17}{49} = 34,7\%$	0,50 $\frac{10}{29} = 34,5\%$		0,12 $\frac{9}{57} = 15,8\%$	0,89 $\frac{8}{14} = 57,1\%$		
CANTIDIO	PROB. FREQ.	0,38 $\frac{16}{18} = 88,9\%$	0,93 $\frac{11}{13} = 84,6\%$	0,29 $\frac{9}{21} = 42,9\%$	0,34 $\frac{12}{20} = 60,0\%$	0,41 $\frac{31}{45} = 68,9\%$		0,47 $\frac{20}{34} = 58,8\%$		
PAULO	PROB. FREQ.					0,28 $\frac{6}{21} = 28,6\%$				
NATANAEL	PROB. FREQ.									

TABELA 5 - CATEGORIZAÇÃO "NEOCODE" GRUPO MORFOLÓGICO
PROBABILIDADES E FREQUÊNCIAS

INFORMANTES		B	T	N	V	G	H	Z	W	J	K	U	Y	X	Q
		COME COMEM	SAI SAEM	FALA FALAM	DIZ DIZEM	VAI VÃO	DÁ DÃO	FOI FORAM	É SÃO	COMEU COMERAM	FALOU FALARAM	VEIO VINHERAM	TROUXE TROUXERAM	FEZ FIZERAM	ERA ERAM
14 INF. 5589	PROB. FREQ.	0,06 70 = 12,3 540	0,06 15 = 17,4 86	0,16 622 = 29,6 2099	0,25 124 = 43,4 286	0,54 234 = 60,9 384	0,59 269 = 70,1 384	0,67 95 = 72,5 131	0,79 457 = 82,3 555	0,75 149 = 78,4 190	0,80 581 = 82,2 707	0,72 25 = 73,5 34	0,92 28 = 93,3 30	0,77 70 = 86,4 81	0,21 29 = 35,4 82
ELVIRA 848 (827)	PROB. FREQ.	0,06 10 = 11,4 88	0,16 3 = 25,0 12	0,22 114 = 34,4 331	0,45 31 = 60,8 51	0,51 31 = 62,0 50	0,51 38 = 66,7 57	0,86 14 = 87,5 16	0,71 36 = 80,0 45	0,58 16 = 61,5 26	0,89 102 = 91,1 112	+	+	0,80 25 = 89,3 28	0,35 5 = 45,5 11
M.L.FAUSTINA 841 (819)	PROB. FREQ.	0,02 1 = 2,1 47	- 0 = 0,0 15	0,11 44 = 11,9 371	0,49 27 = 47,1 55	0,58 29 = 54,7 53	0,54 20 = 51,3 39	0,54 10 = 52,6 19	0,80 69 = 76,7 90	0,82 20 = 83,3 24	0,78 74 = 76,3 97	0,38 1 = 50,0 2	+	0,81 9 = 41,8 11	0,55 6 = 54,5 11
UBIRAJARA 665 (646)	PROB. FREQ.	0,06 11 = 10,2 108	0,05 1 = 9,1 11	0,29 91 = 40,1 227	0,14 11 = 25,0 44	0,72 41 = 75,6 51	0,77 60 = 83,3 72	+	0,87 60 = 88,2 68	0,54 9 = 64,3 14	0,82 33 = 80,5 41	0,90 9 = 90,0 10	+	0,39 3 = 60,0 5	0,72 4 = 80,0 5
SÔNIA 602 (598)	PROB. FREQ.	0,23 27 = 60,0 45	0,02 2 = 15,4 13	0,52 189 = 80,1 236	0,09 2 = 33,3 6	0,61 26 = 89,7 29	0,65 34 = 89,5 38	0,84 19 = 95,0 20	0,79 78 = 92,9 84	0,86 21 = 95,5 22	0,73 77 = 91,7 84	+	+	0,62 11 = 91,7 12	0,51 7 = 77,8 9
ILZENI 442 (374)	PROB. FREQ.	- 0 = 0,0 50	- 0 = 0,0 1	0,05 9 = 5,8 154	- 0 = 0,0 8	0,71 22 = 68,8 32	0,42 14 = 50,0 28	0,51 10 = 47,6 21	0,79 35 = 81,4 43	0,62 6 = 60,0 10	0,70 59 = 68,6 86	0,43 1 = 33,3 3	+	3 = 100,0 3	- 0 = 0,0 2
GEORGE M. 394 (340)	PROB. FREQ.	- 0 = 0,0 48	0,58 2 = 66,7 3	0,07 12 = 9,0 133	0,11 14 = 46,7 30	0,50 23 = 51,1 45	0,80 18 = 78,3 23	+	0,83 52 = 81,3 64	0,86 13 = 86,7 15	0,75 18 = 75,0 24	1 = 100,0 1		4 = 100,0 4	0,19 1 = 33,3 3
MARLENE 376 (368)	PROB. FREQ.	0,11 3 = 8,3 36	- 0 = 0,0 2	0,09 7 = 5,4 129	0,18 2 = 33,3 6	0,70 12 = 54,5 22	0,80 11 = 61,1 18	0,44 1 = 20,0 5	0,83 26 = 63,4 41	0,76 7 = 63,6 11	0,85 63 = 75,9 83	+	0,68 1 = 50,0 2	5 = 100,0 5	0,24 2 = 13,3 15
HENRIQUE 372 (366)	PROB. FREQ.	0,17 5 = 16,1 31	- 0 = 0,0 5	0,18 35 = 22,7 154	0,18 4 = 21,1 19	0,44 11 = 42,3 26	0,45 18 = 54,5 33	0,70 4 = 66,7 6	0,85 23 = 85,2 27	0,94 15 = 88,2 17	0,87 42 = 89,4 47	+	+	0,14 1 = 33,3 3	0,39 1 = 33,3 3
VANILDA 305 (295)	PROB. FREQ.	0,10 11 = 42,3 26	0,14 7 = 53,8 13	0,41 82 = 70,1 117	0,27 18 = 72,0 25	0,50 19 = 79,2 24	0,85 18 = 90,0 20	0,60 4 = 66,7 6	0,61 18 = 81,8 22	0,66 8 = 88,8 9	0,89 31 = 93,9 33			1 = 100,0 1	0 = 0,0 0 = 0,0 9
GEORGE C. 228 (209)	PROB. FREQ.	- 0 = 0,0 15	- 0 = 0,0 3	0,05 7 = 9,5 74	0,16 5 = 35,7 14	0,41 8 = 33,3 24	0,69 15 = 75,0 20	0,80 6 = 85,7 7	0,76 16 = 80,0 20	0,44 6 = 60,0 10	0,78 27 = 87,1 31	-	-	0,82 3 = 75,0 4	0,28 2 = 40,0 5
LIDIO 217 (180)	PROB. FREQ.	- 0 = 0,0 20	- 0 = 0,0 3	0,09 9 = 10,6 5	0,15 1 = 20,0 5	0,63 15 = 57,7 26	0,90 8 = 80,0 10	+	0,20 2 = 50,0 4	0,79 13 = 86,7 15	0,79 29 = 80,6 36	-	-	1 = 100,0 1	0 = 0,0 0 = 0,0 2
CANTIDIO 151 (94)	PROB. FREQ.	0,23 2 = 20,0 10	- 0 = 0,0 1	0,51 21 = 48,8 43	0,78 8 = 44,4 18	+	0,73 10 = 71,4 14	0,72 2 = 66,7 3	+	+	+	+	+	3 = 100,0 3	0,11 1 = 14,3 7
PAULO 98	PROB. FREQ.	- 0 = 0,0 15	- 0 = 0,0 3	- 2 = 7,4 27	- 0 = 0,0 3	- 2 = 40,0 5	- 2 = 28,6 7	- 2 = 40,0 5	- 7 = 70,0 10	- 6 = 85,7 7	- 11 = 78,6 14	-	0 = 0,0 0 = 0,0 1	1 = 100,0 1	-

TABELA 6 - CATEGORIZAÇÃO "NEOCODE"

GRUPOS SEMÂNTICO E POSICIONAL;

INPUT GLOBAL

INFORMANTES		P	I	A	O	S	D	INPUT
14 INF.	PROB. FREQ.	0,39 $\frac{2382}{5072} = 47,0$	0,62 $\frac{386}{517} = 74,7$	0,71 $\frac{1528}{3118} = 49,0$	0,65 $\frac{1122}{2068} = 54,3$	0,44 $\frac{72}{204} = 35,3$	0,22 $\frac{46}{199} = 23,1$	0,53 $\frac{2768}{5589} = 49,5$
ELVIRA	PROB. FREQ.	0,42 $\frac{353}{710} = 49,7$	0,58 $\frac{91}{134} = 67,9$	0,76 $\frac{306}{538} = 56,9$	0,55 $\frac{125}{257} = 48,6$	0,47 $\frac{11}{34} = 32,4$	0,23 $\frac{4}{19} = 21,1$	0,49 $\frac{446}{848} = 52,6$
M.L.FAUSTINA	PROB. FREQ.	0,34 $\frac{300}{820} = 36,6$	0,67 $\frac{17}{21} = 81,0$	0,71 $\frac{190}{498} = 38,2$	0,61 $\frac{118}{309} = 38,2$	0,34 $\frac{4}{11} = 21,1$	0,34 $\frac{5}{15} = 33,3$	0,51 $\frac{317}{841} = 37,7$
UBIRAJARA	PROB. FREQ.	0,43 $\frac{314}{622} = 50,5$	0,57 $\frac{28}{43} = 65,1$	0,79 $\frac{198}{388} = 51,0$	0,68 $\frac{126}{230} = 54,8$	0,62 $\frac{17}{35} = 48,6$	0,07 $\frac{1}{12} = 8,3$	0,41 $\frac{342}{665} = 51,4$
SONIA	PROB. FREQ.	0,26 $\frac{423}{525} = 80,6$	0,74 $\frac{74}{77} = 96,1$	0,78 $\frac{296}{346} = 85,5$	0,72 $\frac{181}{212} = 85,4$	0,58 $\frac{14}{21} = 66,7$	0,07 $\frac{6}{23} = 26,1$	0,83 $\frac{497}{602} = 82,6$
ILZENI	PROB. FREQ.	0,30 $\frac{146}{420} = 34,8$	0,70 $\frac{14}{19} = 73,7$	0,57 $\frac{87}{262} = 33,2$	0,58 $\frac{66}{156} = 42,3$	0,60 $\frac{4}{12} = 33,3$	0,27 $\frac{3}{12} = 25,0$	0,66 $\frac{160}{442} = 36,2$
GEORGE M.	PROB. FREQ.	0,19 $\frac{140}{369} = 38,0$	0,82 $\frac{19}{25} = 76,0$	0,66 $\frac{81}{211} = 38,4$	0,60 $\frac{62}{121} = 51,2$	0,36 $\frac{8}{36} = 22,2$	0,38 $\frac{8}{26} = 30,8$	0,76 $\frac{159}{394} = 40,4$
MARLENE	PROB. FREQ.	0,19 $\frac{115}{347} = 33,1$	0,81 $\frac{26}{29} = 89,7$	0,44 $\frac{44}{184} = 23,9$	0,70 $\frac{91}{163} = 55,8$	0,79 $\frac{2}{4} = 50,0$	0,13 $\frac{4}{25} = 16,0$	0,62 $\frac{141}{376} = 37,5$
HENRIQUE	PROB. FREQ.	0,33 $\frac{128}{330} = 38,8$	0,67 $\frac{32}{42} = 76,2$	0,66 $\frac{74}{194} = 38,1$	0,66 $\frac{78}{148} = 52,7$	0,62 $\frac{4}{10} = 40,0$	0,14 $\frac{4}{20} = 20,0$	0,60 $\frac{160}{372} = 43,0$
VANILDA	PROB. FREQ.	0,48 $\frac{180}{258} = 69,8$	0,52 $\frac{37}{47} = 78,7$	0,77 $\frac{108}{145} = 74,5$	0,77 $\frac{101}{134} = 75,4$	0,25 $\frac{5}{13} = 38,5$	0,22 $\frac{3}{13} = 23,1$	0,55 $\frac{217}{305} = 71,1$
GEORGE C.	PROB. FREQ.	0,14 $\frac{77}{209} = 36,8$	0,86 $\frac{18}{19} = 94,7$	0,78 $\frac{46}{121} = 38,0$	0,69 $\frac{47}{90} = 52,2$	0,20 $\frac{1}{9} = 11,1$	0,34 $\frac{1}{8} = 12,5$	0,70 $\frac{95}{228} = 41,7$
LIDIO	PROB. FREQ.	0,52 $\frac{72}{188} = 38,3$	0,48 $\frac{12}{29} = 41,4$	0,44 $\frac{35}{92} = 38,0$	0,52 $\frac{45}{115} = 39,1$	- $\frac{0}{1} = 0,0$	0,54 $\frac{4}{9} = 44,4$	0,61 $\frac{84}{217} = 38,7$
CANTIDIO	PROB. FREQ.	0,83 $\frac{96}{141} = 68,1$	0,17 $\frac{3}{10} = 30,0$	0,62 $\frac{47}{72} = 65,3$	0,56 $\frac{48}{72} = 66,7$	0,33 $\frac{1}{2} = 50,0$	- $\frac{3}{5} = 60,0$	0,13 $\frac{99}{151} = 65,6$
PAULO	PROB. FREQ.	0,28 $\frac{28}{91} = 30,8$	0,57 $\frac{5}{7} = 71,4$	0,44 $\frac{12}{44} = 27,3$	0,52 $\frac{20}{38} = 52,6$	0,20 $\frac{1}{5} = 20,0$	0,00 $\frac{0}{11} = 0,0$	0,33 $\frac{33}{98} = 33,7$

TABELA 7 - CATEGORIZAÇÃO "NEOCODE"

NÚMERO DA ENTREVISTA;

PROBABILIDADES E FREQUÊNCIAS

INFORMANTES		0	1	2	3	4	5	6	7	8
ELVIRA	PROB. FREQ.		0,53 $\frac{79}{147} = 53,7$	0,55 $\frac{82}{146} = 56,2$	0,61 $\frac{66}{122} = 54,1$	0,25 $\frac{24}{76} = 31,6$	0,48 $\frac{51}{107} = 47,7$	0,54 $\frac{61}{120} = 50,8$	0,57 $\frac{83}{130} = 63,8$	
M.L.FAUSTINA	PROB. FREQ.		0,73 $\frac{46}{82} = 56,1$	0,61 $\frac{59}{116} = 50,9$	0,57 $\frac{43}{100} = 43,0$	0,40 $\frac{30}{117} = 25,6$	0,36 $\frac{48}{153} = 31,4$	0,41 $\frac{38}{115} = 33,0$	0,40 $\frac{53}{158} = 33,5$	
UBIRAJARA	PROB. FREQ.		0,34 $\frac{43}{96} = 44,8$	0,55 $\frac{40}{88} = 45,5$	0,44 $\frac{81}{144} = 56,3$	0,61 $\frac{86}{147} = 58,5$	0,52 $\frac{43}{98} = 43,9$	0,60 $\frac{23}{43} = 53,5$	0,56 $\frac{14}{28} = 50,0$	0,39 $\frac{12}{21} = 57,1$
SONIA	PROB. FREQ.		0,64 $\frac{58}{69} = 84,1$	0,52 $\frac{94}{116} = 81,0$	0,31 $\frac{99}{129} = 76,7$	0,29 $\frac{39}{52} = 75,0$	0,53 $\frac{86}{98} = 87,8$	0,40 $\frac{47}{57} = 82,5$	0,78 $\frac{74}{81} = 91,4$	
ILZENI	PROB. FREQ.	0,30 $\frac{13}{64} = 20,3$		0,60 $\frac{36}{100} = 36,0$		0,70 $\frac{17}{34} = 50,0$	0,52 $\frac{28}{63} = 44,4$	0,54 $\frac{14}{34} = 41,2$	0,32 $\frac{25}{87} = 28,7$	0,53 $\frac{27}{60} = 45,0$
GEORGE M.	PROB. FREQ.			0,54 $\frac{30}{87} = 34,5$	0,34 $\frac{30}{82} = 36,6$		0,46 $\frac{60}{158} = 38,0$	0,66 $\frac{39}{67} = 58,2$		
MARLENE	PROB. FREQ.		0,67 $\frac{51}{99} = 51,5$	0,29 $\frac{15}{64} = 23,4$	0,31 $\frac{16}{72} = 22,2$	0,71 $\frac{19}{39} = 48,7$	0,64 $\frac{8}{13} = 61,5$	0,49 $\frac{12}{38} = 31,6$	0,39 $\frac{20}{51} = 39,2$	
HENRIQUE	PROB. FREQ.		0,59 $\frac{18}{31} = 58,1$	0,44 $\frac{49}{115} = 42,6$	0,36 $\frac{14}{62} = 22,6$	0,73 $\frac{20}{30} = 66,7$	0,31 $\frac{31}{81} = 38,3$	0,57 $\frac{28}{53} = 52,8$		
VANILDA	PROB. FREQ.			0,58 $\frac{85}{123} = 69,1$	0,27 $\frac{18}{38} = 47,4$	0,35 $\frac{23}{35} = 65,7$	0,80 $\frac{65}{78} = 83,3$	0,48 $\frac{26}{31} = 83,9$		
GEORGE C.	PROB. FREQ.		0,29 $\frac{12}{40} = 30,0$	0,73 $\frac{20}{41} = 48,8$	0,56 $\frac{23}{29} = 59,0$	0,55 $\frac{12}{36} = 33,3$	0,54 $\frac{15}{38} = 39,5$	0,10 $\frac{1}{11} = 9,1$	0,82 $\frac{12}{23} = 52,2$	
LIDIO	PROB. FREQ.		0,57 $\frac{8}{17} = 47,1$	0,48 $\frac{32}{51} = 62,7$	0,46 $\frac{17}{49} = 34,7$	0,54 $\frac{10}{29} = 34,5$	0,10 $\frac{9}{57} = 15,8$	0,88 $\frac{8}{14} = 57,1$		
CANTIDIO	PROB. FREQ.		0,42 $\frac{16}{18} = 88,9$	0,92 $\frac{11}{13} = 84,6$	0,22 $\frac{9}{21} = 42,9$	0,35 $\frac{12}{20} = 60,0$	0,43 $\frac{31}{45} = 68,9$		0,51 $\frac{20}{31} = 58,8$	
PAULO	PROB. FREQ.						0,28 $\frac{6}{21} = 28,6$			

TABELA 8 - CATEGORIZAÇÃO "NEAMAL" GRUPO MORFOLÓGICO
MODELO ADITIVO (ANOVA)

INFORMANTES	GRÃ MÉDIA	CONCORDÂNCIA VERBAL - CATEGORIZAÇÃO NEAMAL - MODELO ANOVA										
		B	N	V	G	Z	W	J	Y	U	Q	B MORF.
Elvira	52,6	- 37,2	- 18,8	5,8	10,5	39,0	29,2	35,0	37,0	40,7	- 7,3	0,86
M. L. Faustina	37,9	- 36,7	- 25,9	8,6	15,2	17,5	39,5	41,6	48,5	16,8	16,1	0,93
Ubirajara	51,6	- 42,0	- 11,2	- 29,2	29,1	49,0	39,0	25,3	18,8	41,3	26,0	0,93
Sonia	82,5	- 30,9	- 2,0	- 52,0	6,4	10,3	9,9	8,7	6,9	30,5	0,4	0,60
Ilzeni	36,4	- 35,9	- 30,6	- 35,7	24,9	11,6	46,0	30,3	58,0	5,8	- 36,3	0,94
George M.	40,3	- 33,7	- 30,1	- 15,6	21,1	23,6	45,0	39,4	58,5	63,6	- 6,5	0,85
Marlene	37,6	- 28,2	- 28,6	- 16,2	21,6	- 16,2	30,0	31,0	35,9	69,6	- 25,8	0,75
Henrique	43,1	- 26,8	- 20,1	- 23,3	6,3	29,6	44,8	42,7	- 5,3	57,2	- 9,4	0,85
Vanilda	77,0	- 25,8	- 1,7	- 1,6	14,2	- 1,2	14,6	20,6	27,6	----	- 65,4	0,66
George C.	41,6	- 35,8	- 32,3	- 9,1	13,0	45,2	38,6	32,4	33,3	- 8,7	9,9	0,80
Lidio	38,7	- 39,0	- 28,3	- 16,9	25,4	61,3	11,0	43,7	60,0	- 34,1	- 40,0	0,89
* Cantidio	65,7	- 29,4	- 19,4	- 6,1	9,7	- 2,9	31,4	37,3	30,1	51,5	- 55,2	0,88
Paulo	33,6	- 33,3	- 27,7	- 27,1	8,2	- 0,7	29,0	47,3	24,6	----	----	0,81
14 inf.	49,6	- 35,8	- 20,0	- 9,6	16,6	23,8	33,2	32,7	34,4	27,2	- 14,7	0,72

* Significância de F = 0,999 para os efeitos principais

TABELA 9 - CATEGORIZAÇÃO "NEAMAL"; GRUPOS SEMÂNTICO E POSICIONAL
MODELO ADITIVO (ANOVA)

INFORMANTES	CONCORDÂNCIA VERBAL - CATEGORIZAÇÃO NEAMAL - MODELO ANOVA								
	D	S	O	A	β POS.	P	I	β SEM.	R MÚLTIP.
Elvira	- 31,5	- 11,8	- 9,8	6,6	0,31	- 2,1	11,0	0,16	0,92
M. L. Faustina	- 26,6	- 17,7	- 4,1	4,0	0,18	- 0,6	23,7	0,12	0,95
Ubirajara	- 37,4	- 10,7	- 4,5	4,8	0,23	- 0,8	11,0	0,09	0,94
Sonia	- 54,1	- 10,1	0,2	4,1	0,52	- 1,4	9,8	0,18	0,84
Ilzeni	- 23,5	- 2,9	2,0	- 0,1	0,11	- 1,0	21,9	0,14	0,96
George M.	- 10,5	- 14,3	- 0,9	4,3	0,16	- 3,0	44,8	0,30	0,89
Marlene	- 24,9	12,6	9,1	- 5,0	0,25	- 2,3	27,1	0,21	0,89
Henrique	- 29,6	- 1,0	1,8	1,7	0,22	- 2,1	16,8	0,19	0,92
Vanilda	- 39,5	- 37,8	3,6	3,6	0,43	0,4	- 2,2	0,03	0,80
George C.	- 30,2	- 26,0	- 0,5	4,3	0,22	- 22,7	30,2	0,25	0,90
Lidio	- 0,9	- 10,7	1,0	- 1,0	0,03	0,3	- 2,1	0,02	0,89
* Cantidio	- 18,6	- 20,3	0,4	1,5	0,13	2,6	- 38,9	0,32	0,92
Paulo	- 28,4	- 26,5	8,8	2,5	0,32	0,6	- 7,5	0,05	0,87
14 Inf.	- 29,4	- 13,8	- 0,5	3,1	0,18	- 1,3	12,7	0,11	0,86

2.5 - Conclusões teóricas

Do ponto de vista teórico a regra de concordância verbal é de sumo interesse porque ela é um exemplo de uma regra sintática que se encontra atualmente em processo de mudança. Através de seu estudo, conseguimos captar ao vivo alguns dos fatores que determinam a mudança lingüística, possibilitando um melhor entendimento da própria história do desenvolvimento das línguas. Além de nos revelar desta maneira os caminhos pelos quais uma mudança se difunde, podemos também tentar isolar os fatores que deram início ao processo. Estes dois aspectos são conhecidos na lingüística sob os nomes de "implementação" e "actuação". (Chen, Matthew Y. e William S. Y. Wang, "Sound Change: Actuation and Implementation" - Language, vol. 51, nº 2, junho 1975).

Os betas parciais produzidos pela MCA da análise da variância nos mostram que, de longe, o grupo mais poderoso de fatores determinantes da frequência de concordância é o morfológico, o que corresponde ao fato desse grupo possuir uma gama ampla (de 0,06 a 0,92 em "neo-code") no modelo de Sankoff (probabilidades).

Os resultados numéricos para o amálgama de 14 informantes nos três modelos para a categorização "neamal" são repetidos abaixo.

A tabela 10 nos mostra que a falta de concordância ocorre mais freqüentemente precisamente naqueles casos em que a diferença material entre o singular e o plural é menor, ou menos saliente. Os números revelam um salto notável entre as categorias V e G (=L), que é precisamente o ponto na escala onde a acentuação vai incidir sobre a parte da palavra que contém o elemento diferenciador, i. e. , a desinência. Na primeira parte da tabela notamos a maior falta de concordância na classe B, onde a única diferença entre singular e plural é a mera nasalização da vogal final inacentuada ([kômi] | [kômĩ]). Já na classe N essa diferença aumenta um pouco, uma vez que se apresenta, além da nasalização, ou uma ditongação ([fála] | [fálãw]) ou uma mudança da qualidade vocálica ([fala] | [falũ] ou, às vezes, [fálu]). Na classe V a diferença torna-se mais saliente do que na classe anterior pois se trata da presença ou ausência de um segmento inteiro ([d'iz] | [d'izĩ] ou, às vezes, [d'izi]).

Como já notamos, ao passar de V para G (=L) há um salto bastante grande nos números porque nesta segunda categoria cai o acento sobre a parte da palavra em que se localiza a diferença entre o singular e o plural. Nas classes Z (foi/foram) e W (é/são) aumenta essa diferença fônica e, concomitantemente, a concordância. Na classe W fundem-se raiz e desinência e, quanto ao material fônico segmental, chegamos à maior diferenciação possível: não há nenhum segmento em comum entre é e são. No entanto, a concordância na classe K (=E) é do mesmo nível que na classe W porque naquela classe incide o acento sobre a desinência que, destacada da raiz, fica bem saliente. Na classe y as

duas formas se distinguem por uma mudança de acento, mas este facto parece ser de pouca importância.

Em suma, chegamos à conclusão de que quanto mais saliente for a diferença fônica entre singular e plural, mais provável será a concordância, i. e., a falta de concordância ocorre tanto mais quanto menos ela se faz sentir.

TABELA Nº 11 - Efeito de homogeneização paradigmática

C L A S S E	PROB	EFEITO AJUSTADO	FREQUÊNCIA
N fala/falam	0,17	- 20,0%	29,6%
Q era/eram	0,22	- 14,7%	35,4%

Na tabela 11 apresentamos os resultados relativos à hipótese de uma influência paradigmática. Como se vê, os três modelos refletem a existência desse efeito, já que o fator Q era/eram concorda ligeiramente mais do que N, a classe à qual pertence na hierarquia de saliência morfológica. Isto indica a necessidade de se adicionar aos grupos de fatores um de parentesco paradigmático em estudos futuros da concordância verbal.

Verificamos resultados paralelos aos do grupo morfológico no caso da variável posicional. (V. Tabela nº 12). A categoria menos saliente é aquela em que o sujeito segue ao seu verbo. Neste caso, o elemento determinante da concordância segue ao elemento determinado, fazendo com que a falta de concordância seja me-

nos óbvia. No caso do sujeito separado há uma distância física, com a correspondente demora temporal, entre determinante e determinado. O caso mais saliente é o do sujeito pré-posto, quando o sujeito plural antecede imediatamente ao verbo candidato à concordância.

Além das três categorias relativas à posição do sujeito em sua sentença, incluímos no grupo posicional o caso do sujeito dito "oculto", i. e., fisicamente ausente da sentença. Este caso é diferente dos outros, pois aqui compete ao falante fazer uma ligação mental entre o verbo e um elemento localizado fora do domínio da integração sintática, que opera com sentenças, segundo o modelo proposto em 3.1.2.1, página 53 . Assim sendo, entram em jogo fatores de ordem funcional, tendentes a aumentar a concordância para facilitar as ligações extra-sentenciais.

Tabela Nº 12 - Grupo posicional - amálgama da categorização "neamal"

C L A S S E	DESCRIÇÃO	PROB (VARBRJL)	EFEITO AJUSTA- DO (ANOVA)	FREQUÊNCIA (SWAMINC)
A) Sujeito realizado				
D	Sujeito pós-posto	0,22	-24,4	23,1%
S	Sujeito separado	0,44	-13,8	35,3%
A	Sujeito pré-posto	0,70	3,1	49,0%
B) Sujeito não-realizado				
O	Sujeito oculto	0,65	-0,5 $\beta = 0,18$	54,3%

Finalmente, chegamos à variável relativa ao tipo de sujeito. À primeira vista esta variável parece ser de grande importância, sendo que os efeitos não ajustados são de -2,6% para sujeitos plurais e 25% para impessoais. No entanto, ao ajustar pelos efeitos dos outros fatores, aqueles efeitos caem para -1,3% e 12,7%, respectivamente, porque na nossa amostra a categoria impessoal ocorre mais com sujeito oculto e com verbos das categorias K e Y.

De fato, o tipo de sujeito que estamos chamando de "impessoal" é quase sempre semanticamente plural (como em construíram um edifício novo), ou se refere a um grupo composto de mais de um indivíduo (como em, tiraram o feriado, onde se entende "o governo", etc). Por outro lado, no grupo que rotulamos "plural", a pluralidade ao invés de ser uma característica inerente, é um mero traço accidental (em os meninos chegaram não há pluralidade inerente; podia ter chegado um só menino). Talvez a ligação mais íntima do "impessoal" com a idéia de pluralidade explique a maior freqüência de ocorrência de verbos plurais com aquela categoria e o fato de a falta da marca de plural ser tão saliente nesse caso (cp. dizem que vai chover - diz que vai chover). De qualquer maneira, esse é o grupo de fatores que menos influência tem na concordância.

TABELA Nº 13 - Grupo semântico - Amálgama da categorização "neamal"

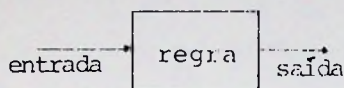
C L A S S E	DESCRIÇÃO	PROB (VARBRUL)	EFETO AJUSTADO (ANOVA)	FREQUÊNCIA (SWAMINC)
P	plural	0,39	-1,3%	47,0%
I	impessoal	0,61	12,7%	74,7%
			$\beta = 0,11$	

Resumindo, pois, concluímos que, dada uma certa forma verbal num certo contexto, a realização ou a ausência de concordância não é um fenômeno aleatório. Pelo contrário, a possibilidade desta ocorrência concordar é determinada por um princípio básico: quanto menos saliente for a diferença entre singular e plural, mais provável será a falta de concordância. A saliência de que se trata é, no entanto, um conceito geral: engloba aspectos tão diversos como diferenciação física entre formas, posição relativa dos elementos determinante e determinado, grau de inerência da noção semântica em jogo.

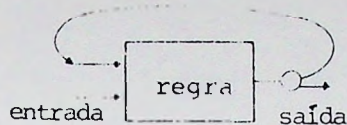
É possível que a noção da saliência possa ser estendida a fatores extra-linguísticos. Em geral, é de se supor que em conversas descontraídas entre amigos deva baixar o nível de atenção à linguagem usada, e assim a saliência de quaisquer não-aplicações das regras da gramática internalizada do falante. Na nossa amostra, tentamos testar esse fator social, variando tanto o local das entrevistas quanto o entrevistador. Houve também uma tentativa, nem sempre bem sucedida, de obter gravações

ã revelia. De um modo geral, podemos dizer que não obtivemos os resultados esperados: os informantes exibiram as mais diversas reações a situações sociais semelhantes. Por exemplo, no caso da gravação ã revelia (entrevista número 6), o informante Lídio teve o seu mais baixo nível de concordância ($p = 0,12$; freqüência = 15,8%) como seria de se esperar, enquanto o informante Ubirajara teve, nestas mesmas circunstâncias, seu nível mais alto ($p = 0,60$; freq = 53,5%). Da mesma sorte, da primeira para a segunda entrevista o informante George C. aumentou notavelmente (de $p = 0,27$ para $p = 0,76$, ou, em percentagem, de 30,0% para 48,8%), enquanto a informante Marlene baixou (de $p = 0,67$ para $p = 0,28$, ou freq = 51,5% para freq. = 23,4%), e a informante Elvira ficou quase inalterada ($p = 0,53$ e 0,55; freq = 53,7% e 56,6%). Em vista desses resultados, temos de abandonar, pelo momento, a tentativa de estender a idéia de saliência ã dimensão social. (Cf. tabelas 4 e 7).

Até aqui provamos que no modelo do funcionamento sincrônico da gramática deve ser introduzido o conceito de saliência. Uma regra gramatical será aplicada mais ou menos dependendo da saliência dos efeitos provocados. Do ponto de vista da gramática gerativa e outros modelos formais atuais, isto é um resultado duplamente surpreendente. Em primeiro lugar, prova que, além da sua entrada, ou descrição estrutural, a aplicação de uma regra pode, ao nível do desempenho, depender igualmente da sua saída;



Modelo clássico de
regras gramaticais



Modelo proposto de
regras gramaticais

O modelo proposto encontra paralelos computacionais aproximados no sistema "end-around carry", às vezes usado na unidade aritmética. Na própria lingüística, as chamadas regras globais, introduzidas pela corrente da semântica gerativa, também fazem a saída depender de certas de suas próprias características, mas de uma maneira bem mais abstrata. Em segundo lugar, os nossos resultados focalizam a importância da noção de oposição, definida em termos de fatores da superfície. Essa noção, central na lingüística estrutural, veio a ser relegada ao esquecimento devido à orientação abstrata ao nível explicatório na gramática gerativa.

No plano sincrônico, já tivemos ocasião de notar que a incidência mais baixa de concordância se dá na classe B, onde a única marca de plural é uma nasalização da vogal final inacentuada. Este fato por si só já sugere que, no plano diacrônico, o fenômeno deve ter se iniciado nesta mesma classe morfológica. De fato, entre os nossos informantes, aqueles que atingem os níveis mais altos de concordância são a Sônia, a Vanilda e o Cantídio (82,5%, 71,1%, 65,0%, 65,9%, respectivamente) e são precisamente esses informantes que exibem o maior distanciamento entre a classe B e o seu vizinho mais próximo, a classe N (0,13/0,48, 0,12/0,41, 0,24/0,51, respectivamente (Cf. Tabela 2)). Por outro

lado, os três informantes do nível mais baixo de concordância (exceção feita de Paulo e Natanael, cujos dados não foram suficientemente numerosos para rodar em VARBRUL) foram Ilzeni, Marlene, e M. L. Faustina (36,4%, 37,5%, 37,9%) e esses informantes mostram uma distância bem pequena entre B e N (0,00/0,06, 0,12/0,10, 0,02/0,12), como, aliás, também o fazem todos os informantes com uma percentagem global perto de 40% ou menos. Essas observações revelam que nas etapas iniciais do processo de perda da concordância verbal a classe B fica bem na vanguarda dos não-concordantes, enquanto que, à medida que o processo vai se difundido através da língua toda, a classe B perde o extremismo de sua posição especial. De fato, a perda da nasalização final inacentuada é um fenômeno bem mais amplo na língua portuguesa, estendendo-se, na língua falada, a todas as classes gramaticais (cf. homem [õmi] , ontem [õci], virgem [virgi], Miriam [Míria] , etc) É um fenômeno conhecido dos dois lados do Atlântico, em dialetos antigos e modernos. Assim, concluímos que a actuação da mudança em direção a um sistema sem concordância verbal foi fundamentalmente fonológica, enquanto que a sua implementação se deu através de uma difusão no eixo da saliência, sendo a principal coordenada a morfológica. Esta visão do fenômeno mostra-se bem de acordo com a visão geral da mudança sintática como um fenômeno de superfície (proposta em Naro 1971, 1976) já que, segundo esta última, as mudanças se originam em contextos que admitem mais de uma análise sintática sem alteração na cadeia de superfície. Assim sendo, a mudança se introduz no ponto zero de saliência e vai se difundido em contextos cada vez mais salientes.

2.6 - Conclusões Práticas

Numa tentativa de ensinar ao Mobralense o uso categórico da regra de concordância, de acordo com a norma culta, deve ser colocada mais ênfase naqueles aspectos em que seu dialeto mais difere daquela norma. Assim, nos verbos, os exercícios deverão insistir mais nos regulares no presente e no imperfeito e menos nas formas mais diferenciadas do perfeito; em relação à posição do sujeito, os exercícios serão de preferência com os sujeitos pós-postos e aqueles distanciados do verbo, como era de se esperar. No entanto, o caso do sujeito em outra oração que não aquela que contém o verbo não precisa ser enfatizado.

Em geral, recomendaríamos como princípio didático que as explicações e os primeiros exercícios aproveitassem os contextos mais altos na hierarquia de saliência, já que nestes casos o dialeto do Mobralense está mais próximo da norma que se pretende ensinar. Passada essa fase inicial de explicação, os treinos devem focalizar os pontos de discrepância, de acordo com a hierarquia.

3 - COMPLEXIDADE SINTÁTICA

3.1 - O problema da avaliação

3.1.1 - Fundamentação teórica

No estudo da complexidade sintática, optamos por nos atermos à estrutura superficial das sentenças. A justificativa teórica de tal opção nos vem de duas áreas: a psicolinguística e a teoria gramatical.

Em psicolinguística, houve uma fase de experimentações nas quais se fizeram tentativas de relacionar o desempenho no processamento das sentenças com a distância derivacional entre a estrutura superficial e a suposta estrutura profunda. Apesar de alguns aparentes sucessos iniciais, a hipótese foi desconfirmada por experimentos posteriores, e os resultados dos experimentos iniciais puderam ser reinterpretaados em termos de mera estrutura superficial. As evidências da realidade psicológica das estruturas profundas propostas pelas análises linguísticas são poucas e frágeis. Para uma revisão do estado atual da matéria, vide The Psychology of Language, de J. Fodor, T. G. Bever e M. F. Garrett (New York: Mc Graw-Hill, 1974).

Em teoria gramatical, a direção que vem tomando a linha da semântica interpretativa (também conhecida como teoria lexicalista ou teoria standard estendida), inspirada nos mais recentes trabalhos de Chomsky, Jackendoff e numerosos outros, está tornando claro que muitas das generalizações gramaticais que vinham sendo

feitas através do uso de regras transformacionais podem ser feitas pelo aproveitamento adequado do léxico, que era uma parte da descrição lingüística mal explorada até recentemente. Atribuindo ao léxico a responsabilidade de dar conta de numerosas sinônimas construcionais, a estrutura sintática pode ser mantida bastante próxima da forma superficial das sentenças, e possivelmente dispensado o nível da estrutura profunda e as transformações. A orientação teórica acima esquematizada pode ser exemplificada através dos seguintes estudos: a "Remarks on Nominalizations", Noam Chomsky, Readings in English Transformational Grammar, R. Jacobs e P. Rosenbaum, eds. (Waltham, Mass, Ginn and Co; 1970. Pp.124-221). b. Semantic Interpretation in Generative Grammar, Ray Jackendoff (Cambridge, MIT Press, 1972). c. "Conditions on Rules of Grammar", Noam Chomsky, Linguistic Analysis, vol. II, nº 4, 1976.

d. "The analysis of passives", Robert Freidin, Language, vol. 51, nº 1, 1976, Pp. 384-405.

e. "Constructions Superlatives et Structures Apparentes", Philippe Barbaud, Linguistic Analysis, vol. 2, nº 2, 1976. Pp. 125-174.

f. Semantic Conditions on Surface Structure, Robert Wilson Fiengo, tese doutoral, Massachusetts Institute of Technology, 1974.

3.1.2 - A construção de uma medida de complexidade sintática

3.1.2.1 - Modelo de desempenho da integração sintática

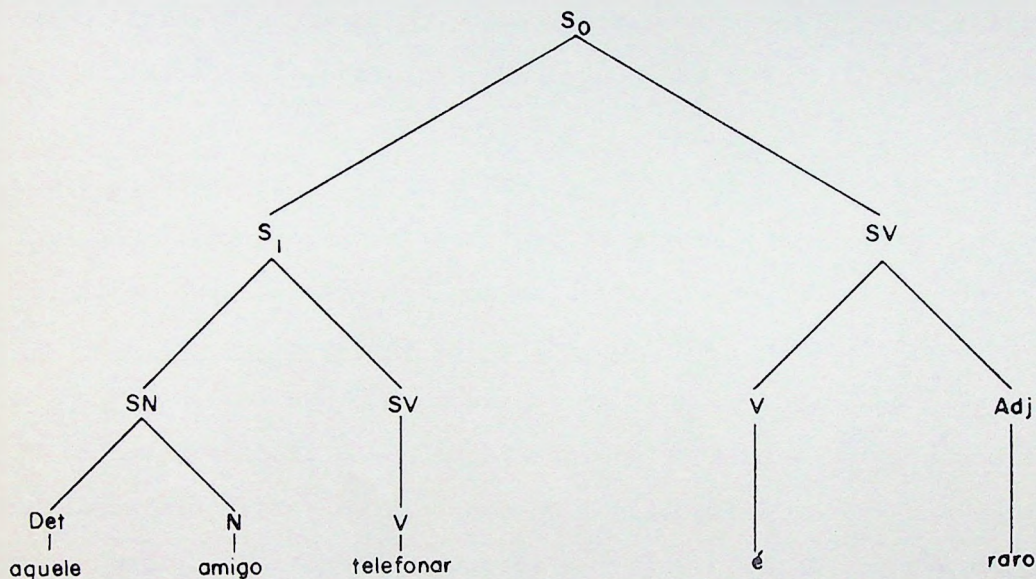
Com base nas justificativas teóricas sumarizadas no item anterior, partimos do pressuposto de que a complexidade sintática poderia ser medida diretamente na estrutura sintática superficial das sentenças.

O método que construímos para quantificá-la encontra apoio em dois modelos de desempenho da integração sintática. O primeiro, desenvolvido por Victor Yngue em "A Model and a Hypothesis for Language Structure" (em Proceedings of the American Philosophical Society 104, 1960, pp. 444-466), discute o conceito de profundidade de encaixe, aventando a hipótese de que o processamento de nódulos encaixados à esquerda e ao centro (com referência ao diagrama em árvore representativo da estrutura sintática) difere do processamento de nódulos encaixados à direita. A diferença reside na maior exigência imposta à memória a curto prazo para o processamento dos constituintes de esquerda, uma vez que estes exigem o registro cumulativo dos vários constituintes em processamento simultâneo. Um ponto de vista muito semelhante é o proposto por J. Kimball em "Seven Principles of Surface Structure Parsing in Natural Language" (em Cognition 2, 1973, pp. 15-48). Pensando na produção ou compreensão de uma sentença como um cálculo que se processa linearmente da esquerda para a direita e de cima para baixo na árvore que representa a cadeia lingüística, Kimball propõe os conceitos-chave de nóduo aberto e nóduo fechado, e define nóduo fechado como aquele cujo cálculo atingiu a etapa de processamento do seu último (da esquerda para a direita) nóduo

constituente .

Para ilustração, vejamos a aplicação desses conceitos ao processamento do cálculo da sentença i, abaixo diagramada:

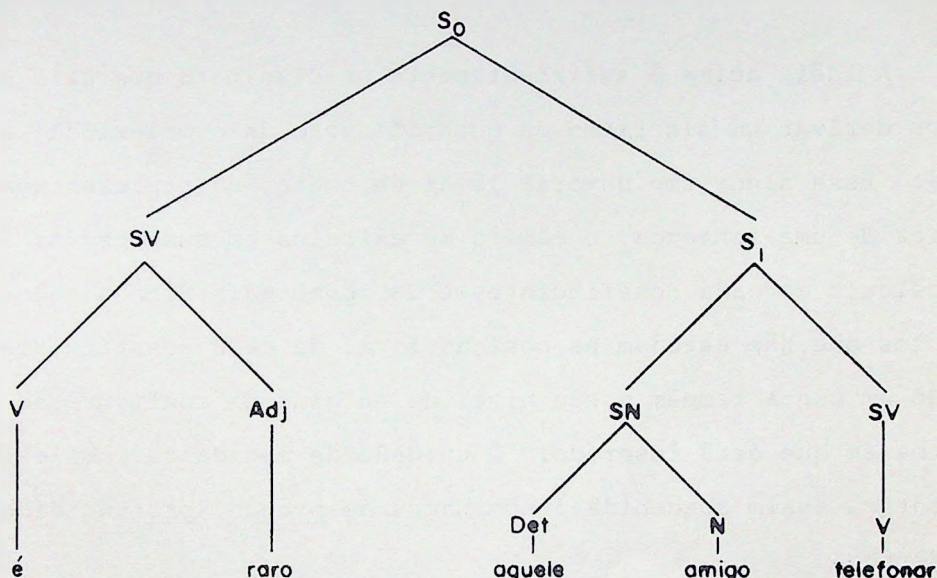
i. Aquele amigo telefonar é raro.



Ao enunciarmos o determinante aquele, três nódulos estão simultaneamente abertos: SN, S₁ e S₀. Atingida a enunciação do nome amigo, sendo este o último nódulo constituinte SN, o cálculo de SN está fechado; continuam abertos os nódulos S₁ e S₀. No momento de enunciar o SV telefonar, sendo este o último constituinte de S₁, S₁ é um nódulo fechado, mas ainda resta aberto o cálculo de S₀. A enunciação do verbo é corresponde um só nódulo aberto, o SV que o domina, uma vez que S₀, atingido o processamento de SV, seu último constituinte, já está fechado. Finalmente, a enunciação do adjetivo raro correspondem só nódulos fechados, uma vez que ele é o constituinte final de um constituinte final.

Comparemos o processamento acima descrito com o da sentença ii, semanticamente equivalente a i.

ii. É raro aquele amigo telefonar.



A enunciação do verbo é correspondem dois nódulos abertos: SV e S₀. O adjetivo raro tem, quando do seu processamento neste caso, um nódulo aberto dominando-o, S₀, pois SV atingiu o processamento do último nódulo-filho. O processamento do determinante aquele corresponde a dois nódulos abertos (SN e S₁), o do nome amigo a um nódulo aberto (S₁), e o do verbo telefonar a zero nódulos abertos.

A idéia fundamental tanto no artigo de Yngve quanto no de Kimball é a de que, em termos das exigências de memória imediata impostas ao organismo no desempenho da tarefa de produção ou percepção, o fator estrutural que contribui para a complexidade sintática é a ramificação à esquerda, pois esta é a configuração da qual resulta uma sobreposição de nódulos abertos, ou seja, processamentos a serem

feitos cumulativamente.

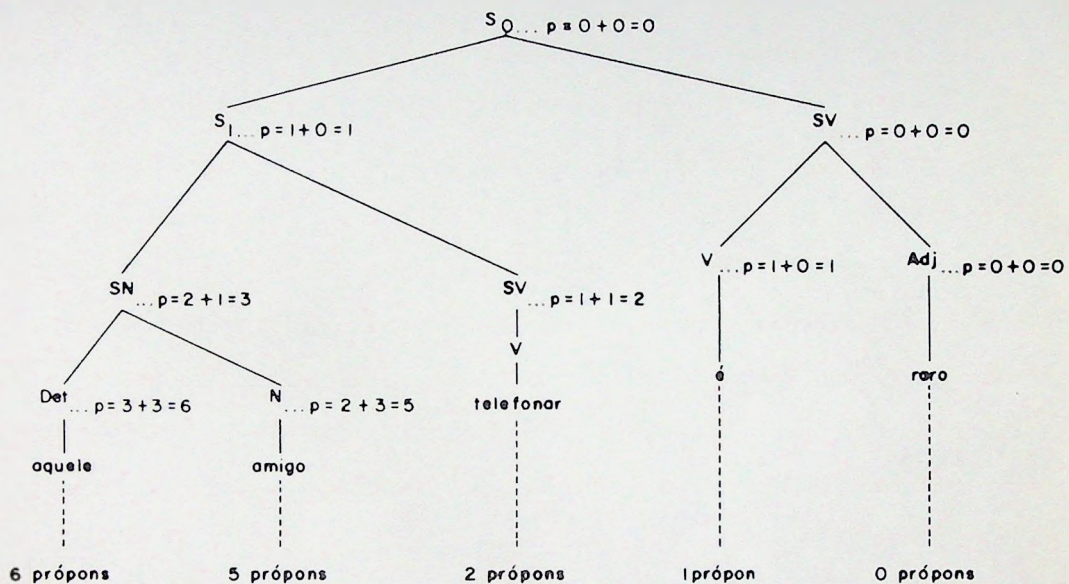
3.1.2.2 - O método de quantificação

A idéia acima é suficientemente precisa para que dela pudésemos derivar um algoritmo de quantificação da complexidade sintática. Esse algoritmo deveria levar em conta, na representação sintática de uma sentença, o número de cálculos em suspenso na altura do cálculo de cada constituinte, e dar peso mais alto a todos os nódulos que não estejam na posição final de cada constituinte, levando em conta também o seu nível de encaixe na configuração sintática em que está inserido. A unidade de medida de complexidade sintática assim concebida denominamos de própon (profundidade ponderada).

A complexidade sintática total de uma sentença é a soma do número de própons de todas as palavras que a constituem. O peso em própons de cada palavra é uma quantia obtida aditivamente e cumulativamente, no percurso dos ramos da árvore desde o topo até a palavra em avaliação, atribuindo-se a cada nódulo do percurso um peso que consiste na soma de duas parcelas: o número de nódulos abertos que o dominam mais o número de própons atribuído ao seu nódulo-pai.

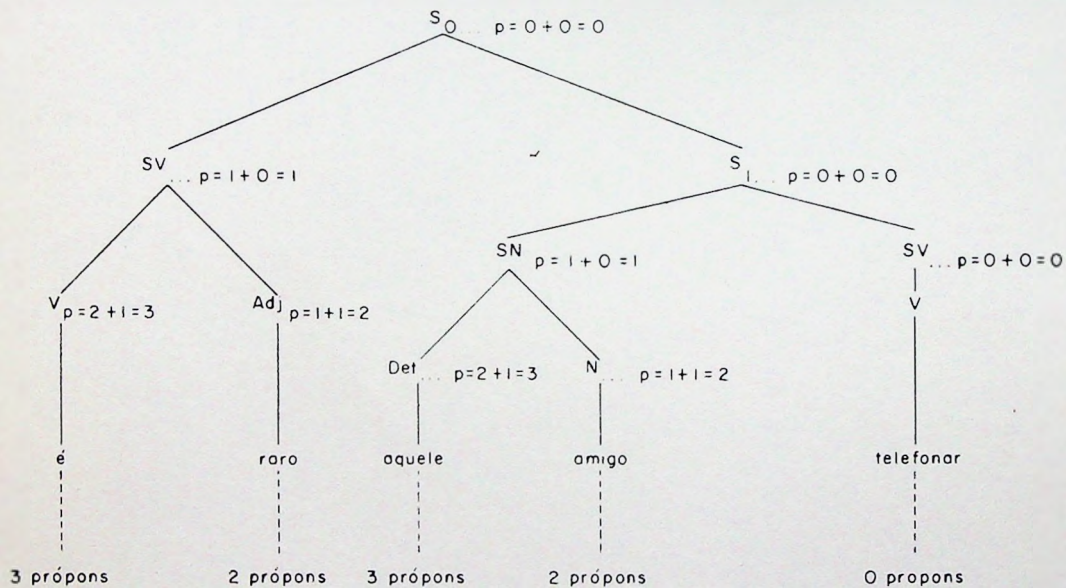
Para maior clareza, redesenharemos as sentenças i e ii, apresentando o cálculo do seu peso em própons. (Em cada nódulo, indicamos com p o cálculo parcial do peso até aquele nódulo)

i.



Peso total da sentença: $6 + 5 + 2 + 1 + 0 = 14$ própons

ii.



Peso total da sentença: $3 + 2 + 3 + 2 + 0 = 10$ própons

Vê-se que, fundamentalmente, o algoritmo leva em conta, na ponderação de cada nódulo, a carga de peso que ele herda da posição em que se insere e mais a carga de nódulos que estão mantidos em aberto em seu próprio processamento. Portanto, cada palavra terá, cumulativamente, peso tanto maior quanto mais abaixo e à esquerda na árvore ela se localize. Esse método de contagem é uma tentativa de traduzir quantitativamente a concepção acima descrita do processamento da computação mental da sentença.

3.1.2.3 - Comprovação do método de quantificação

Considerando-se que o método de quantificação em própons da complexidade sintática propõe-se a medir um cálculo mental efetivamente realizado pelos falantes, ele deve poder ser empiricamente validado pelas intuições lingüísticas desses mesmos falantes. Oito tipos de comprovação serão dados a seguir.

Abaixo estão alguns exemplos de sentenças que se equivalem semanticamente, mas diferem em suas configurações sintáticas, para que o leitor possa confrontar os seus próprios julgamentos de complexidade sintática com os pesos atribuídos a essas sentenças através do método acima exposto.

Primeiro exemplo: Aproveitemos as sentenças i e ii de 3.1.2.2.

- | | |
|------------------------------------|------------|
| i. Aquele amigo telefonar é raro. | 14 própons |
| ii. É raro aquele amigo telefonar. | 10 própons |

A sentença i, com a oração subordinada subjetiva à esquerda do predicado da principal foi avaliada como mais complexa do que

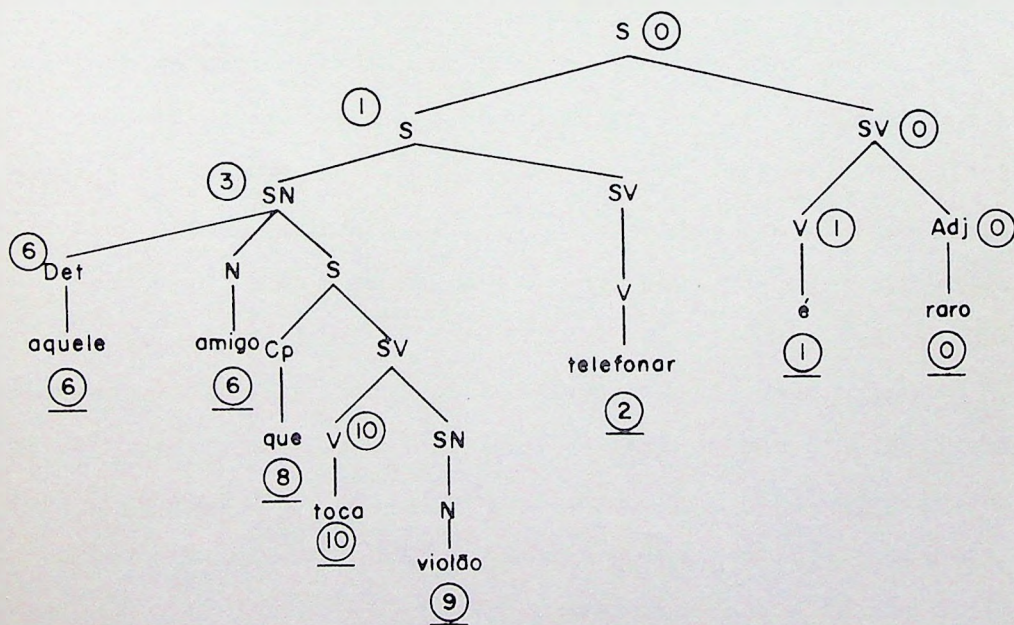
a sentença ii, cuja oração subjetiva vem à direita da principal. E, de fato, as avaliações numérica e intuitiva parecem ter boa correspondência.

Segundo exemplo: Vamos expandir as sentenças i e ii, acrescentando uma sentença encaixada ao sujeito da oração subjetiva:

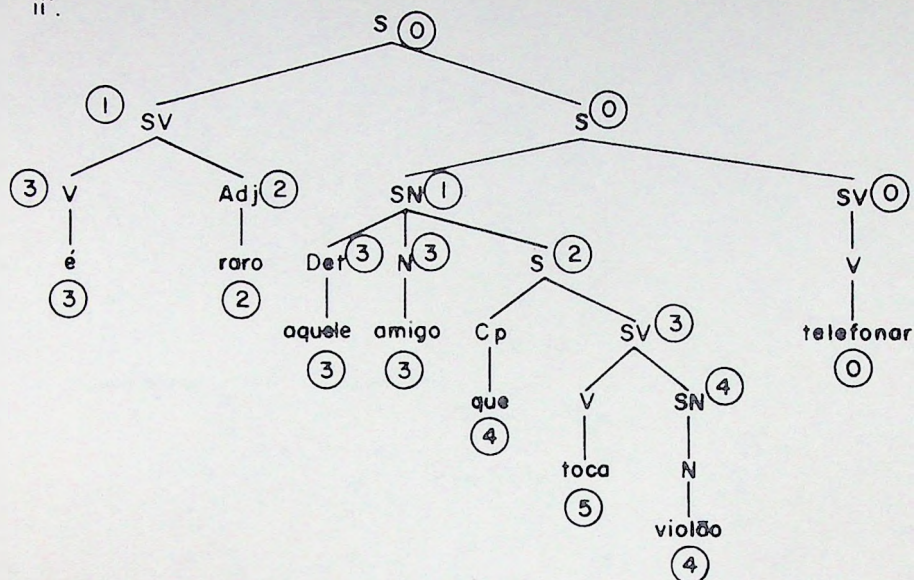
i'. Aquele amigo que toca violão telefonar é raro

ii'. É raro aquele amigo que toca violão telefonar

Peso total: 42 prôpons



ii'.



Note-se que entre o peso de i (14 prôpons) e o peso de ii (10 prôpons), a razão é de 1,4, e entre o peso de i' (42 prôpons) e o peso de ii' (24 prôpons), a razão é de 1,7. Compare-se agora essa razão numérica com uma avaliação introspectiva: se i é mais difícil do que ii, i' não o é ainda mais, em relação a ii'?

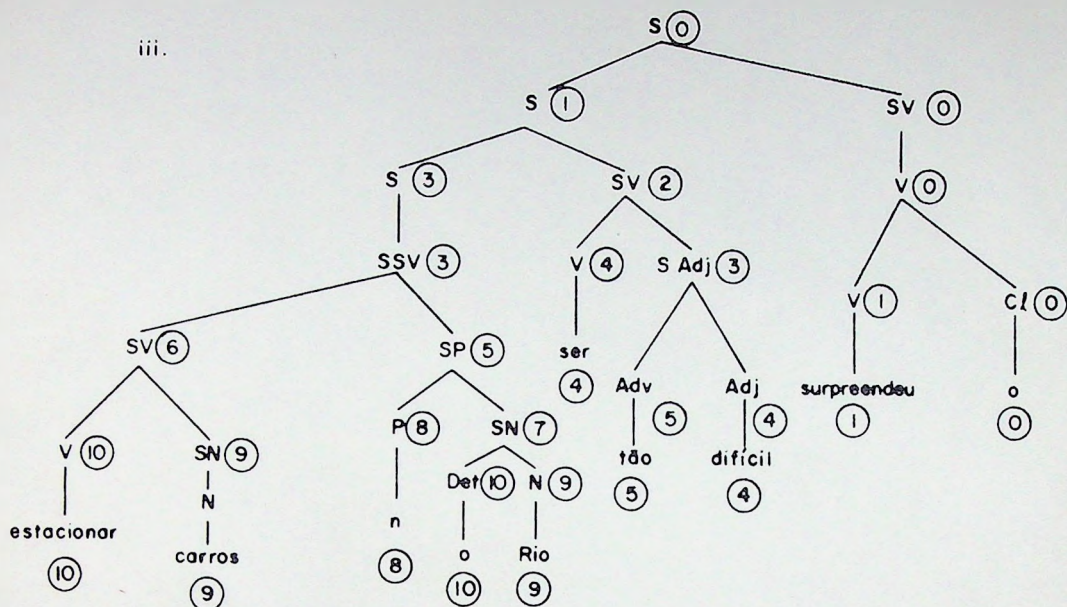
Terceiro exemplo: Comparem-se as sentenças abaixo:

- iii. Estacionar carros no Rio ser tão difícil surpreendeu-o. (60)
- iv. Ser tão difícil estacionar carros no Rio surpreendeu-o. (51)
- v. Surpreendeu-o estacionar carros no Rio ser tão difícil. (31)
- vi. Surpreendeu-o ser tão difícil estacionar carros no Rio. (22)

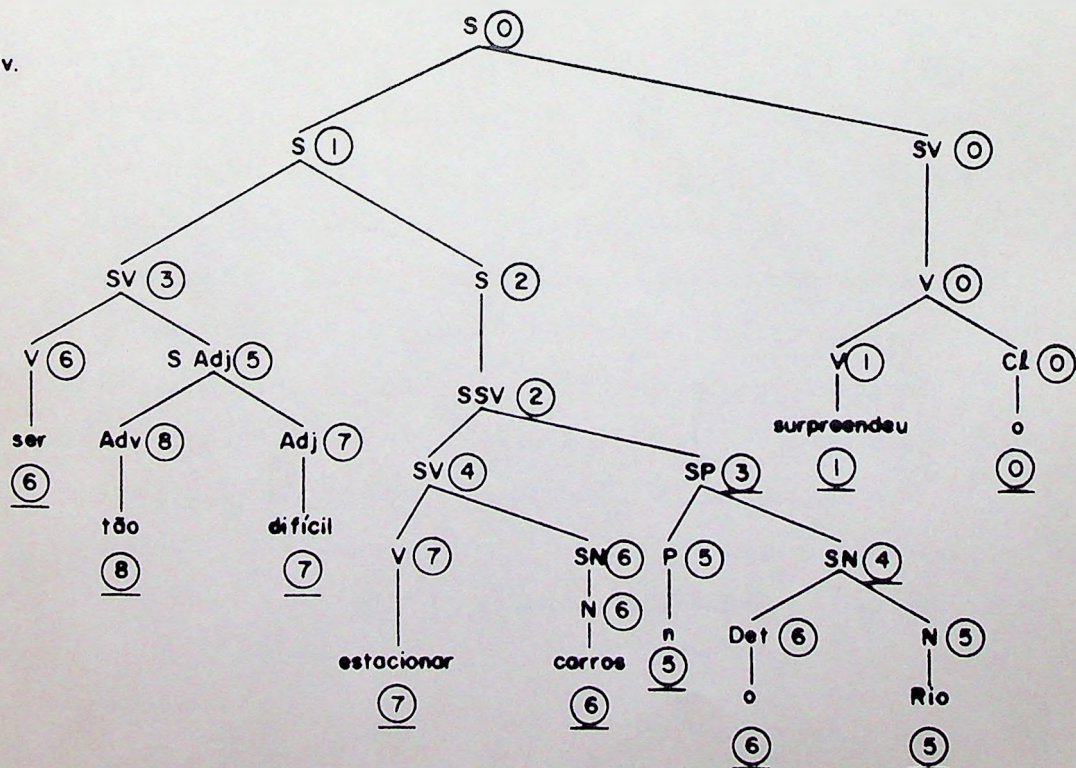
Intuitivamente, a gradação de complexidade das sentenças acima parece corresponder à ordem em que foram dadas. Vejamos como o nosso método de quantificação as gradua:

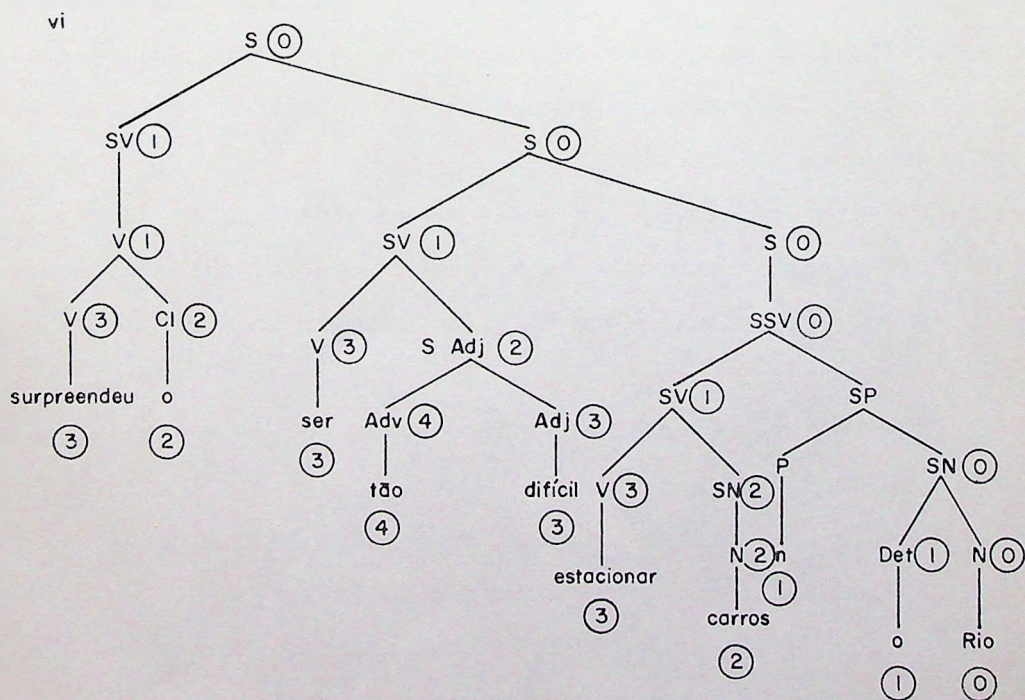
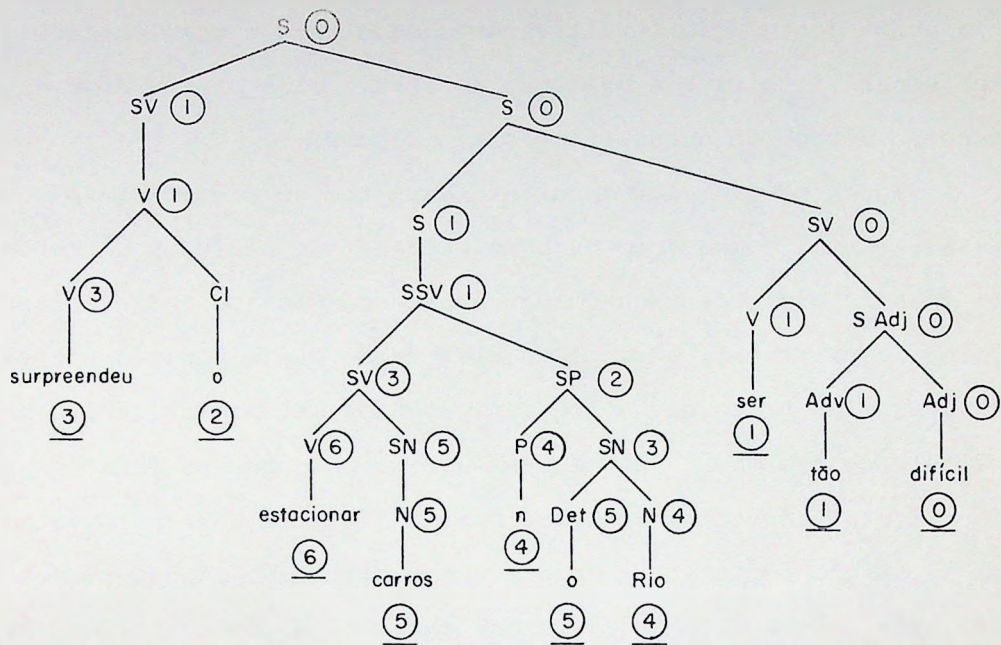
peso total: 24 prôpons

iii.



iv.





Mais uma vez, podemos verificar que há uma boa correspondência entre a gradação intuitivamente dada para a complexidade das sentenças iii a vi e a gradação do peso a elas atribuído segundo o nosso método de quantificação em própons.

Todos os exemplos dados até aqui têm em comum o fato de consistirem numa comparação de complexidades de períodos contendo orações subordinadas subjetivas à esquerda ou à direita dos predicados da principal, e em todos eles vemos que a posição à direita é favorecida tanto por nossos julgamentos intuitivos quanto pelo cálculo de própons. Vale a pena acrescentar que as frequências de uso das duas alternativas também são nitidamente mais altas para a posição à direita - embora essa afirmativa esteja sendo feita com base em observações assistemáticas, i. e. sem contagens em amostras reais e estatisticamente validadas.

Quarto exemplo: Também neste caso, excluída a busca de efeitos estilísticos especiais de ênfase, os julgamentos de relativa complexidade das sentenças abaixo são corroborados pela avaliação dos seus pesos em própons.

vii. O secretário enviou um telex ao novo embaixador da França

vii'. O secretário enviou ao novo embaixador da França um telex.

viii. O secretário enviou um telex provocador ao novo embaixador da França.

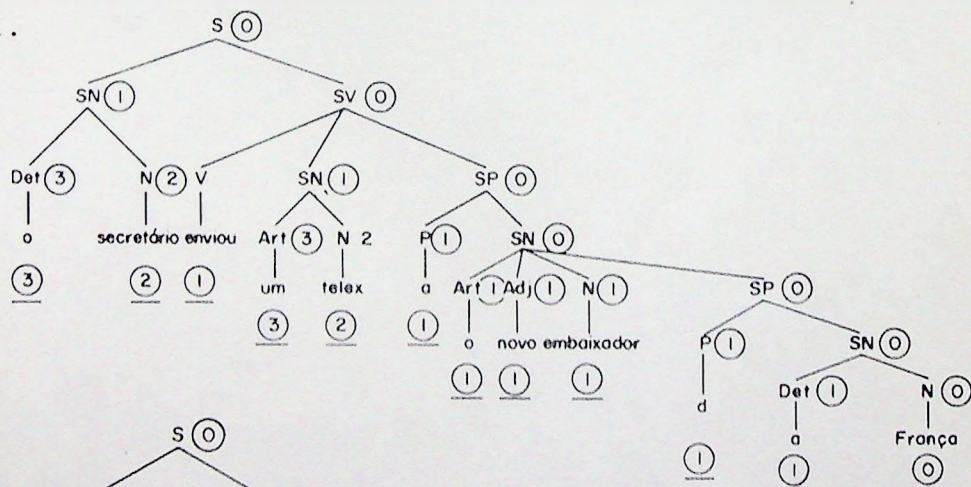
viii'. O secretário enviou ao novo embaixador da França um telex provocador.

ix. O secretário enviou um telex provocador de uma grave questão internacional ao novo embaixador da França.

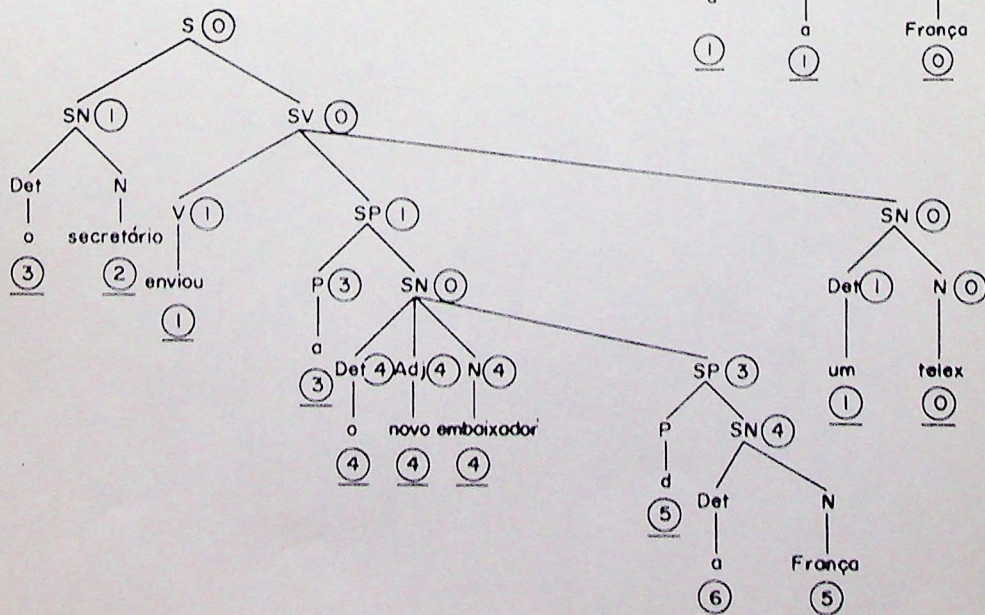
ix'. O secretário enviou ao novo embaixador da França um telex provocador de uma grave questão internacional.

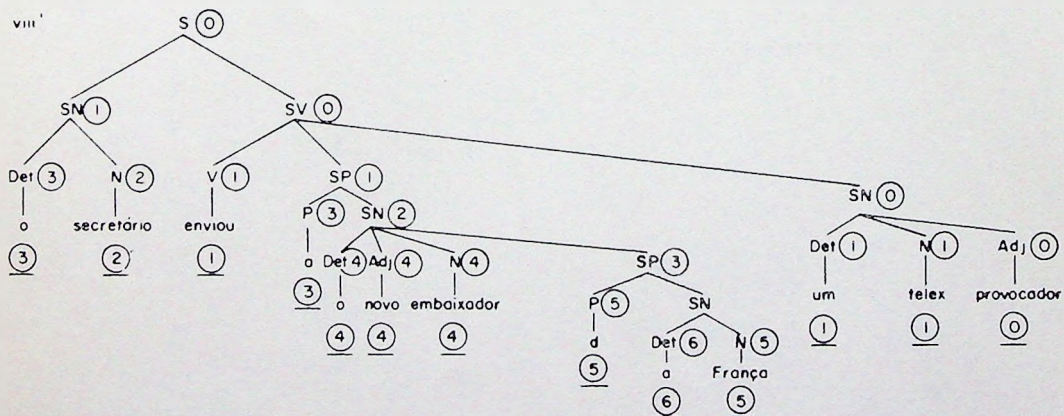
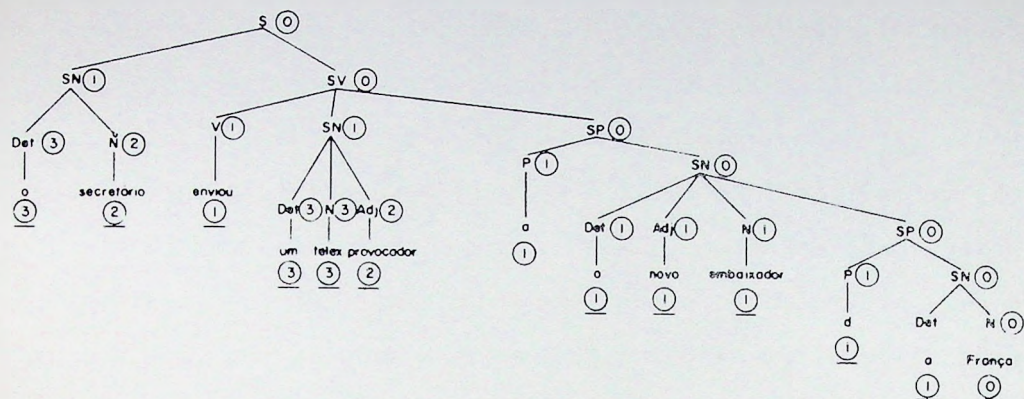
Observe-se que, nos exemplos acima, as sentenças sinônimas de cada par diferem uma da outra apenas pela respectiva ordenação do objeto direto e objeto indireto. Observe-se, porém, que no par vii e vii' damos preferência a vii, com a ordem objeto direto-objeto indireto. Entre viii e viii' essa preferência continua; porém menos acentuada. Já entre ix e ix' damos preferência a ix', onde a ordem é objeto indireto - objeto direto. A contagem de prótons dos três pares, que daremos a seguir, vem ao encontro desses julgamentos, e nos fornece um princípio que os explica.

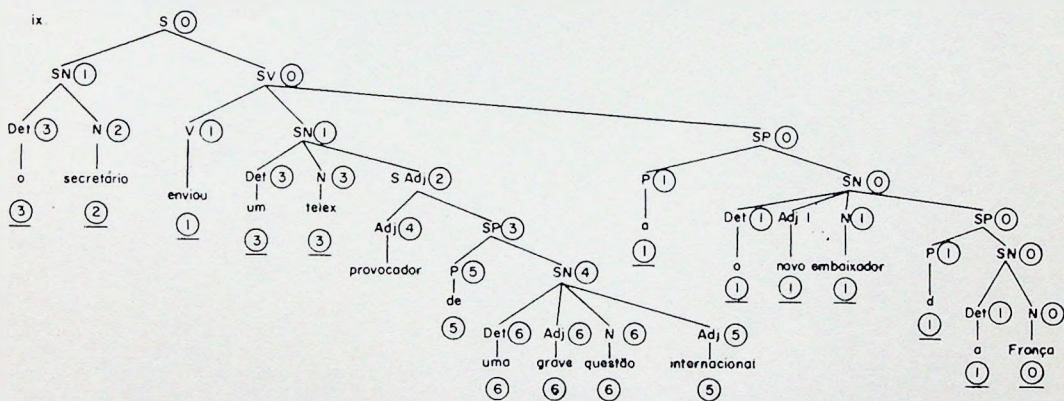
vii.

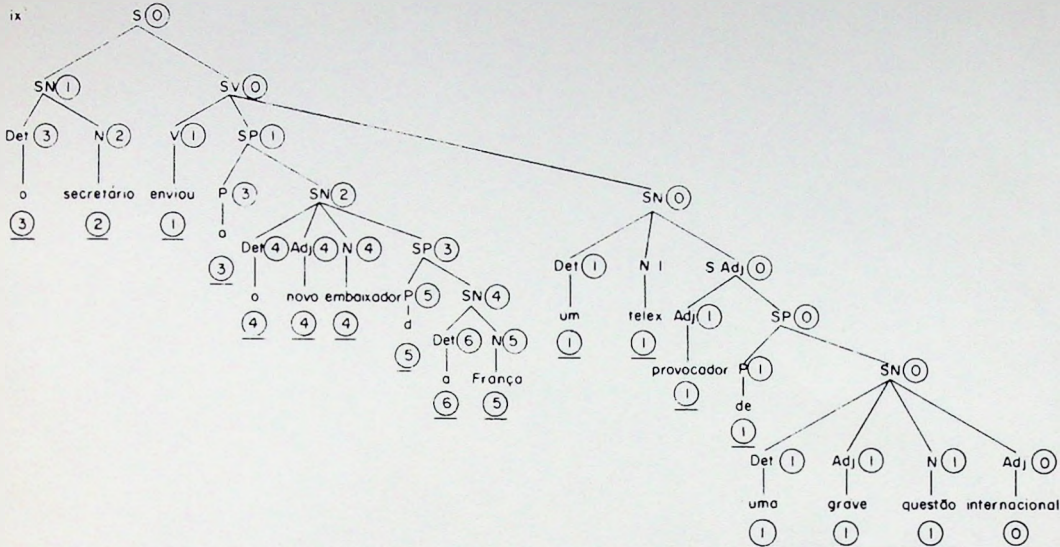


vii'.









Note-se que uma descrição dos julgamentos de grau de complexidade relativa dessas sentenças pode ser feita simplesmente em termos de uma contagem de palavras: são preferidas as ordenações de sintagmas nas quais os de maior número de palavras estão à direita. Contudo, tal descrição carece de qualquer explicação plausível, ao passo que a descrição baseada no mecanismo processamental tenta não apenas a descrição dos fatos, mas a descrição explicada dos fatos.

Quinto exemplo: Neste exemplo mostraremos a correspondência existente entre o grau de desiderabilidade de certas vírgulas facultativas e a quantificação da complexidade sintática, segundo o cálculo de própons das sentenças com e sem tais vírgulas.

Antes, vide à pg. 108 como foram arborizados e pesados os sintagmas coordenados.

A sentença x abaixo apareceu num noticiário do Jornal do Brasil de outubro de 76, com a vírgula assinalada dentro do círculo pontilhado. Note-se que a ausência dessa vírgula em nada altera o sentido básico do texto, e que a sua omissão não cria agramaticalidade. Porém ela parece bastante desejável, em termos da maior "leveza" que dá ao texto. Agora confrontem-se com a sentença x as sentenças xi e xii, as quais diferem de x por uma gradativa diminuição da sentença original. Atente-se novamente nas vírgulas assinaladas, e verificar-se-á que a sua presença é cada vez mais dispensável, tornando-se mesmo indesejável no caso xii.

x. A substituição dessa estratégia por uma cooperação militar mais estreita com a OTAN é a tese defendida pelo Chefe do Estado Maior das Forças Armadas da França, em artigo na revista Défense Nationale.

xi. ... é a tese defendida pelo general Guy Mery, em artigo na revista Défense Nationale.

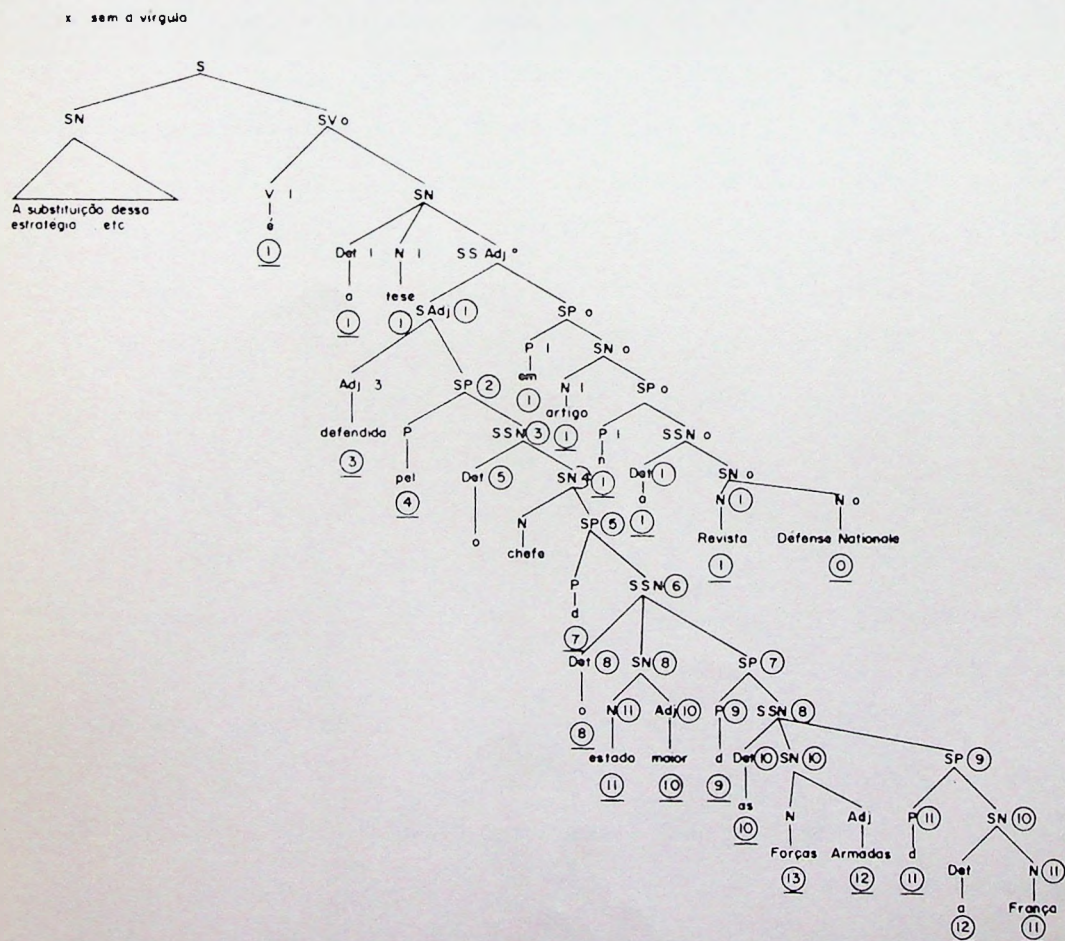
xii. ... é a tese por ele defendida, em artigo na revista Défense Nationale.

Como as seis árvores abaixo demonstram, a aproximação cada vez maior dos valores em própons das sentenças com a vírgula e

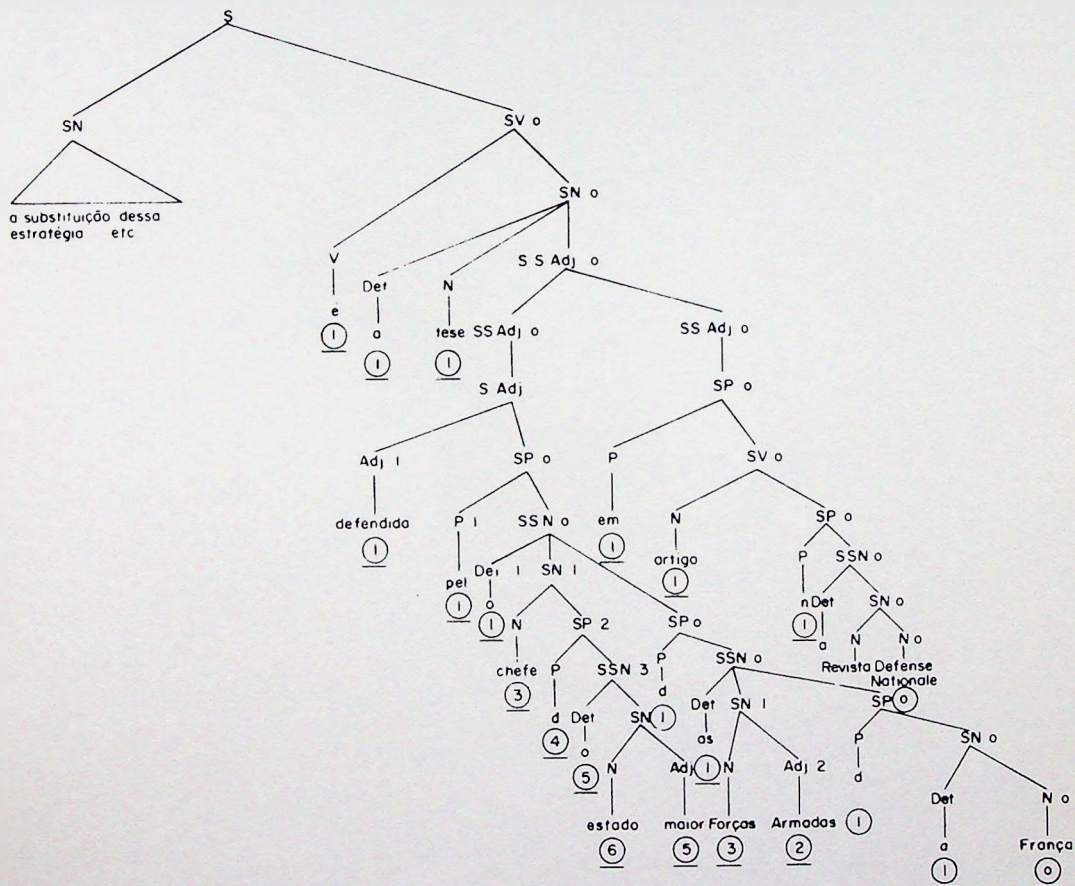
sem ela em cada par constitui uma notável correspondência entre as nossas percepções de desiderabilidade da vírgula e a quantificação em própons das respectivas sentenças:

x. sem a vírgula

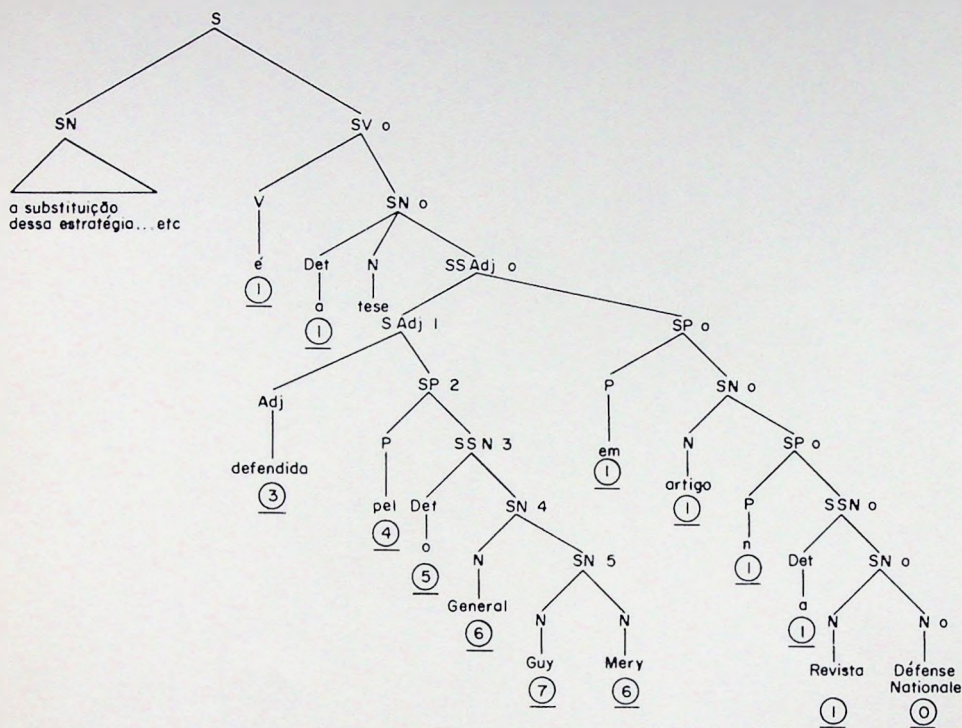
Peso total: 140 própons



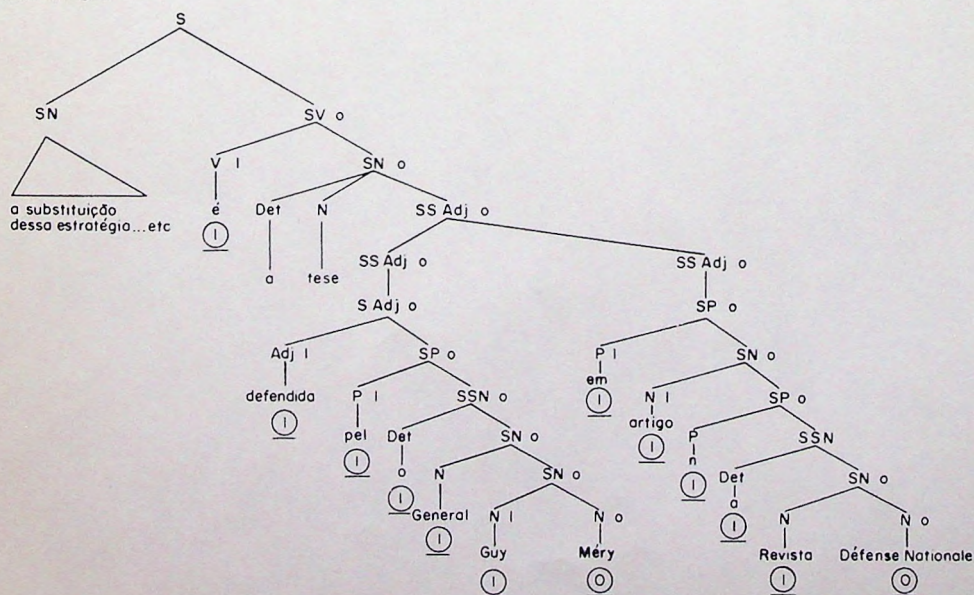
x com a vírgula



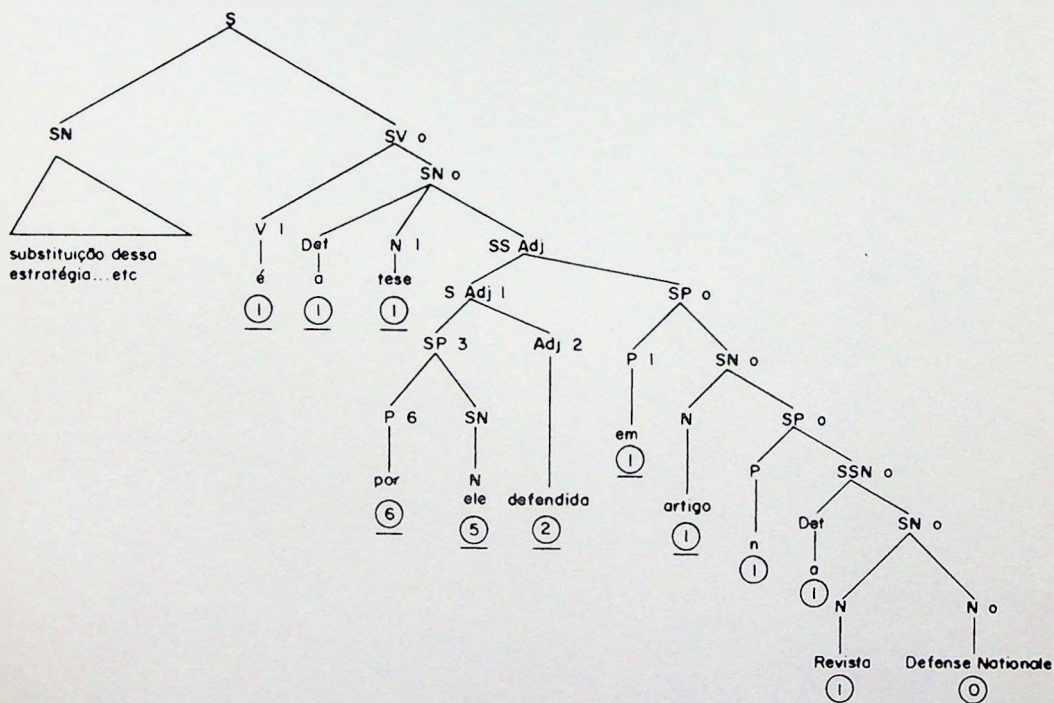
xi. sem a vírgula

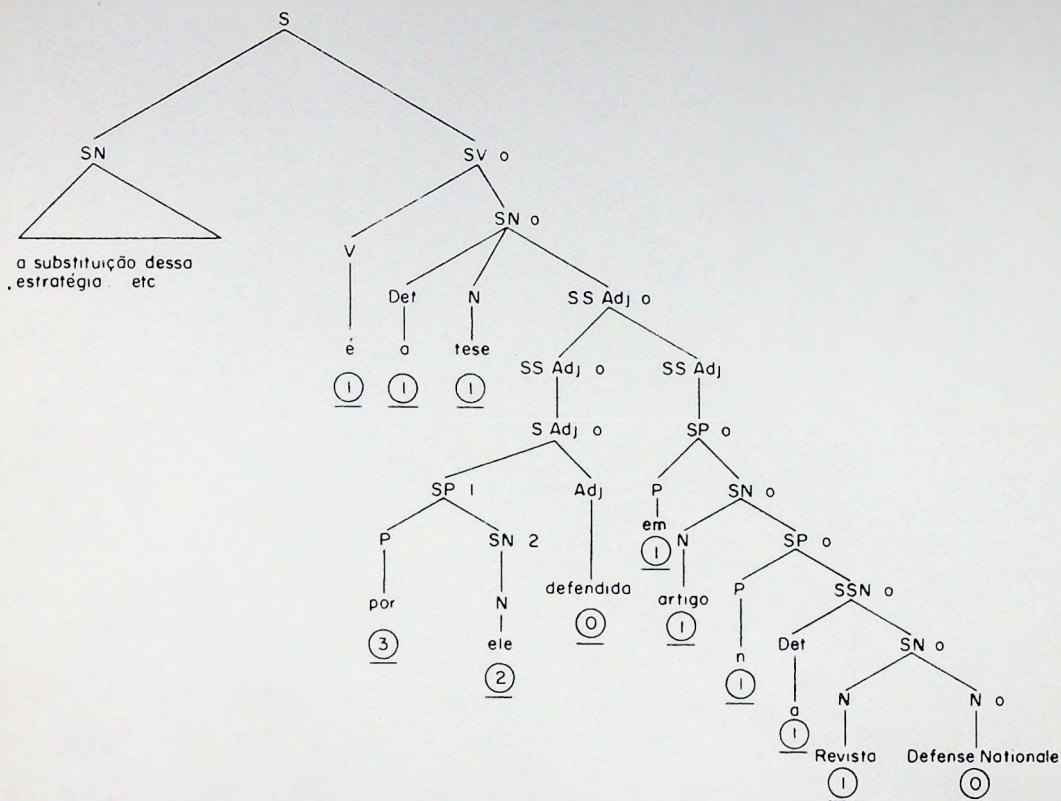


xi. com a vírgula



xii. sem a vírgula





Temos, portanto:

	<u>x</u>	<u>xi</u>	<u>xii</u>
sem vírgula	140	39	21
com vírgula	42	14	13
diferença	98	25	8
razão	3,3	2,7	1,6

Comparemos a fissão sintática representada pela vírgula a uma instrução ao nosso computador mental: processe o constituinte

entrante em fases subseqüentes do cálculo. É natural supormos que essa possibilidade de escalonamento da computação seja tanto mais aproveitada quanto mais difícil venha a ser a computação global. O que as nossas preferências intuitivas sobre as sentenças x a xii indicam, e que nossos números retratam muito bem, é justamente isso: quanto mais importante a facilitação do cálculo que a vírgula nos proporciona, mais a desejamos. Daí a ordem crescente da nossa preferência pela forma com a vírgula ir de xii a xi a x.

Esta explicação para a gradação de desiderabilidade das vírgulas encontra uma comprovação interessante se invertermos, em xii, xi e x, a posição do sintagma nominal do predicado, passando-o para uma posição à esquerda, já por si mais "pesada": x'. A tese defendida pelo Chefe do Estado Maior das Forças Armadas da França, em artigo na revista Défense Nationale, é a substituição dessa estratégia por uma cooperação militar mais estreita com a OTAN.

xi'. A tese defendida pelo general Guy Mery, em artigo na revista Défense Nationale, é a substituição dessa estratégia por uma cooperação militar mais estreita com a OTAN.

xii'. A tese por ele defendida, em artigo na revista Défense Nationale, é a substituição ... etc.

Comparando agora a desiderabilidade da vírgula assinalada pelo círculo em x e x', xi e xi', xii e xii', verificamos que ela é mais desejável nos segundos membros dos pares.

Dispensando-nos, por economia de espaço, de apresentar as

respectivas árvores, daremos a tabela de valores do peso sintático do constituinte relevante para o segundo trio de sentenças:

	<u>x'</u>	<u>xi'</u>	<u>xii'</u>
sem vírgula	269	118	76
com vírgula	159	80	60
diferença	110	38	16
razão	1,7	1,4	1,2

O fato de que a presença da vírgula em questão se faz sentir mais necessária no grupo x', xi' e xii' faz-nos hipotetizar, comparando as duas tabelas de pesos, que o fator preponderante na estratégia computacional dos falantes seja a economia puramente aritmética de própons, e não a sua razão. Ademais, comparando xii' com xi, aquele com uma economia de própons acarretada pela vírgula inferior à deste, porém ocasionando uma reação mais fortemente favorável ao uso dela, podemos hipotetizar que acima de algum limiar de complexidade (cujo valor deverá ser empiricamente estabelecido - e provavelmente varia segundo o falante e segundo o estilo), toda economia nas exigências do cálculo é buscada pelos falantes, ainda que aritmeticamente pequena.

Outra conclusão importante pode ser tirada dessa constatação de que o grau de utilidade das vírgulas difere em se tratando de galhos de direita ou de esquerda. Essa constatação nos permite eliminar, como alternativa de explicação para tais vírgulas, a que se limita a verificar a existência de uma razão direta entre a considerabilidade da vírgula e o número de palavras à esquerda dela. Se, com igual número de palavras à esquerda da vírgula, ela é mais necessária quando fissiona constituintes de esquerda do que de direita, é porque a direção da ramificação é relevante.

Sexto exemplo: Em "Universais lingüísticos: uma reinterpretação" (por M. Lemle, comunicação ao Sétimo Congresso Brasileiro de Língua e Literatura, Rio de Janeiro, julho 1975) são discutidos dois princípios propostos por John Robert Ross como restrições universalmente válidas para regras transformacionais: a restrição de sintagma nominal complexo e a restrição de sujeito oracional. Foi aventado, nessa comunicação, um princípio explicativo subjacente a ambos os universais de Ross. Testes de gradação de aceitabilidade de sentenças foram aplicados a falantes do português, nos quais se lhes pedia que dessem uma escala de preferências (em termos de maior ou menor "facilidade das frases") para orações interrogativas e orações relativas. Essas orações apareciam em cinco tipos de posições sintáticas em relação à sentença principal; subjetiva à esquerda, subjetiva à direita, complemento nominal à esquerda do verbo principal, complemento nominal à direita do verbo principal, objetiva direta. Verificou-se que: a) as orações subordinadas à direita da principal foram as mais bem aceitas; b) qualquer que fosse a posição da oração subordinada, foram mais bem aceitos os casos nos quais o sintagma nominal transformacionalmente movido (quer por interrogação, quer por relativização) provinha de um ramo de direita, dentro da sua sentença. Essas constatações conduziram à conclusão de que as restrições para transformações propostas por Ross são apenas casos particulares da limitação das capacidades de estocagem de estrutura sintática na memória imediata.

Sétimo exemplo: Num estudo apresentado sob a forma de comunicação ao Primeiro Congresso Nacional de Lingüística e Literatura da Sociedade Universitária Augusto Motta (Rio de Janeiro, janeiro

1976, sob o título "Hierarquias de complexidade e regras gramaticais" (por M. Lemle e A. J. Naro), foram testadas as interpretações preferencialmente dadas por falantes nativos a sentenças contendo topicalizações e relativizações de interpretação ambígua, onde a ambigüidade se deve à dupla regência (intransitivos ou transitivos diretos) de certos verbos do português. Assim, nas frases 1, 2 e 3 dadas abaixo,

- 1) Aqueles ladrões, eles afirmam que mataram.
- 2) Quero visitar as obras que os soldados disseram que era difícil acabarem antes de junho.
- 3) Aquele veneno mata os bichos que os zoólogos dizem que não conseguem ver.

constatamos que uma percentagem significativa de informantes interpretou a frase 1 com aqueles ladrões como objeto de matar, a frase 2 com as obras como sujeito de acabarem, e a frase 3 com os bichos como objeto de ver.

A interpretação dada a esses fatos foi a de que o fator determinante do viés foi a maior facilidade (índice de posição estrutural mais baixo) das posições com as quais os elos de co-referência foram estabelecidos.

Oitavo exemplo: Ainda na supra-citada comunicação, investigamos o uso da regra de supressão do sujeito oracional (a regra responsável pelo que tradicionalmente se denomina sujeito oculto). O problema colocado foi o da determinação dos contextos sintáticos favoráveis e desfavoráveis à supressão do pronome de 3a. pessoa, hipotetizando-se haver alguma relação entre a preferência pela supressão do pronome e as posições estruturais do pronome sujeito e

do sintagma nominal que é o seu antecedente.

As frases utilizadas no teste apresentavam várias combinações possíveis de índices de posição estrutural do par Sintagma nominal antecedente - pronome pessoal co-referente. Verificou-se que os contextos mais favoráveis à supressão do pronome são aqueles em que o sintagma nominal antecedente do pronome e o pronome candidato à supressão têm valores semelhantes nos seus índices posicionais, havendo uma boa relação linear entre a tendência a suprimir o pronome e a semelhança das posições estruturais do sintagma nominal antecedente e o pronome. Dois exemplos extremos seriam Kissinger fica em contato permanente com a Casa Branca quando ele viaja (supressão máxima), versus Se tivessem sido conhecidos os poemas que Santos Dumont escreveu, ele teria se tornado um poeta famoso (supressão mínima).

Ademais, a tendência a suprimir o pronome diminui na medida em que o par antecedente - pronome está em posições mais difíceis (i.e., índices de posição estrutural (ISP) mais altos).

A relevância das medidas estruturais aqui propostas para a regra de interpretação de topicalização e relativização e a regra de formação de sujeito oculto é evidente.

3.1.3 - Outras medidas de complexidade

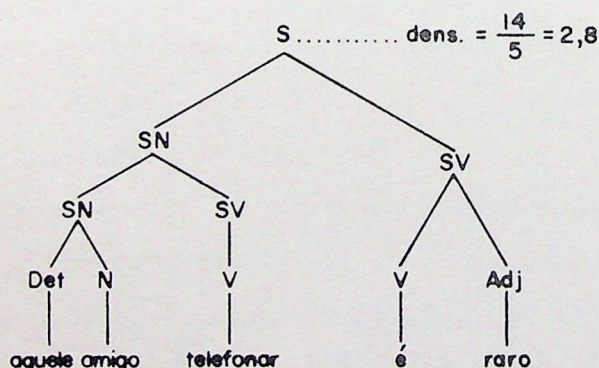
Do método de mensuração sintática acima descrito, derivamos, além do peso sintático por palavra, as seguintes outras medidas:

Índice de posição estrutural (ISP) de um nóculo é o cálculo do peso em prótons desde a raiz da árvore até esse nóculo. Nos exemplos

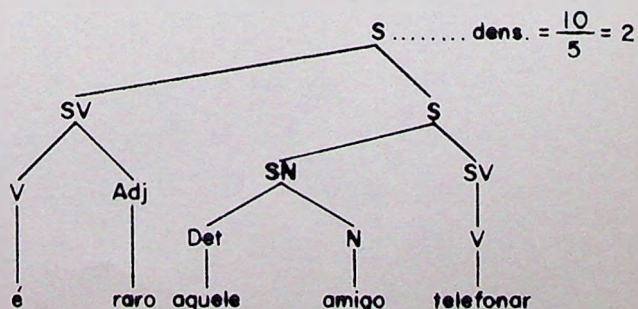
i e ii das pag.57/58, o ISP de cada nódulo é o seu p. A medida de ISP de cada nódulo reflete o seu grau de encaixe e retrata a carga computacional que incide nele em decorrência da posição que ocupa na árvore. Essa maneira de caracterizar a posição estrutural de um constituinte nos abre uma outra hipótese interessante, a ser testada na pesquisa: a de que quanto mais alto o ISP de um sintagma (i. e. quanto mais abaixo e à esquerda na árvore), tanto menos complexo ele tenderá a ser.

Densidade de um sintagma - é o seu peso em prótons dividido pelo seu número de palavras. Por exemplo, para as sentenças i e ii das pag.57 e 68 a densidade da S principal é de $14/5 = 2,8$ para a primeira e $10/5 = 2$ para a segunda.

i.



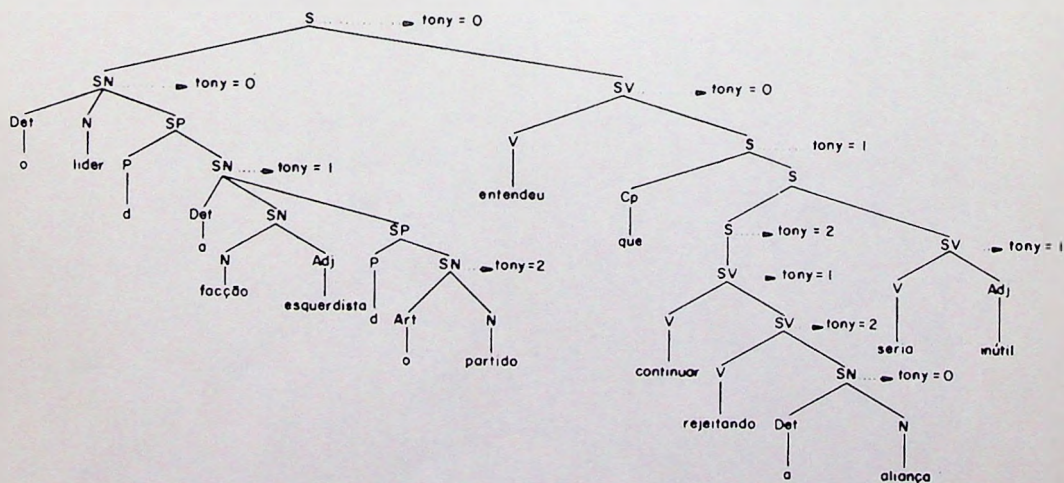
ii.



Como se vê, a sentença i, com a oração subjetiva à esquerda do predicado principal tem densidade maior do que a ii, com a oração subjetiva à direita. Assim, podemos com essa medida distinguir os sintagmas através do tipo de configuração pelo qual atingem o seu peso: tanto mais alta será a densidade de um sintagma quanto mais ele se ramificar para a esquerda.

Temos ainda, como caracterizações da localização dos nós na árvore:

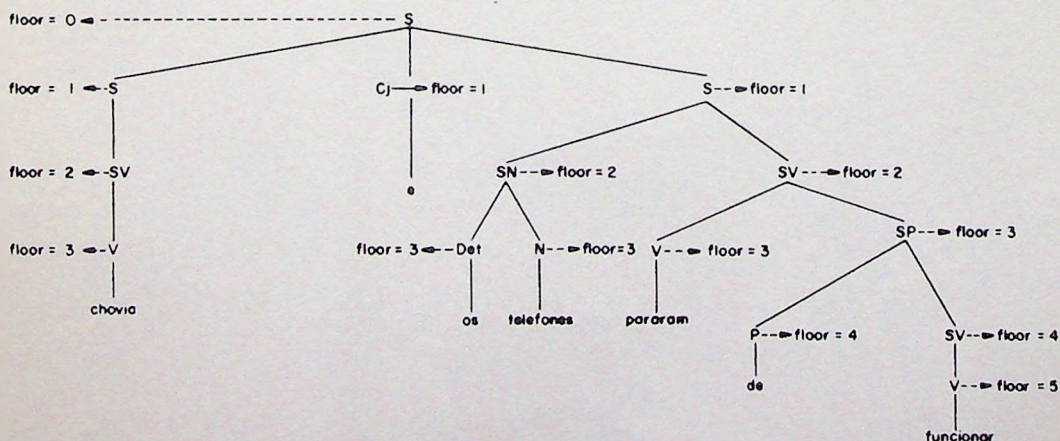
Tony de um nóculo - é o número de nóculos identicamente rotulados que o dominam, mediata ou imediatamente. É uma medida de auto-encaixe,



e não de encaixe. O uso que fizemos dessa medida foi meramente instrumental. No estudo da relação entre ISP e complexidade, serviu-nos para peneirar apenas os sintagmas não auto-encaixados, de modo a evitar o risco de que a relação inversa entre ISP e complexidade pudesse vir a ser atribuída meramente ao fato de ser uma grande parte dos sintagmas de ISP alta constituída de subsintagmas dos de ISP mais baixa. A seleção dos sintagmas de Tony = 0 permitiu-nos evitar esse problema.

Floor de um nóculo - é o número de nóculos que o dominam, a partir da raiz da árvore.

Sintagmas de um só constituinte, retratados por superposições de nóculos não ramificados, também são contados.



Floor difere de ISP pelo fato de que em seu cálculo a direção da ramificação é indiferente, ao passo que em ISP a ramificação para a esquerda conta mais do que para a direita.

3.2 - Caracterização da amostra

O critério de seleção da amostra lingüística para o estudo da complexidade sintática decorreu da pergunta que nos fizemos, e que foi: atingirão os mobralenses, em suas sentenças mais complexas, o nível de complexidade média dos jornais? Tal pergunta, note-se, decorreu por sua vez do pressuposto de que haja alguma relação entre a capacidade de desempenho ativo do falante e a sua capacidade de desempenho como leitor, no sentido de que uma dada variedade de linguagem escrita estará tanto mais ao alcance das capacidades de leitura de um indivíduo quanto mais ela estiver aquém dos seus próprios limites de uso como falante. Portanto, o que nos interessava era comparar a complexidade máxima das sentenças do corpus de linguagem falada dos mobralenses com a complexidade média das variedades de linguagem escrita. Isso posto, devíamos utilizar um critério aleatório para o estabelecimento da amostra de sentenças da linguagem escrita, mas escolher as sentenças mais complexas de cada informante para a amostra de linguagem falada.

O método de seleção aleatória utilizado foi o seguinte: tomamos a tabela de números aleatórios existente em Blalock, H.M., Social Statistics (McGraw-Hill, 1960/72), pg. 554 a 557, e estabelecemos as seguintes regras de correspondência:

a. Amostra de sentenças de jornais: 1º dígito, qual o jornal;

2º e 3º dígito, qual a página do jornal; 4º dígito, qual o artigo na página, 5º dígito, qual a coluna do artigo; 6º e 7º dígito, qual a sentença na coluna.

b. Amostra de sentenças de fotonovelas: 1º dígito, qual a revista; 2º e 3º dígito, qual a página na revista; 4º dígito, qual o quadrinho (ou coluna) na página; 5º dígito, qual a sentença no quadrinho ou coluna.

c. Amostra de sentenças de histórias em quadrinhos: 1º dígito, qual a revista; 2º e 3º dígito, qual a página; 4º e 5º dígito, qual o quadrinho na página; 6º dígito, qual a sentença no quadrinho.

d. Amostra de literatura nacional. Foram tomadas, em cada um dos dez livros selecionados, 40 sentenças, e a correspondência com a lista de números aleatórios foi a seguinte: 1º, 2º e 3º dígitos, qual a página; 4º e 5º dígitos, qual a sentença na página.

Ocorrendo um número mais alto do que o máximo de casos existentes, calculava-se o resto sob divisão do número em questão pelo número de casos existentes.

Este método de escolha não é, no entanto, estritamente aleatório. Uma sentença comprida ocupa mais espaço no papel do que uma curta. Portanto, dentro do espaço físico representado pela unidade "página", as sentenças curtas receberão mais números do que as longas, por unidade de espaço ocupada. Isto faz com que o método de seleção favoreça as sentenças menos compridas, quando estas alternam com as mais compridas mais ou menos sistematicamente. Notamos esse efeito especialmente na amostra da literatura nacional, onde, por razões estilísticas, tal alternância de sentenças compridas e curtas mais se dá.

Daremos, a seguir, a indicação das publicações de onde foram retiradas as sentenças da amostra.

Amostra de literatura nacional: 400 sentenças

(por sugestão de Affonso Romano de Sant'Anna)

Feliz Ano Novo - Rubem Fonseca (Artenova)

Deixa o Alfredo Falar - Fernando Sabino (Record)

Cemitério de Elefantes - Dalton Trevisan (Civ. Brasil.)

Contos Novos - Mário de Andrade (Martins)

A Morte e a Morte de Quincas Berro d'Água - Jorge Amado
(Record)

O Pirotécnico Zacarias - Murilo Ribião (Ed. Atica)

Leão de Chácara - João Antônio (Civ. Bras.)

Uma Vida em Segredo - Autran Dourado (Ed de Ouro)

Olhai os Lírios do Campo - Érico Veríssimo (Globo)

Contos Consagrados - Machado de Assis (Edições de Ouro (Editora Technoprint)

Amostra de Histórias em Quadrinhos (60 sentenças)

Recruta Zero - nº 161 (Rio Gráfica e Editora)

Bolinha - Ano II nº 7 (Abril)

Pinóquio - Nº 2 (Vecchi)

A Turma do Pererê - Ano I nº 9 (Abril)

Thor - Ano II - nº 13 - (Bloch Infanto-Juvenil)

Amostra de Fotonovelas (140 sentenças)

Superálbum Carícia - Ano I nº 13A (Abril)

Capricho - Ano XXIII nº 400 (Abril)

Sétimo Céu - nº 42 - (Bloch)

Romântica - nº 204 - Ano XV (Vecchi)

Grande Hotel - nº 1496 - Ano XXIX (Vecchi)

Amostra de jornais: jornais do dia 25 de março de 1976 (400 sentenças)

A - O Globo, Jornal do Brasil, Diário de Notícias

B - O Dia, A Notícia, Jornal dos Sports

3.3 - Método de arborização

3.3.1 - O Sintagma Nominal

Admitimos as seguintes regras:

$$\begin{aligned} \text{SSN} &\rightarrow (\text{Quant}) (\text{Art}) (\text{Pos}) (\text{Id}) (\text{Car}) (\text{Ord}) (\text{SAdj}) \text{SN} \left(\begin{array}{c} \text{SAdj} \\ \text{SPrep} \\ \text{S} \end{array} \right) \\ \text{SN} &\rightarrow \text{N} \left(\begin{array}{c} \text{SAdj} \\ \text{SPrep} \\ \text{S} \end{array} \right) \end{aligned}$$

onde:

Quant : todo, - a, - os, - as, ambos, - as.

Art : o, este, esse, aquele, um, algum, certo, cada, qualquer

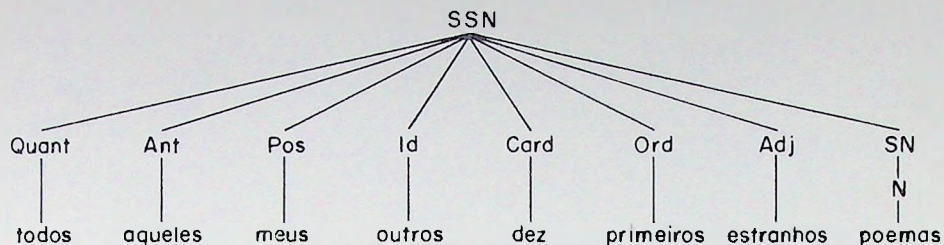
Pos : meu, teu, seu, seu, nosso, vosso

Id : outro, mesmo

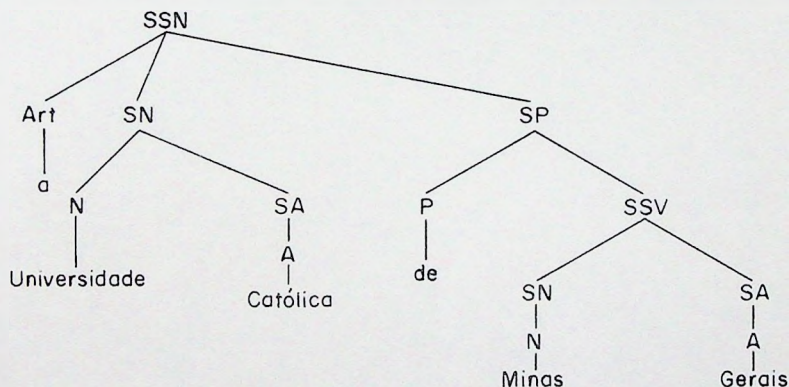
Car : um, dois, três, ... vários, diversos, numerosos

Ord : primeiro, segundo, terceiro, ... último

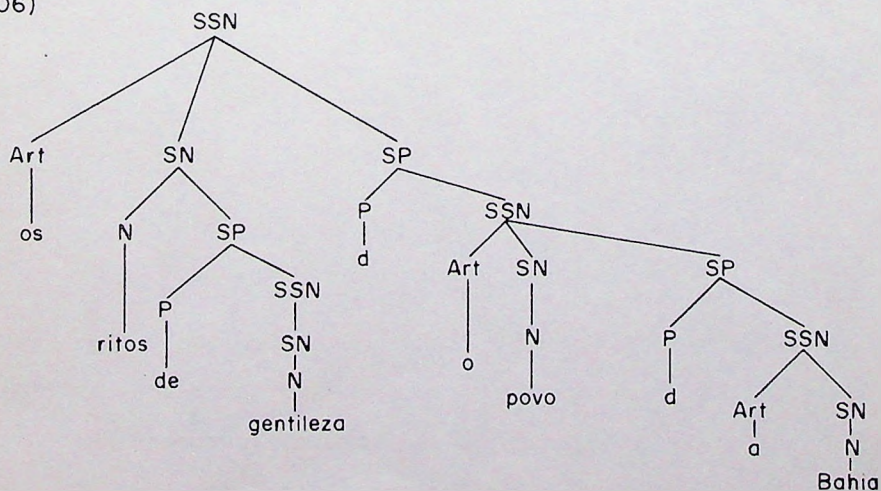
Um exemplo de Sintagma Nominal com as sete posições pronominais preenchidas seria:



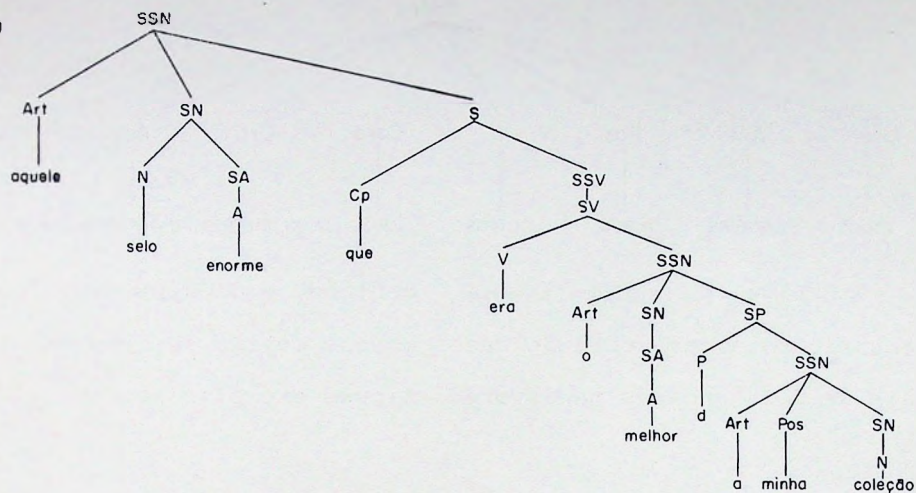
O nível de supersintagma, contendo o de sintagma, foi necessário para dar conta de casos em que deviam ser representados mais de dois andares na árvore. Alguns exemplos seguem:



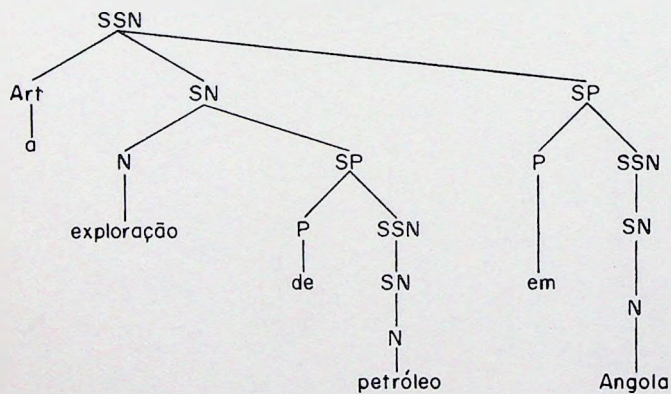
(706)



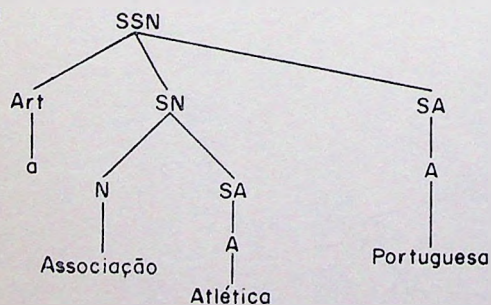
(920)



(176)



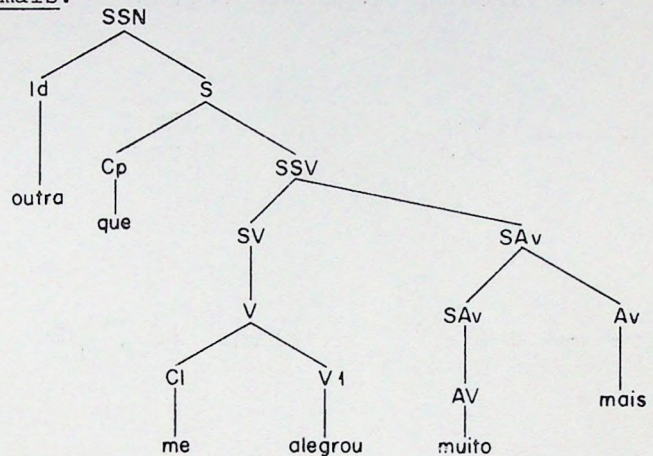
(92)



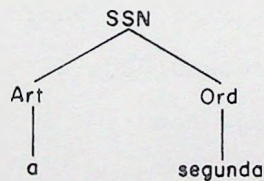
Há casos de Sintagma Nominal sem o nome que seria o seu termo central. Semanticamente, o contexto o fornece. Sintaticamente, representamo-lo tal como se apresenta, carente de nome.

(920) o melhor da minha coleção (vide acima)

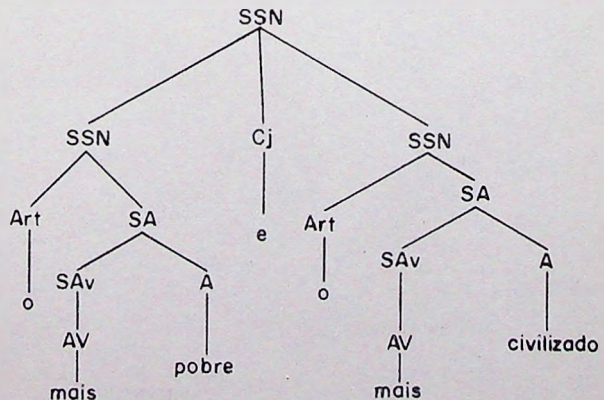
(1222) Você fez uma coisa importante pra mim que eu mais gosto, e fora disso você fez outra que me alegrou muito mais.



(891) cumprida a primeira parte da gafe, saio impávido para a segunda.



(706) cumpriam-se os ritos de gentileza do povo de Bahia, o mais pobre e o mais civilizado.



NOTA: Os números são os que identificam as árvores na amostra.

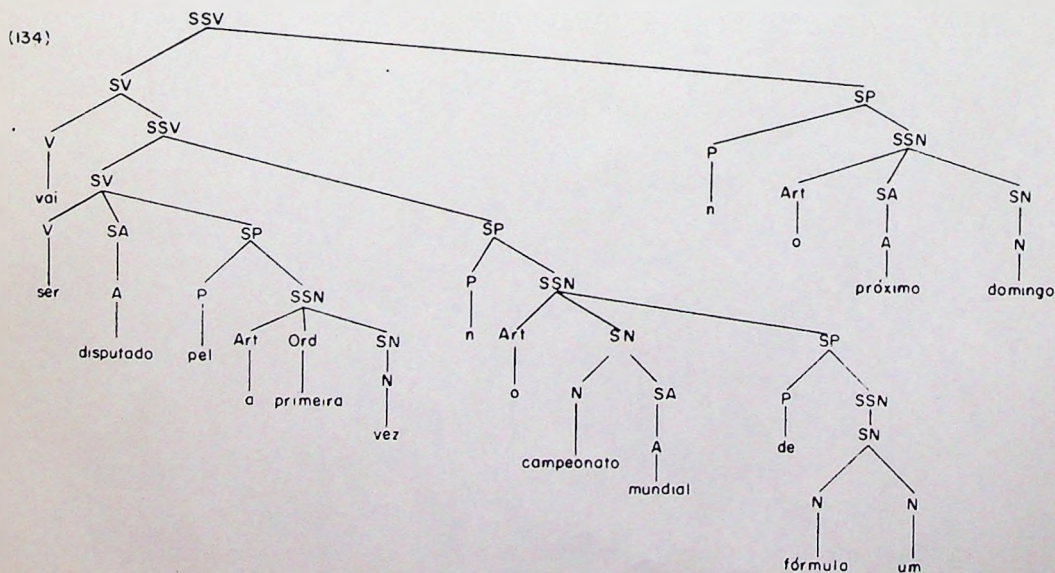
3.3.2 - O Sintagma Verbal

Admitimos as seguintes regras:

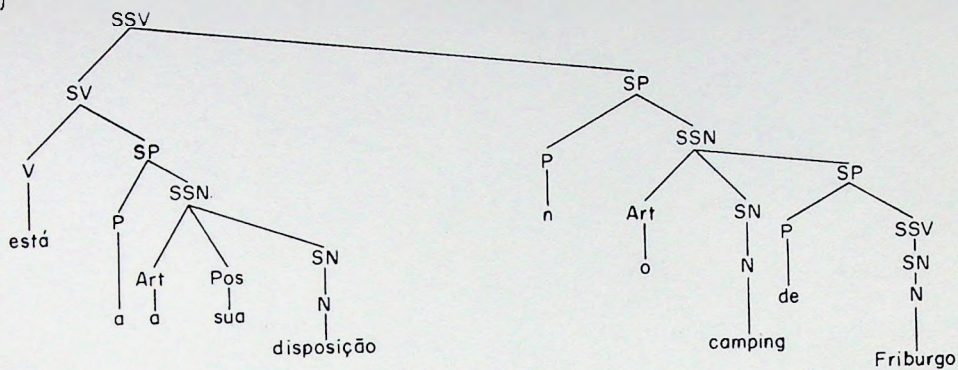
SSV $\longrightarrow$ ({ Neg }) SV ({ SAV })
SAV
SPrep

SSN
SV $\longrightarrow$ V ({ SSV }) SPrep ({ SAV })
SAdj
S

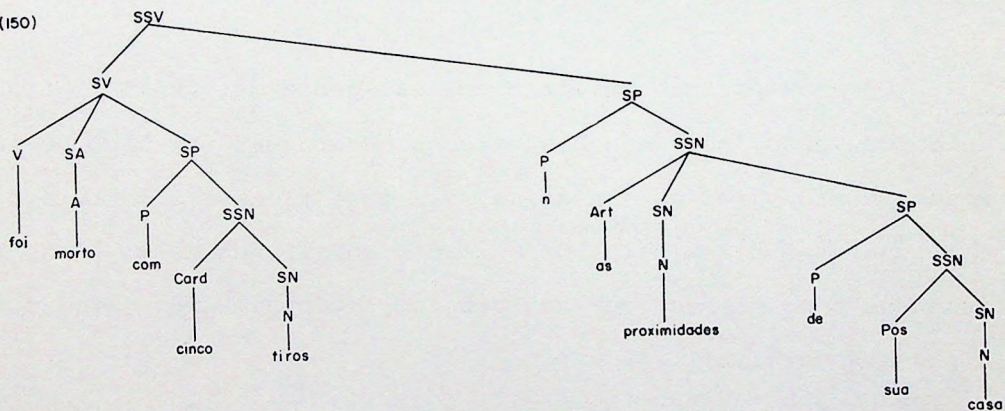
Como no caso do Sintagma Nominal, o que motivou o nível de supersintagma foi a necessidade de subagrupamentos hierárquicos em mais de dois andares, como:



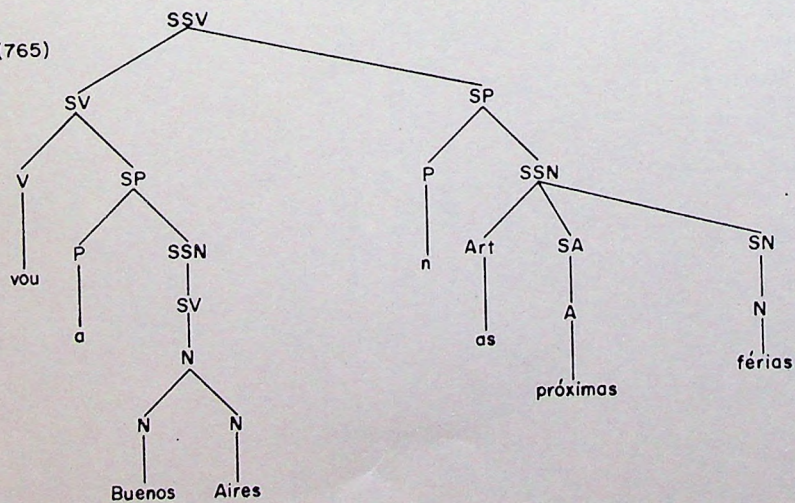
(146)

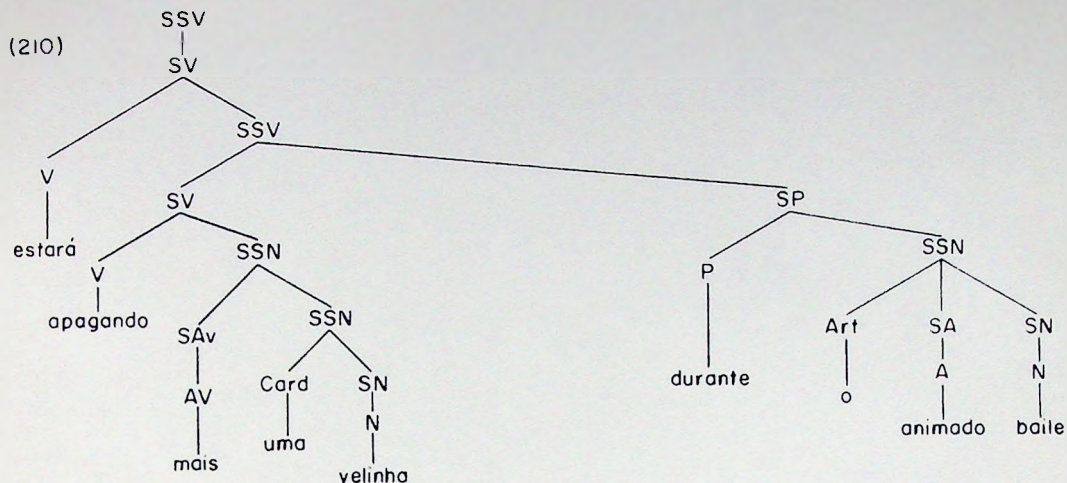


(150)

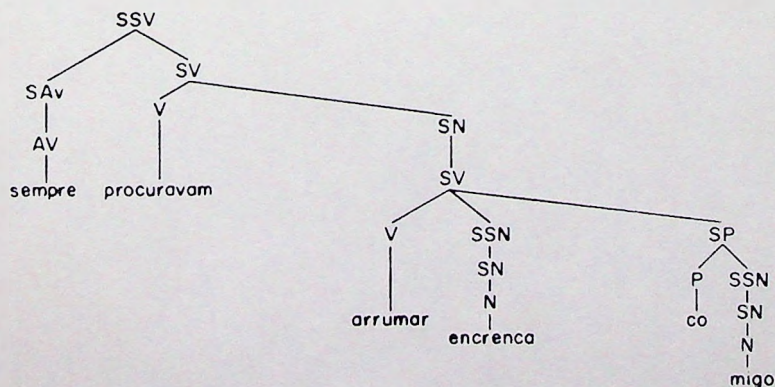


(765)

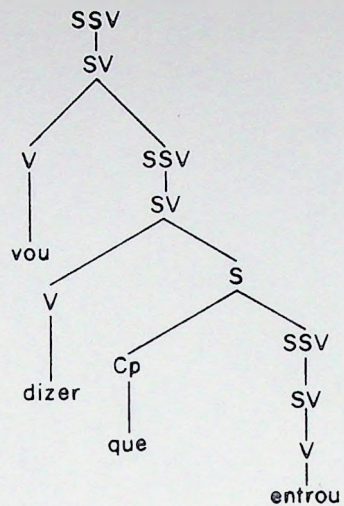




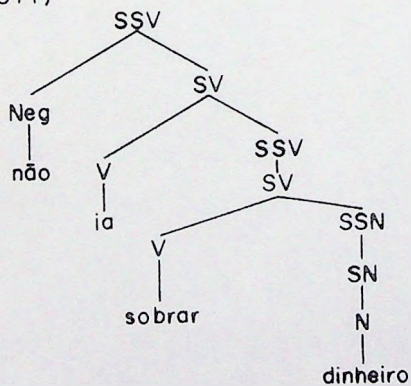
Optamos por não utilizar uma categoria de Auxiliar, diferente da de Verbo, pois não nos pareceu haver qualquer tipo de informação, morfológica ou sintática, que justificasse a postulação de mais uma classe lexical. Os chamados auxiliares estão sendo representados como verbos, por sua vez complementados por novos supersintagmas verbais:



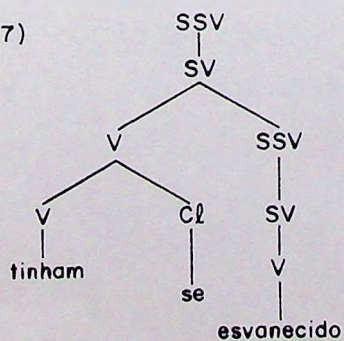
(1162)

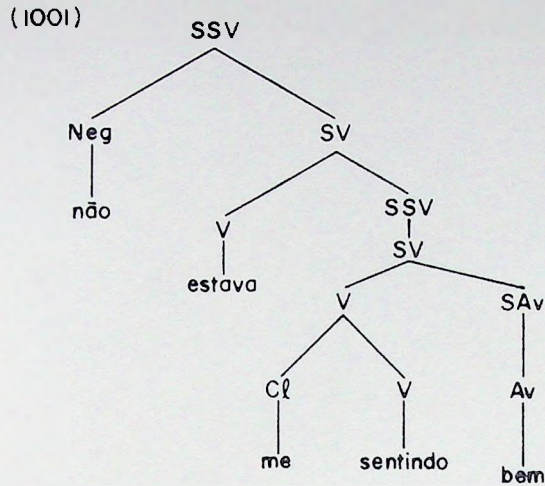


(1071)



(817)





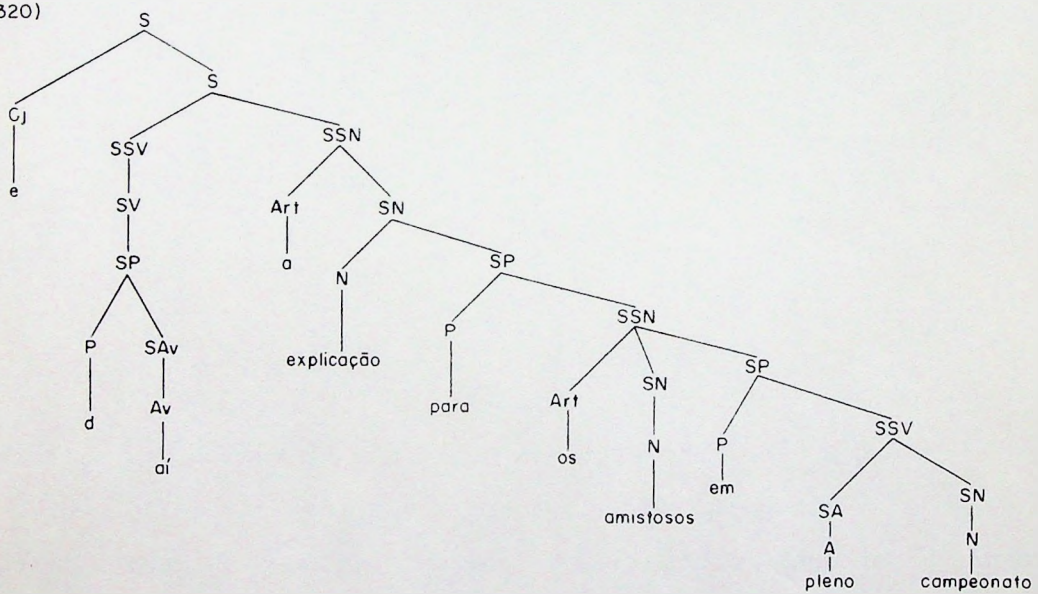
Quanto às exigências de desinências particulares impostas por cada um dos ditos auxiliares aos verbos a eles ligados (estar + V-ndo, ir + V-r, ter + V-do), julgamos ser esse um problema de natureza semelhante à da concordância verbal ou nominal ou à da seleção de casos, ou seja, interdependências sintático morfológicas, que não justificam por si só a postulação de categorias lexicais distintas.

Para os fins da presente pesquisa, contudo, o problema de rotulação dos auxiliares não é relevante, pois o método utilizado de avaliação de complexidade é indiferente às distinções de classe lexical.

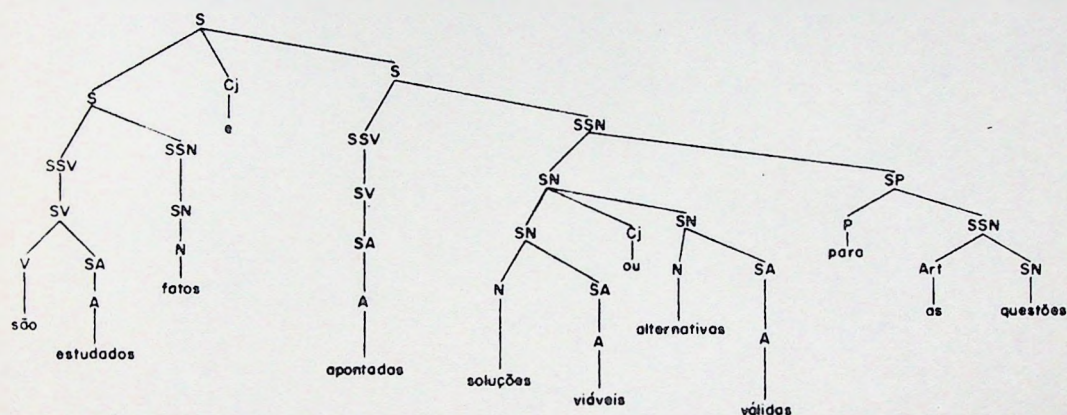
Como no Sintagma Nominal, onde há sintagmas sem nome, há também Sintagmas Verbais sem verbo.

Exemplos:

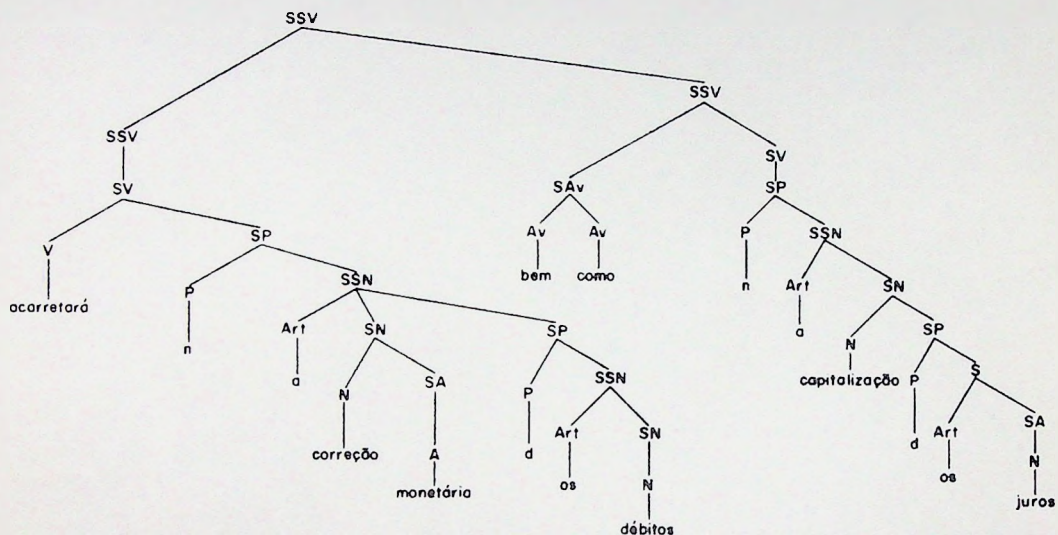
(320)



(323) Os problemas brasileiros são analisados no Conselho Técnico da CNC, quando são estudados fatos e apontados soluções viáveis ou alternativas válidas para as questões que interessem governos, empresários e povo.



(362) O não cumprimento da obrigação de acordo com o estabelecido no artigo 19 do Decreto-Lei 1432, de 5-12-1975, acarretará na correção monetária dos débitos, bem como na capitalização dos juros.



3.3.3 - O Sintagma Adjetivo

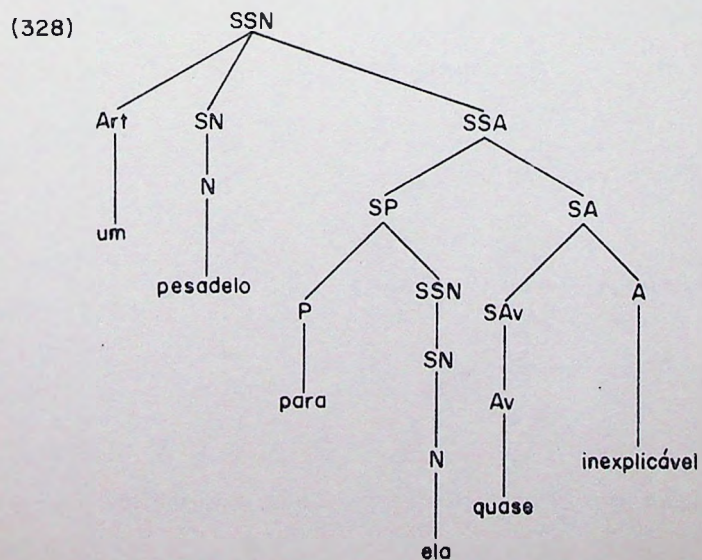
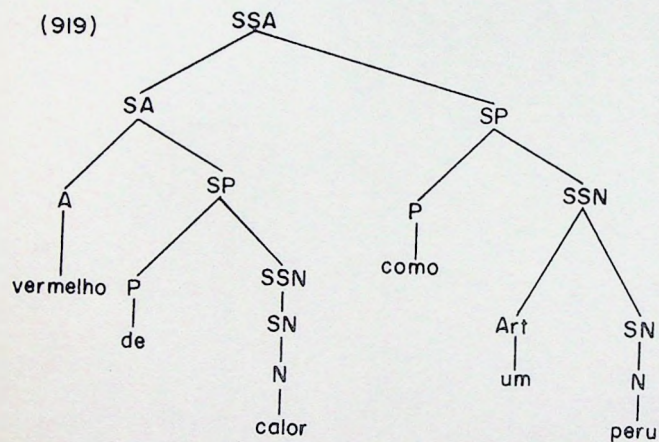
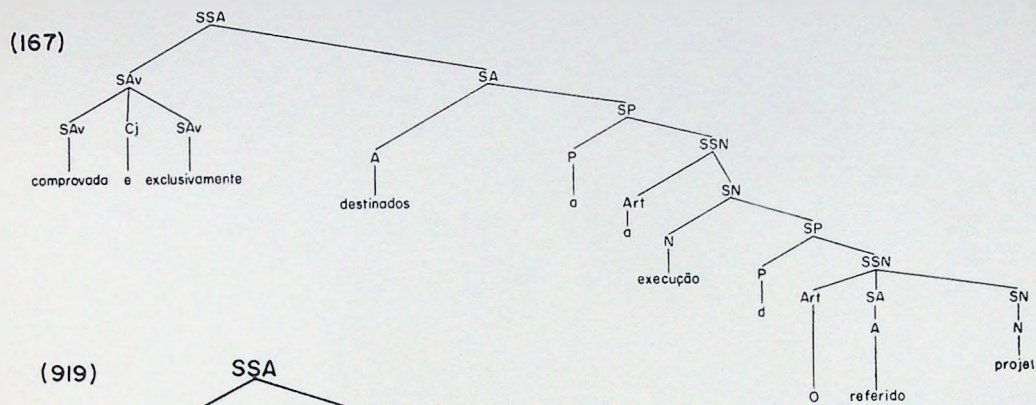
Admitimos as regras:

SSAdj $\longrightarrow$ (SAv) SAdj

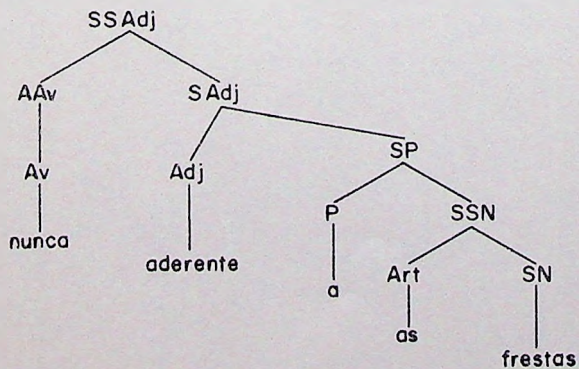
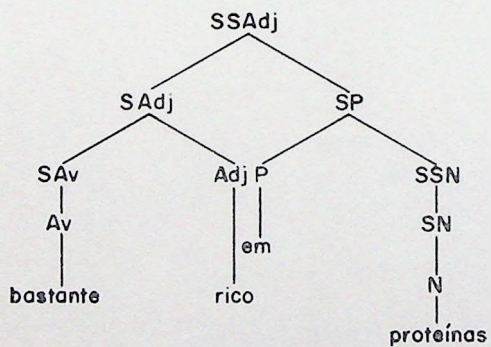
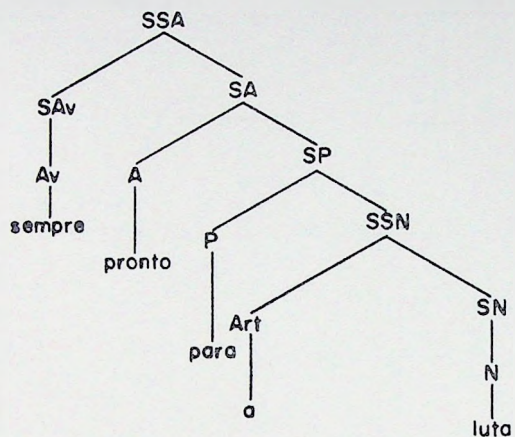
SAdj $\longrightarrow$ Adj (SPrep)

Os tipos de árvores de mais de dois andares que justificam a postulação do nível de supersintagma podem ser exemplificadas pe-

las seguintes:



Outros exemplos seriam:

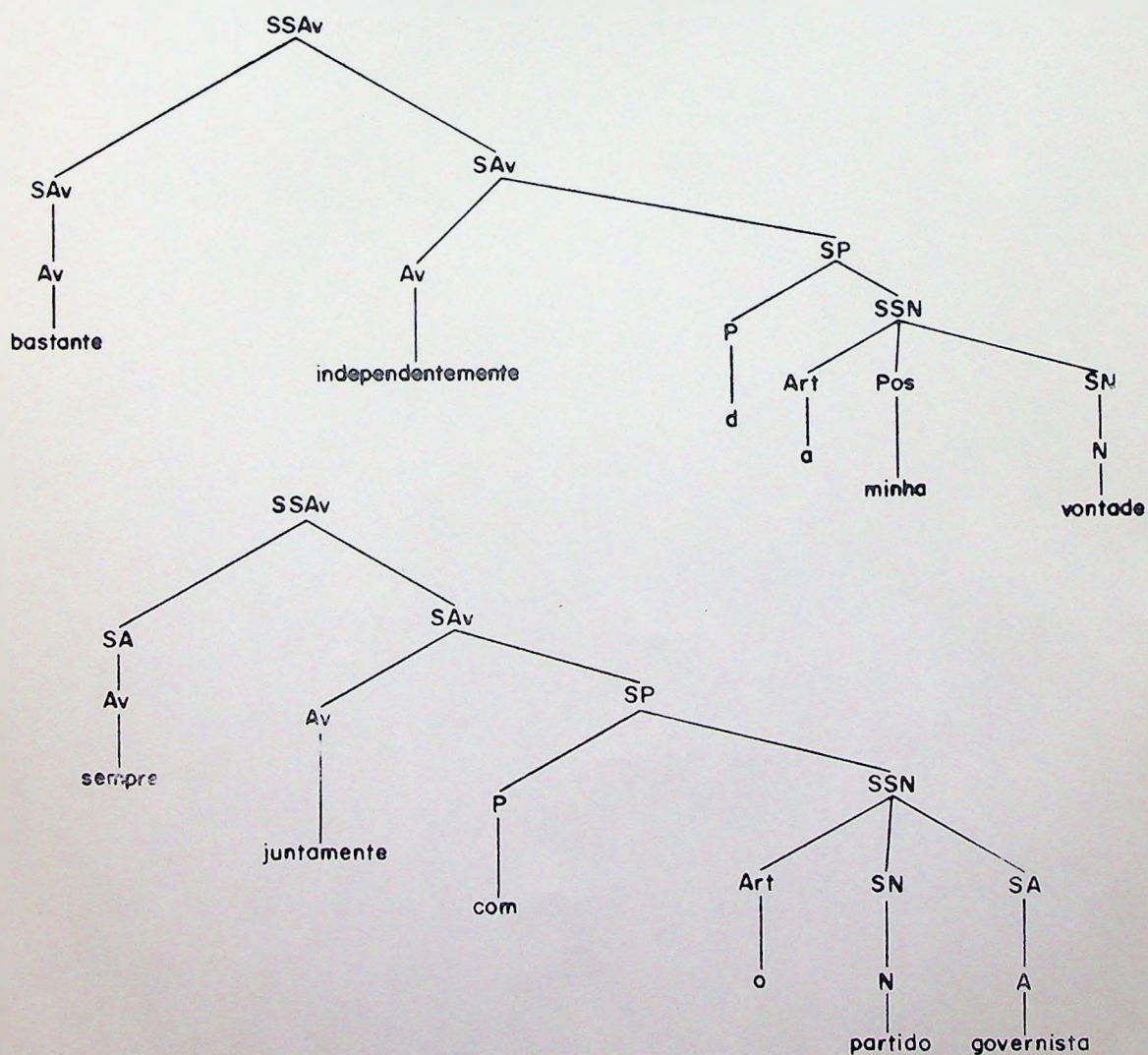


3.3.4 - O Sintagma Adverbial

SSAv $\longrightarrow$ (SAv) SAv

SAv $\longrightarrow$ Av (SPrep)

O aproveitamento integral das regras está demonstrado nos seguintes exemplos:



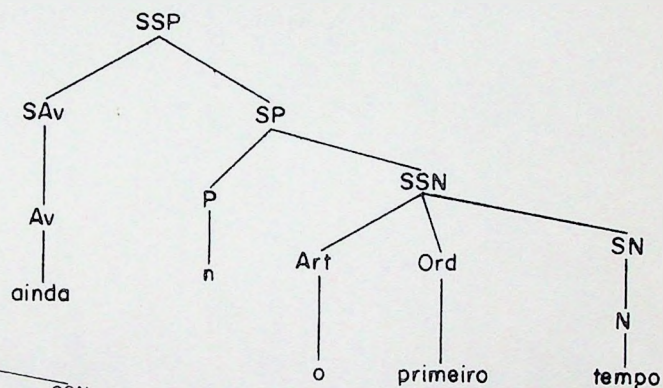
3.3.5 - O Sintagma Preposicional

$SSP \rightarrow (SAv) SP$

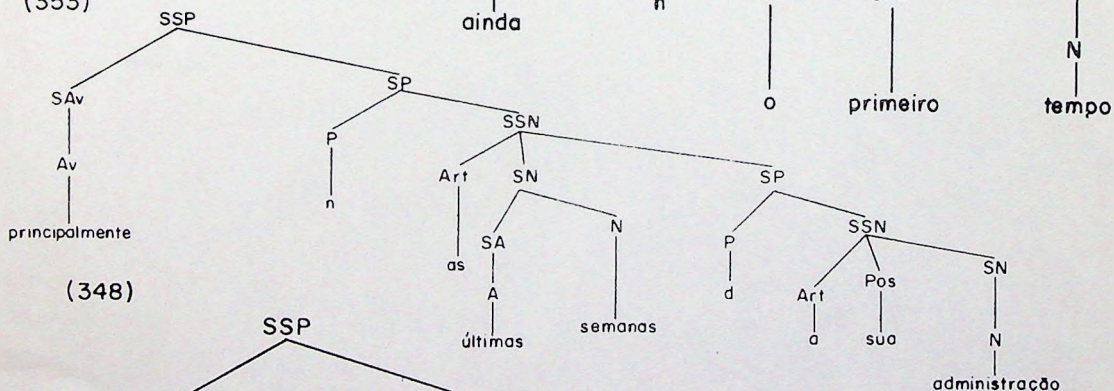
$SP \rightarrow P \begin{cases} SSN \\ SSV \\ SP \\ S \end{cases}$

Alguns exemplos da utilização plena das regras são:

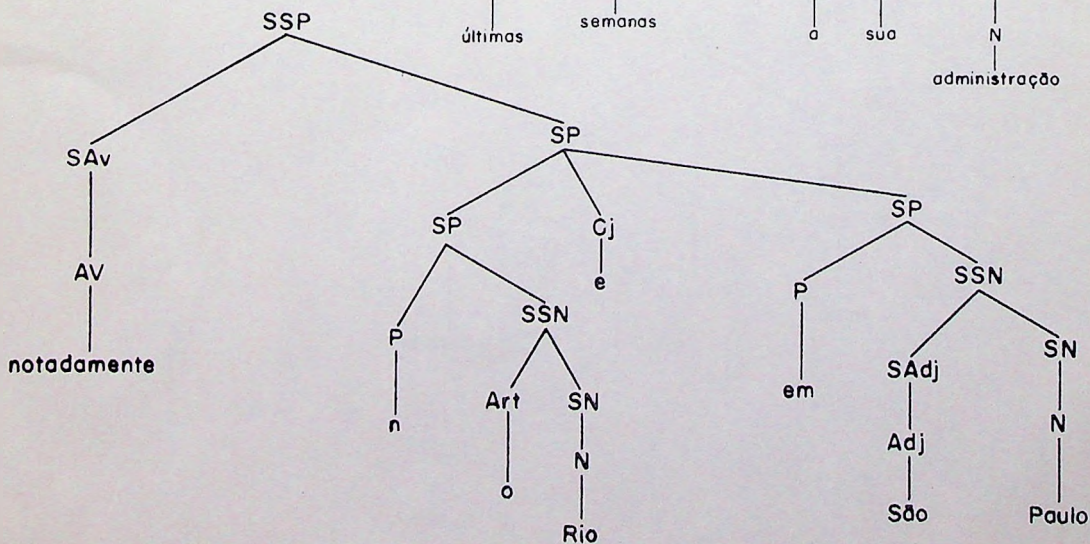
(304)



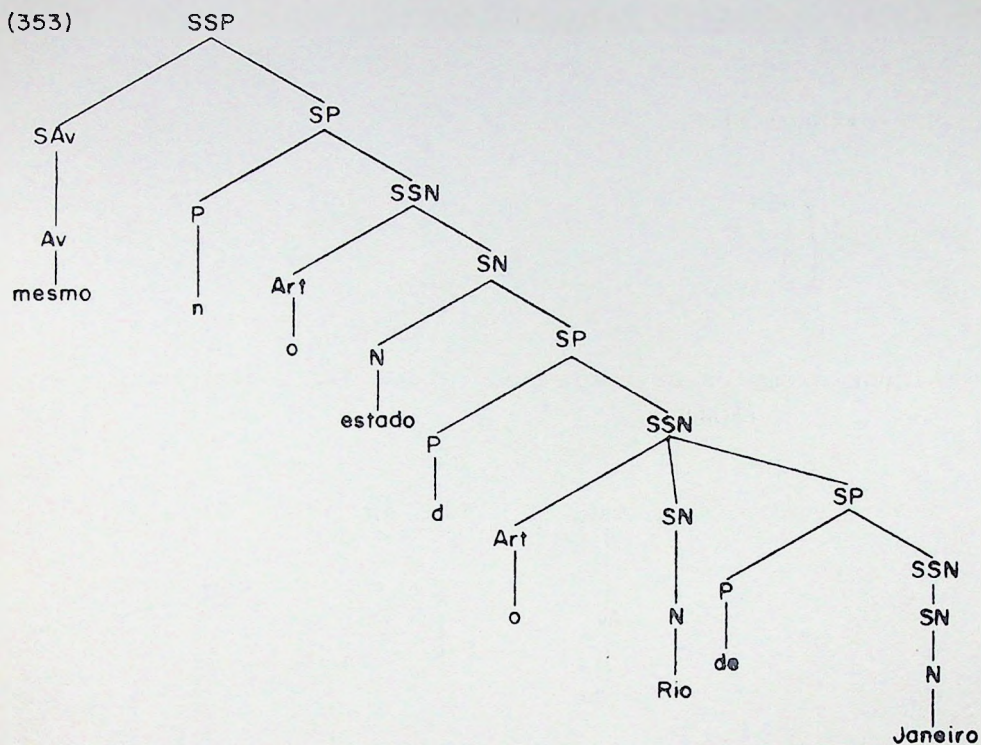
(353)



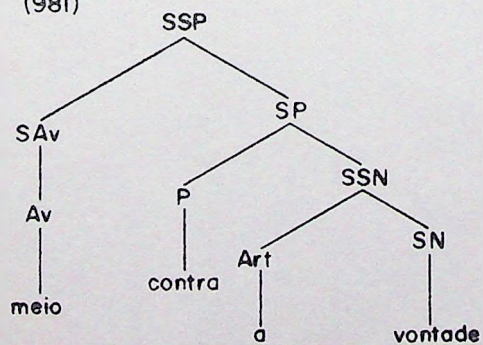
(348)



(353)



(981)



Note-se que um Sintagma Adverbial e um Supersintagma preposicional consistem da mesma seqüência de elementos: um advérbio e um Sintagma Preposicional. No caso do Sintagma Adverbial, é o advérbio que é o termo essencial e o Sintagma Preposicional complementa o advérbio, enquanto que no Supersintagma preposicional, o centro da construção é o Sintagma Preposicional, sendo o advérbio um mero modificador deste.

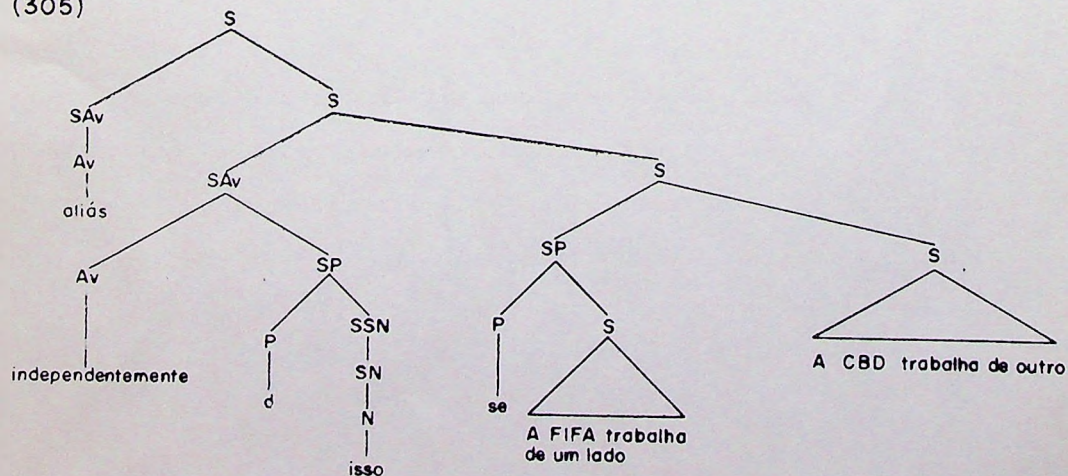
Entre as preposições, consideramos alguns itens que aparecem em gramáticas tradicionais classificados como conjunções subordinativas: se, quando, embora.

3.3.6 - A Sentença

Todos os modificadores sentenciais antepostos à sentença nuclear foram representados como ramificações à esquerda, podendo haver casos com superposições numerosas desses modificadores.

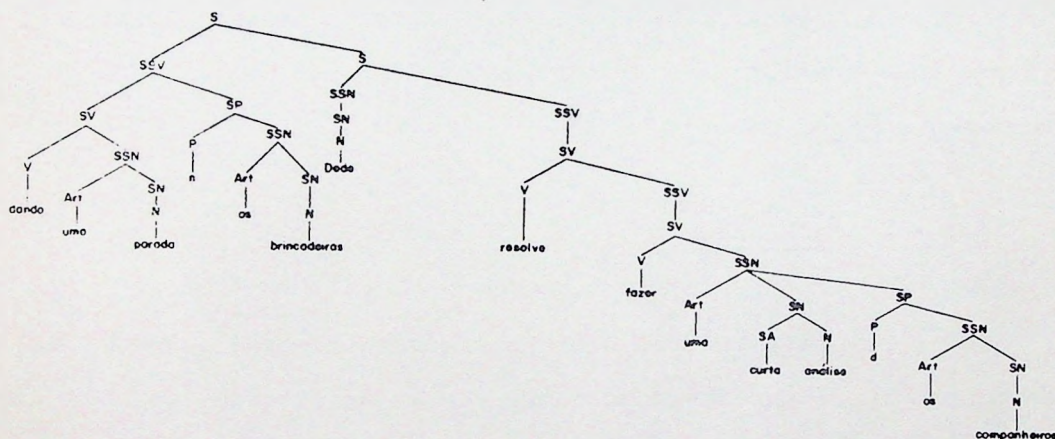
Exemplos:

(305)

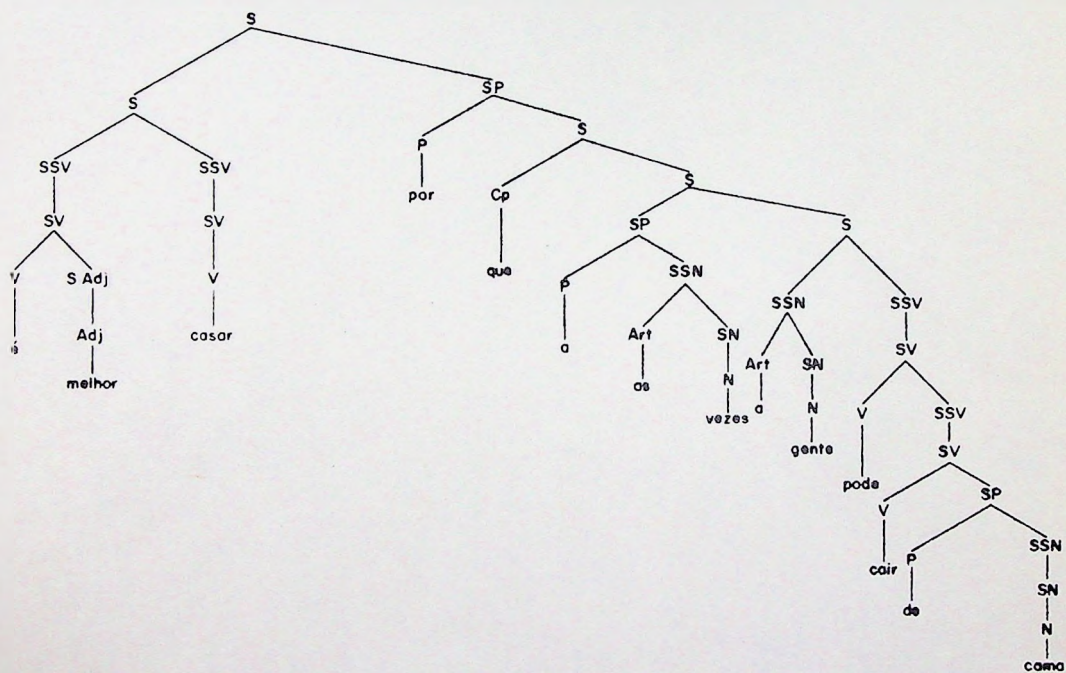


As chamadas orações reduzidas de gerúndio foram representadas como sintagmas verbais, ficando sua função de modificador sentencial expressa pela configuração da árvore.

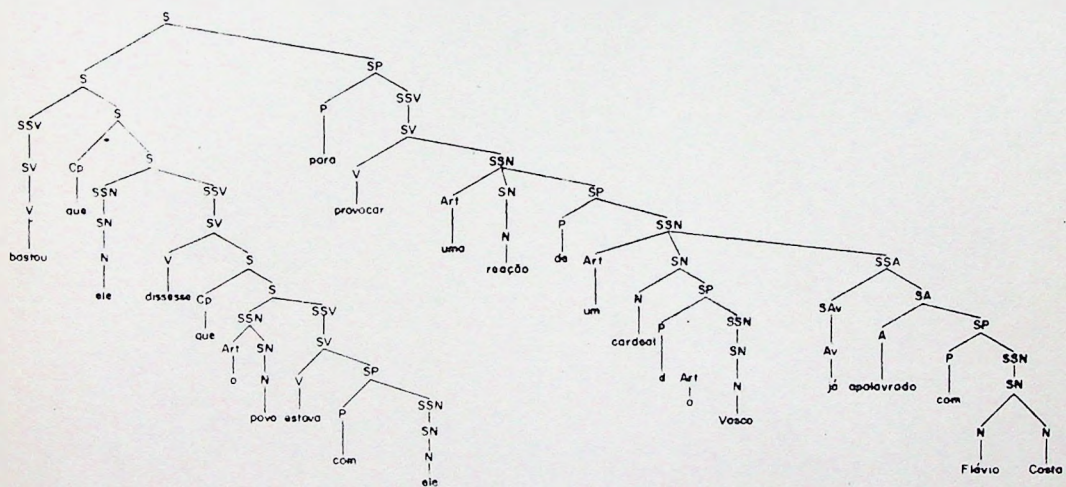
(586)



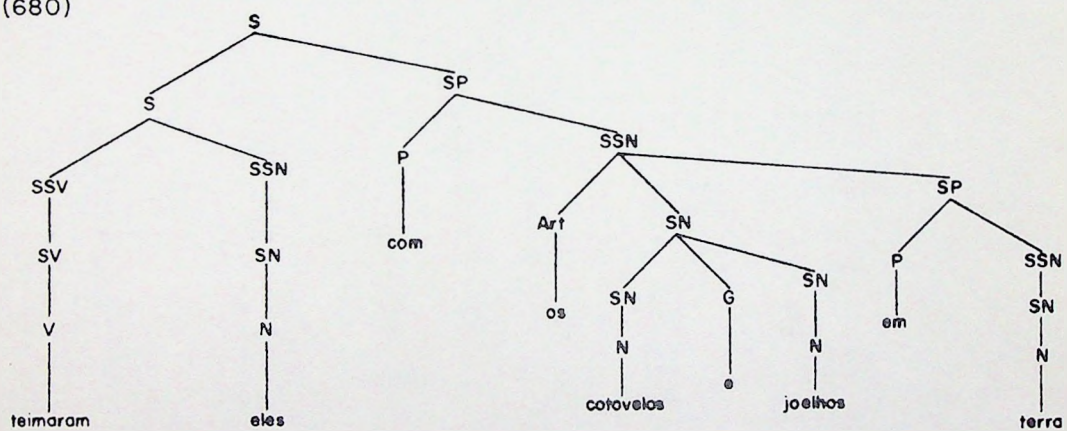
Há também modificadores sentenciais à direita da sentença nuclear. Os desse tipo, porém, nunca vêm em sobreposições numerosas. O motivo da assimetria esquerda-direita é claro: evitam-se os encaixes muito profundos à esquerda.



(209)



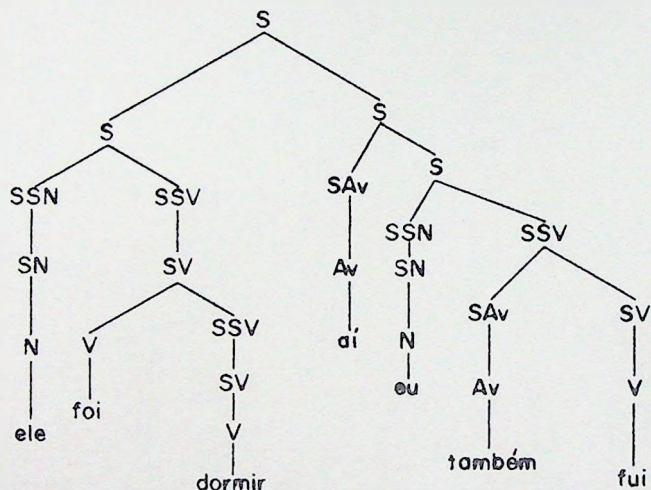
(680)



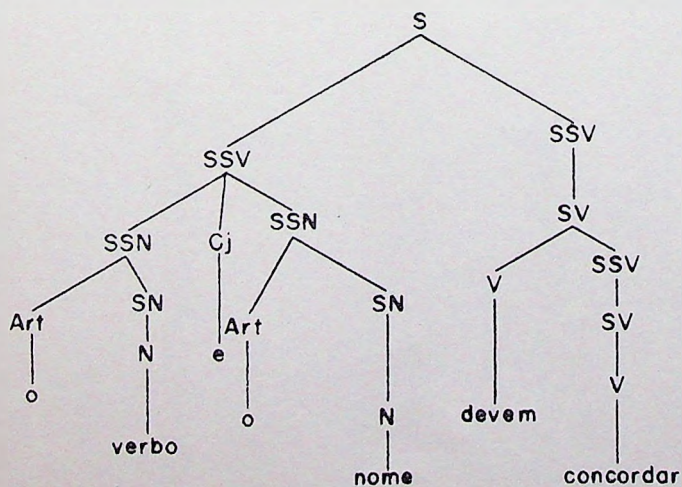
3.3.7 - A coordenação

Representamos a coordenação de constituintes reduplicando o nódulo constituído por sintagmas coordenados. A coordenação pode ser a qualquer nível sintático, conforme ilustraremos a seguir.

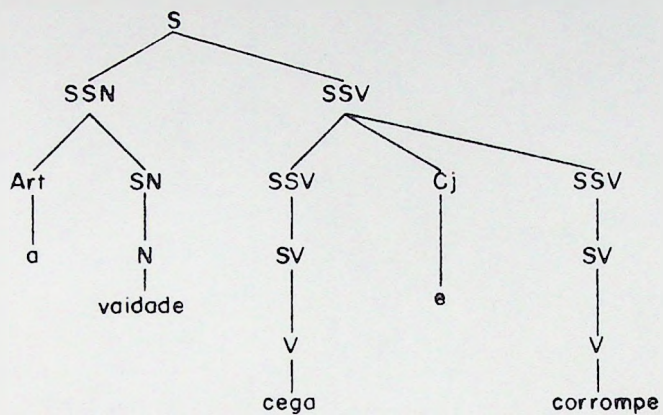
Coordenação de Sentenças:



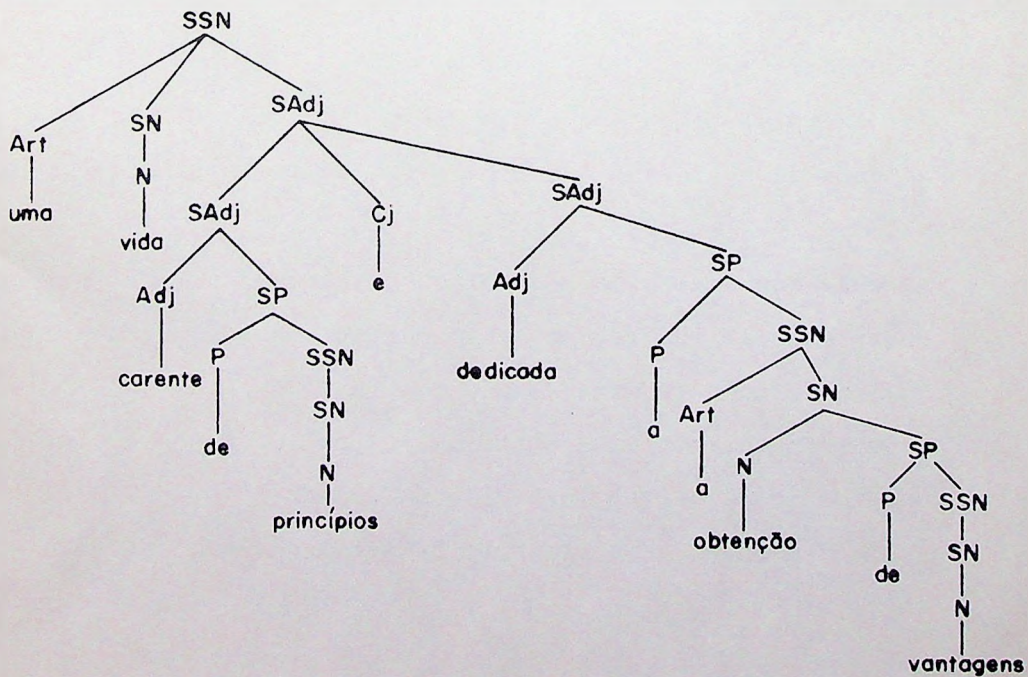
Coordenação de Sintagmas Nominais



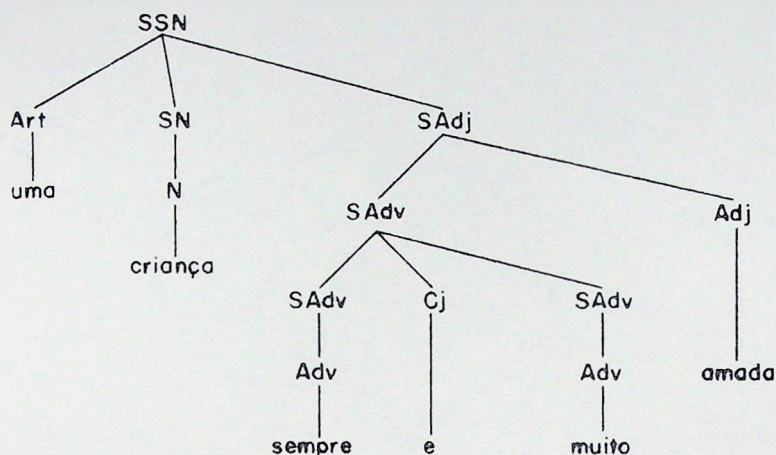
Coordenação de Sintagmas Verbais



Coordenação de Sintagmas Adjetivos



Coordenação de Sintagma Adverbial:



Consideramos como coordenação os casos de certos modificadores que, na escrita, vêm precedidos por vírgulas, vírgulas essas que não são de uso obrigatório, e que interpretamos como indicadores de uma coordenação de sintagmas da mesma espécie, o segundo dos quais truncado.

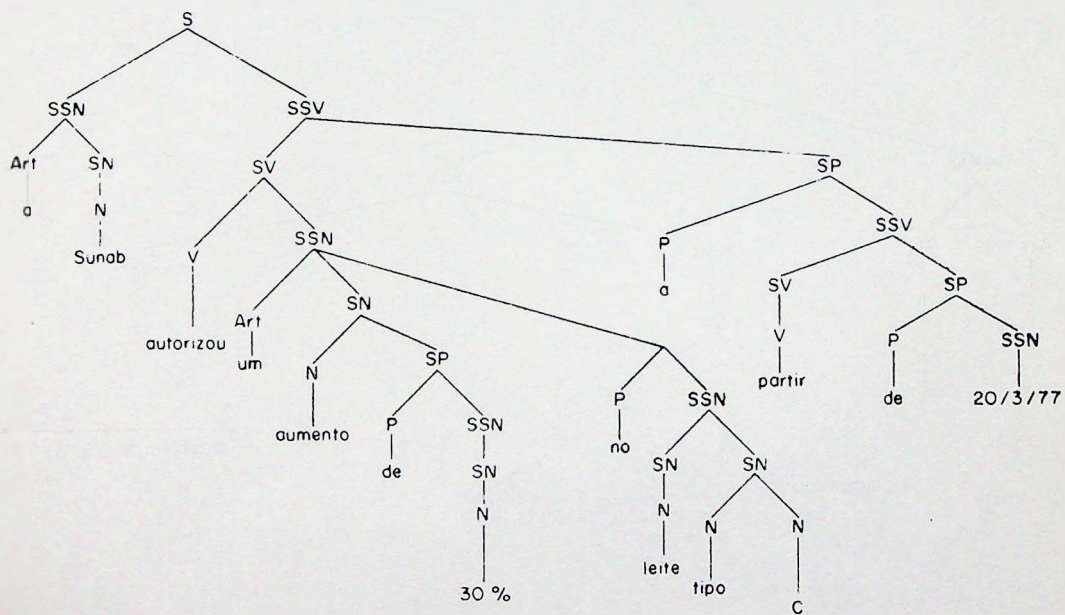
Por exemplo, tome-se a sentença

A Sunab autorizou um aumento de 30% no leite tipo C⁽¹⁾ a partir de 20/3/77.

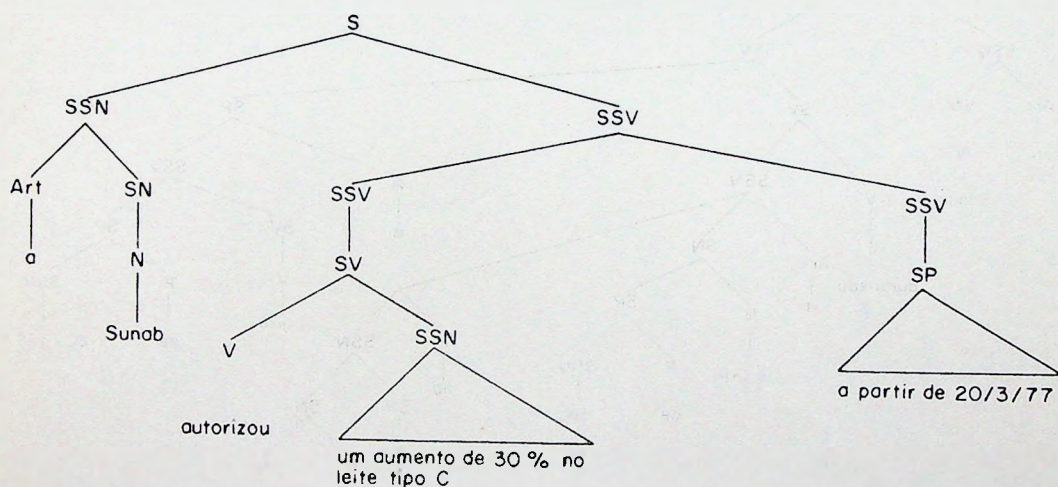
Note-se que a vírgula assinalada no círculo é facultativa. Como interpretá-la? Parece-nos que podemos considerar essa sentença equivalente, semanticamente, a

A Sunab autorizou um aumento de 30% no leite tipo C, e autorizou um aumento de 30% no leite tipo C a partir de 20/3/77.

Assim pensando, a arborização da sentença sem a vírgula seria



e a arborização com a vírgula seria



Outros exemplos do mesmo tipo são:

- a) Muitos preferem passar o domingo em casa, para ver televisão.

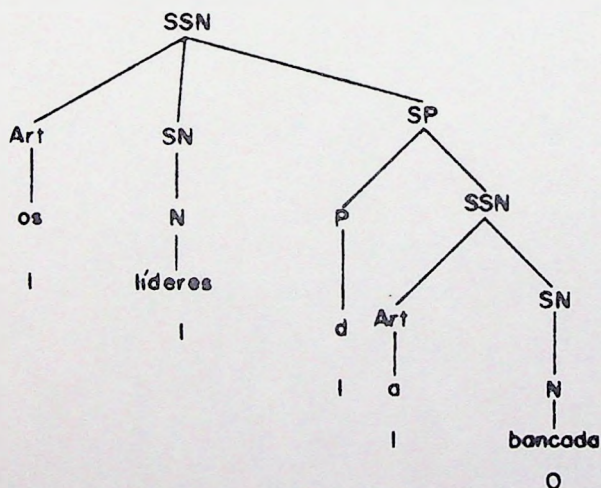
Nesse caso, diríamos que a versão com vírgula da sentença é equivalente a

Muitos preferem passar o domingo em casa e preferem passar

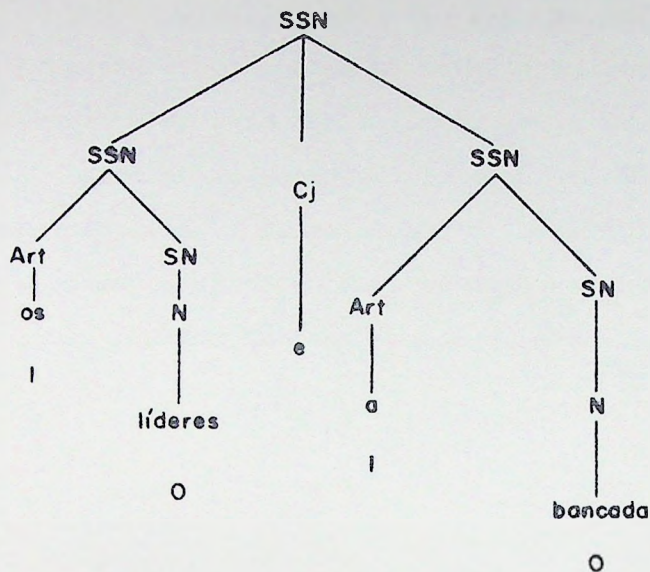
o domingo em casa para ver televisão.

- b) O sistema telefônico da cidade ficou bastante prejudicado⁽⁶⁾ devido ao forte temporal de ontem (= e ficou bastante prejudicado devido ao forte temporal de ontem)

Quanto à contagem do peso sintático, os constituintes coordenados foram avaliados como se cada um fosse o único, e seus pesos somados. Assim, a avaliação do sintagma nominal complexo os líderes da bancada



é de 4 própons, enquanto a do sintagma coordenado os líderes e a bancada



é de 2 própons, pois cada um dos seus elementos formadores, sozinho, pesa 1 própon.

3.4 - Operacionalização da quantificação: descrição do programa TREEBUILD (por Gregory R. Guy)

3.4.1 - Introdução

O programa TREEBUILD foi criado para o projeto de pesquisa "Competências Básicas do Português" a fim de facilitar o cálculo de diversas medidas da complexidade sintática das sentenças. Os dados são apresentados ao programa como árvores de estrutura superficial, em forma de parentização rotulada. O programa reconstrói uma representação interna da árvore para cada sentença, e em seguida calcula as medidas para os tipos de nódulos desejados.

(Instruções preliminares em cada rodada possibilitam ao usuário selecionar os tipos de nódulos para os quais as medidas deverão ser calculadas - por exemplo, Sintagma Nominal, Sintagma Verbal, Sentença - e possibilitam também suprimir ou permitir a perfuração da saída). Todos os valores resultantes são impressos em forma tabular para cada sentença da entrada, e poderão ser também perfurados em cartões, caso haja interesse. Ao fim de cada rodada são também impressas tabelas globais, incluindo os valores para todas as sentenças analisadas.

As medidas de complexidade sintática calculadas pelo programa são as seguintes: ISP (Índice de posição estrutural) - quantifica a posição de um nódulo na árvore; VEC (complexidade vertical ou peso) quantifica a complexidade estrutural de uma árvore inteira ou de qualquer subparte; HOC (complexidade horizontal) - o número de sintagmas constituintes de um dado nódulo; DENS (densidade) - o peso por número de palavras; TONY - número de nódulos similares dominando um dado nódulo; FLOOR - número de nódulos de qualquer tipo dominando dado nódulo.

O programa está escrito na linguagem de programação SNOBOL/SPITBOL.

3.4.2 - Estrutura geral (vide também Fluxograma)

- 1 - Preparações: a. inicialização, definição de sub-rotinas (linhas 1-26); b. criação de arrays e patterns (dependente da entrada - linhas 27 - 51).

- 2 - Entrada do texto de uma sentença (linhas 52 - 62). Caso não sejam encontradas outras sentenças mais, pula a 7.
- 3 - Reconstrução da árvore a partir da forma em parênteses rotulados
 - a. Preparações e procura de erros (linhas 63-72, e sub-rotina PARCO, 255-270).
 - b. Reconstrução da árvore
 - i. inicialização dos valores (74-77)
 - ii. reconstrução da árvore e cálculo de HOC, TONY e FLOOR para cada nóculo (Sub-rotina Treebuild, 184-222)
- 4 - Cálculo de ISP para todos os nóculos (Sub-rotina ISPCAL 149-183)
- 5 - Entrada nas tabelas de valores para os nóculos selecionados
 - a. Inicialização: 80-82, 85-87
 - b. Procura dos tipos de nóculos desejados (84, 105-106)
 - c. Cálculo de VEC para os nóculos desejados (Sub-rotina TERDAWT, 223-234)
 - d. Cálculo de DENS (89-90)
 - e. Entradas dos dados em tabelas locais (91-104)
- 6 - Saída dos dados para a sentença em análise, e entrada nas tabelas globais.
 - a. Imprimir cabeçalhos (107-111)
 - b. Imprimir dados e incluir em tabelas globais(112-126)

7 - Saída de tabelas globais (dados combinados para todas as sentenças) - quando todas as sentenças estiverem rodadas.

a. Ordenação com respeito a VEC e ISP (281-299)

b. Saída (300-302)

8 - END (302)

3.4.3 - Sub-rotinas principais

1. TREEBUILD. Executa a tarefa central de reconstituir uma representação utilizável da árvore na forma de um grafo ordenado. Esta sub-rotina é recursiva, utilizando uma pilha simulada (LIFO- 'last in first out') para armazenamento. Ela calcula também os valores de HOC, TONY e FLOOR, e reconstrói nódulos omitidos (segundo a convenção pela qual nódulos N, V, A e AV devem ser imediatamente dominados por SN, SV, SA e SAV, respectivamente, nódulos estes que por sua vez devem ser dominados por nódulos SSN, SSV, SSA e SSAV).

2. ISPCAL. Calcula os valores de ISP para todos os nódulos na árvore. Também esta sub-rotina é recursiva, e utiliza uma pilha. Contém rotinas para incluir exceções para nódulos não ramificados e nódulos coordenados.

3. TERDAWT. Calcula os valores de VEC para nódulos desejados, como se se encontrassem em posições de ISP zero. Para isso, localizam-se todos os nódulos terminais que são descendentes do nódulo em questão e adicionam-se os seus valores ajustados de ISP. É também uma sub-rotina recursiva e funciona através da simulação de uma pilha.

4. PARCOUNT. Conta parênteses à esquerda e à direita, a fim de estabelecer se a cadeia de entrada é bem formada. Produz mensagens de erro com respeito a erros de parentização, rotulação e erros em itens lexicais.

5. SWITCH e SWAP. Essas duas sub-rotinas têm efeitos idênticos: intercambiar os valores encontrados em dois pontos diversos da tabela. SWAP o faz para a tabela PU usada para saída perfurada, e SWITCH para as tabelas globais. Ambas as sub-rotinas são chamadas no decorrer das operações de ordenação de bolhas, cujo propósito é o de organizar e ordenar suas respectivas tabelas.

6. COMP. A função básica é a de eliminar, de cada registro de entrada, o número da sentença e letras indicadoras de sequência, deixando apenas a cadeia de parentização rotulada. Realiza também algumas funções de detecção de erros.

3.4.4 - Uso

Os dados são apresentados ao programa na forma de cadeias em parentização rotulada, perfuradas em cartões de computador. O formato dos cartões é o seguinte: colunas 1 a 4, número da sentença, justificado à direita; coluna 5, letra indicando a posição do cartão na sequência de cartões correspondentes à sentença em apreço (o último cartão de cada sentença recebe sempre a letra Z, ainda que seja o único cartão da sentença); coluna 6, vazia; colunas 7-80, dados. O deck de dados para o programa consiste de todos os cartões de dados em sequência apropriada, precedidos por um ou dois cartões de controle. Os cartões de controle, que deverão apa-

recer nessa ordem, são: NOPUNCH (facultativo), que é usada para evitar que os resultados sejam não só impressos comotambém per-
furados, e a lista de seleção de nódulos, que informa ao progra-
ma quais os nódulos que deverão ter suas medidas de complexida-
de calculadas.

Os cartões de controle apareceriam, portanto, como no exem-
plo seguinte:

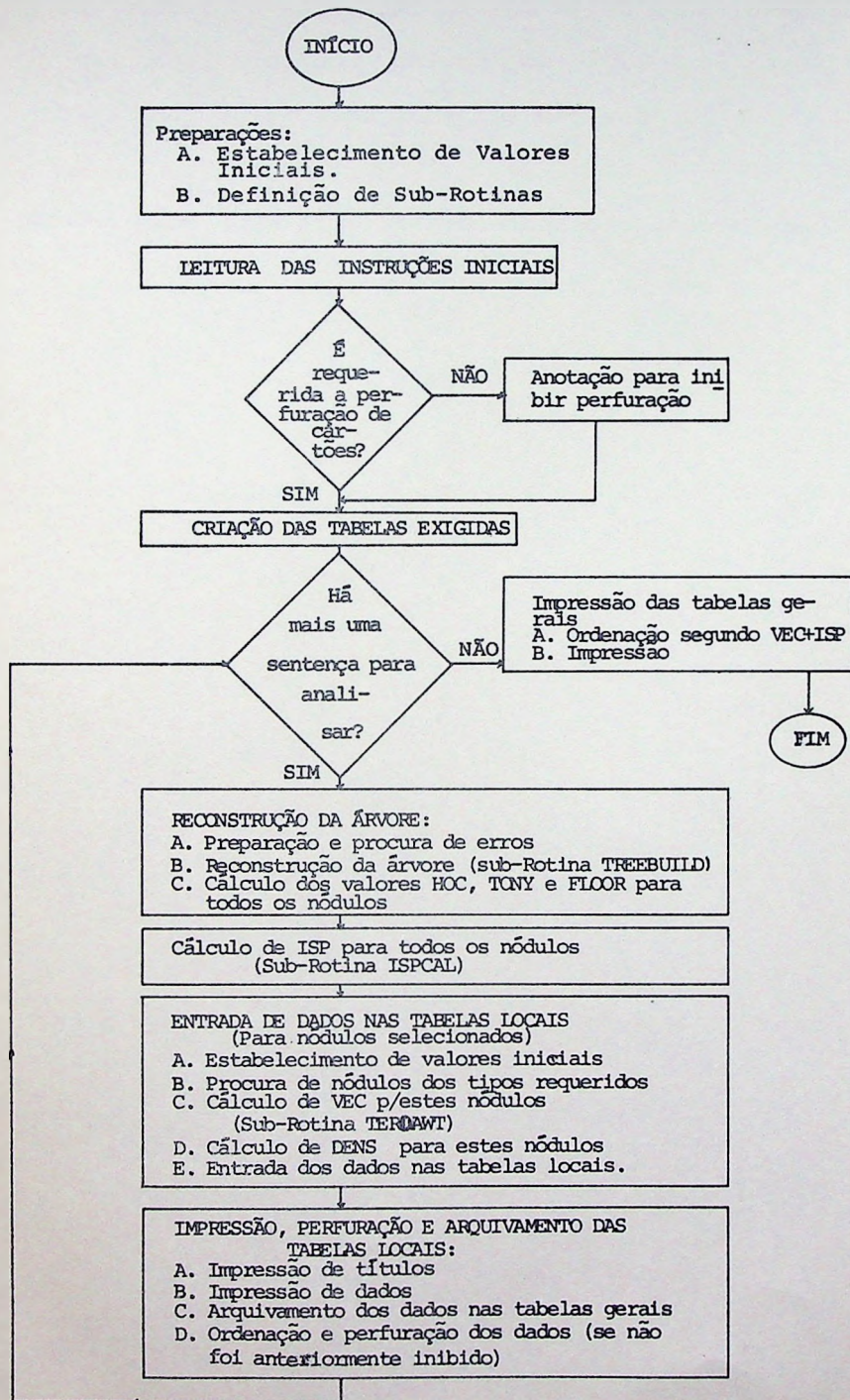
Coluna

1 2 3 4

NOPUNCH

So, S, SN, SV, ...

3.4.5 - FLUXOGRAMA GERAL - PROGRAMA TREEBUILD



3.5 - Etapas analíticas

A análise comparativa da complexidade sintática foi realizada através das seguintes etapas:

- 1^a etapa: seleção das sentenças da amostra. Vide descrição em 3.2.
- 2^a etapa: cada sentença foi recortada e colada numa ficha de cartolina, com um rótulo numérico e indicação da publicação e página à qual pertencia.
- 3^a etapa: análise sintática: cada sentença foi arborizada, segundo método descrito em 3.3.
- 4^a etapa: as árvores foram transpostas para a representação equivalente em parênteses rotulados, em folhas de codificação. O rótulo numérico de sentença foi mantido, e cada linha da parentização foi rotulada alfabeticamente, a última sempre Z.
- 5^a etapa: perfuração de cartões, a cada linha da parentização rotulada correspondendo um cartão.
- 6^a etapa: os cartões dão entrada no programa TREEBUILD (de Gregory R. Guy).
- 7^a etapa: a saída do programa imprime a parentização rotulada de cada árvore, e dá, para cada nóduo Sentença Principal, Sentença Encaixada, Sintagma Nominal, Sintagma Verbal, as seguintes medidas: ISP (Índice de Posição Estrutural), TONY (número de nóduos dominantes da mesma natureza) VEC (complexidade vertical ou peso), DENS(densidade, ou peso dividido por número de palavras), FLOOR (número de nódu-

los dominantes), HOC (Complexidade Horizontal, número de ramificações de um nódulo).

8^a etapa: a saída do programa TREEBUILD para as medidas de ISP e FLOOR em cada sentença serve de entrada para o subprograma AGGREGATE do Statistical Package for the Social Sciences, o qual calcula as médias e máximas de ISP e FLOOR por sentença (vide 3.6.1)

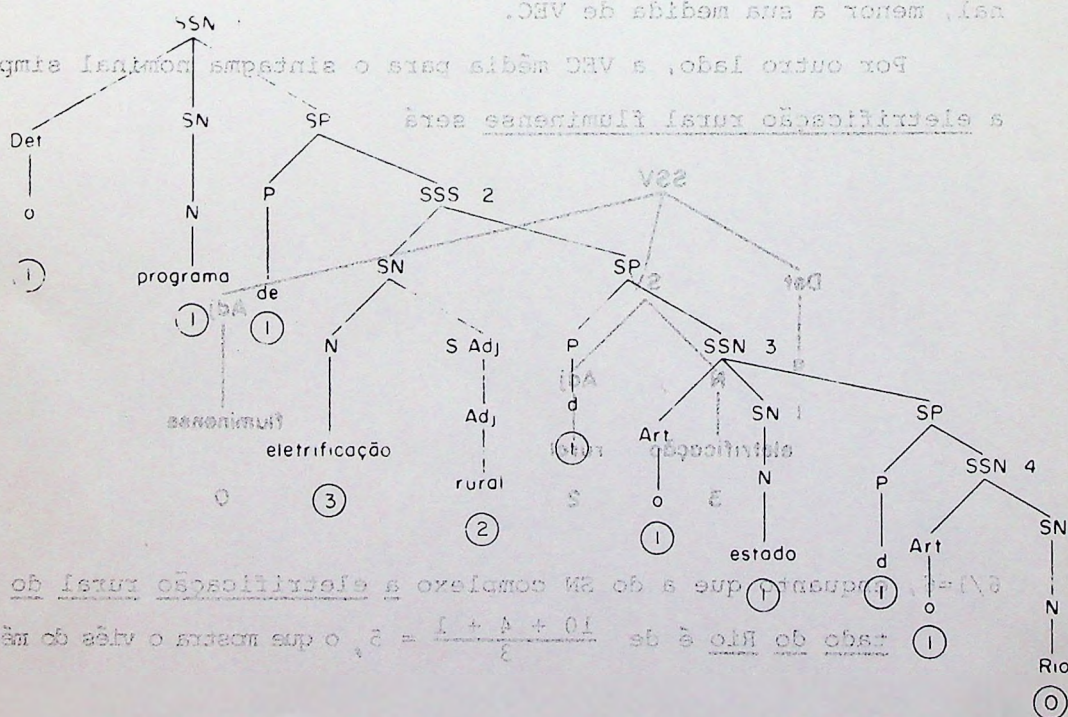
9^a etapa: a saída do programa TREEBUILD (e para as medidas de ISP e FLOOR, do subprograma AGGREGATE) serviu de entrada para o programa ONE WAY do Statistical Package for the Social Sciences H, versão 6.02. Esse programa calculou, para cada variedade de língua, para cada nódulo Sentença Principal, Sentença Encaixada, Sintagma Nominal, Sintagma Verbal, a probabilidade de F e, aproveitando uma opção, as médias, máximas, número de casos e intervalo de confiança de 95% para a média das variáveis em estudo (densidade e VEC).

10^a etapa: O teste de t de Student foi aplicado a vários pares de resultados para verificar se esses pares eram significativamente diferentes. Utilizou-se o subprograma T-TEST de SPSS.

3.6 - Resultados

3.6.1 - Descrição dos cálculos

Para os cálculos da densidade média e do peso médio de cada tipo sintagmático foram selecionados apenas os nódulos não auto-encaixados, i.e., os de $TONY = 0$. O motivo dessa seleção é que qualquer sintagma encaixado tem, necessariamente, peso menor do que o do sintagma que o contém e, portanto, se levássemos indiscriminadamente em conta os todos juntamente com suas partes, estaríamos indevidamente diminuindo as avaliações de complexidade das amostras que tivessem maior número de encaixes. O problema aparece concretamente ilustrado no exemplo abaixo.



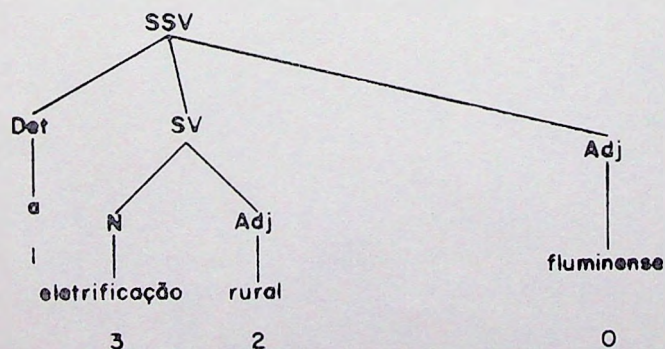
	TONY	VEC	DENS
SSN ₁	0	13	13/11 = 1,18
SSN ₂	1	10	10/8 = 1,25
SSN ₃	2	4	4/5 = 0,80
SSN ₄	3	1	1/2 = 0,50

$$\text{VEC Média} = (13 + 10 + 4 + 1) / 4 = 28 / 4 = 7,00$$

$$\text{DENS Média} = (1,18 + 1,25 + 0,80 + 0,50) / 4 = 0,93$$

A tabela mostra que quanto mais encaixado o sintagma nominal, menor a sua medida de VEC.

Por outro lado, a VEC média para o sintagma nominal simples a eletrificação rural fluminense será



6/1=6, enquanto que a do SN complexo a eletrificação rural do estado do Rio é de $\frac{10 + 4 + 1}{3} = 5$, o que mostra o viés do método.

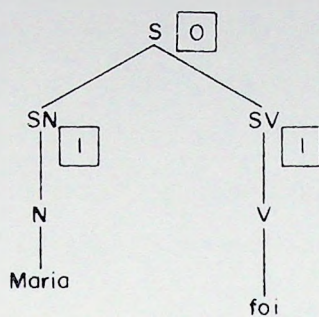
Como o problema é causado pelo fato de estar sendo calculada uma média de um todo com suas partes, a solução está em basear o cálculo da média de peso (VEC) de cada tipo de nódulo exclusivamente nos nódulos não auto-encaixados, i. e., os de TONY = 0. A diferença entre os dois métodos fica clara na tabela a seguir, onde é evidente que a média dos jornais, onde os encaixes são mais freqüentemente usados, é a que fica mais alterada.

Médias de VEC		
	Todos os SN	Só SN com TONY = 0
Mobralenses	2,11	2,33
Literatura & Foto-		
novelas	2,94	2,95
Jornais A & B	7,06	9,29

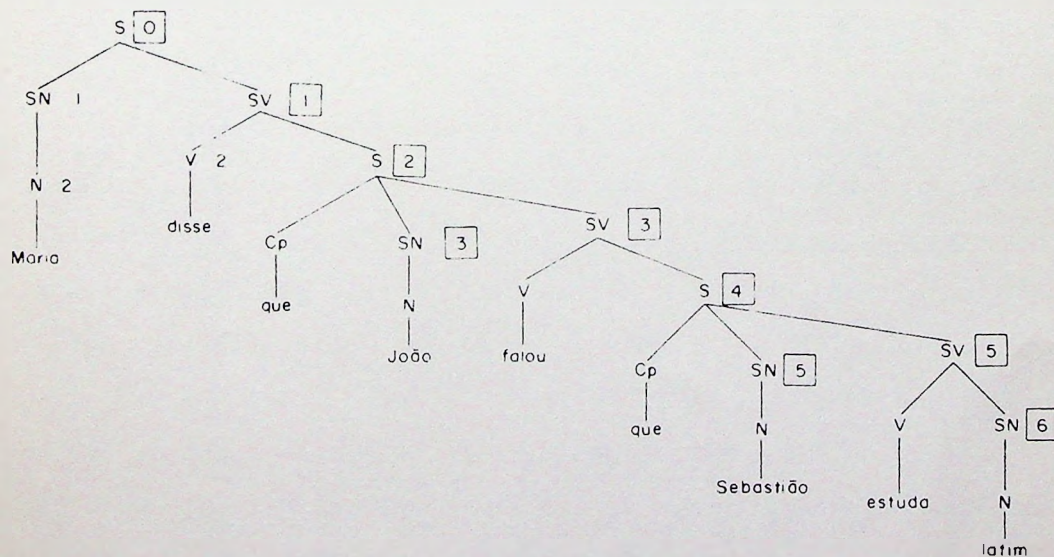
Na medida de densidade, existem casos em que um sintagma encaixado é mais denso. No sintagma nominal dado acima, o programa de eletrificação rural do Estado do Rio de Janeiro, por exemplo, SSN₂, encaixado em SSN₁, é o mais denso. Esse tipo de caso está sendo captado através do cálculo do máximo, pois aí não impomos a restrição de TONY = 0.

No cálculo das médias de ISP e FLOOR quisemos evitar que as sentenças mais ramificadas da amostra contribuíssem mais para a média do que as sentenças com menos nódulos, já que o que nos interessa saber é a profundidade a que chega cada grupo em uma sentença típica. Considere-se os seguintes exemplos de FLOOR. É preciso manter presente o fato de que dispomos de dados relativos

apenas aos nódulos principais.



$$FLOOR_{MED} = \frac{0 + 1 + 1}{3} = \frac{2}{3} = 0,7$$



$$FLOOR_{MED} = \frac{0 + 1 + 1 + 2 + 3 + 3 + 4 + 5 + 5 + 6}{10} = \frac{30}{10} = 3.0$$

Há dois métodos possíveis de calcular as médias de FLOOR para essa amostra de duas sentenças: 1 - somar as duas médias sentenciais e dividir por 2: $0.7 + 3.0 = 3.7/2 = 1.9$ 2 - somar os escores de cada nóculo das duas sentenças e dividir pelo número total de nósculos: $\frac{2 + 30}{3 + 10} = 2.5$. Obviamente o segundo método é aquele que apresenta o resultado indesejável de dar mais peso às sentenças mais profundamente encaixadas da amostra. Por essa razão, o cálculo das médias de encaixe foi feito em duas etapas: Primeiro calcularam-se as médias e máximas sentenciais através do subprograma AGGREGATE de SSPSS. O cálculo da média sentencial de ISP consiste da soma dos ISPs dos nósculos SN, SV e Sencaixada de cada sentença dividido pelo número desses nósculos. A máxima sentencial de ISP é o valor máximo de ISP em cada sentença. Do mesmo modo foram calculadas as médias e máximas sentenciais de FLOOR. Esses resultados foram usados como entrada do subprograma ONEWAY. Os cálculos feitos aí foram os seguintes:

média sentencial média de ISP é a soma das médias sentenciais de ISP dividido pelo número de sentenças.

média sentencial máxima de ISP é o maior valor encontrado para as médias sentenciais.

máxima sentencial média de ISP é a soma das máximas sentenciais de ISP dividido pelo número de sentença.

máxima sentencial máxima de ISP é o maior valor encontrado para as máximas sentenciais.

Pelo mesmo método foram calculadas as correspondentes medidas de FLOOR. Os resultados desses cálculos podem ser apreciados na

Tabela 18.

Nota-se imediatamente que o ISP médio dos Mobralenses é o mais alto, o que é uma consequência direta do método não-aleatório de amostragem usado com esse grupo, visando precisamente a colher as sentenças mais complexas. No entanto, tendo notado essa distribuição de encaixe nas amostras, levanta-se a possibilidade de os resultados apresentados sobre complexidade serem uma consequência direta de tal tipo de distribuição. Para resolver essa dúvida re-fizemos os cálculos de VEC e DENS_{média} para cada nódulo em quatro níveis de ISP (0, 1 a 3, 4 a 6, e acima de 6). Infelizmente essa atomização dos dados tem o efeito de diminuir consideravelmente o número de dados em cada célula. De fato, nem todos os cálculos deram probabilidades de F aceitáveis. Nas tabelas resumidas dadas a seguir, cada coluna representa duas rodadas em ONEWAY - uma para VEC e outra para DENS. Quando a probabilidade de F é diferente de zero, esses valores vão anotados na linha mais baixa da tabela. Como se vê, os resultados para histórias em quadrinhos não têm validade estatística para qualquer dos nódulos, enquanto que para as sentenças encaixadas o grupo literatura mais fotonovela também carece de validade estatística. Os resultados dados nessas colunas são, no entanto, consistentes com os das outras colunas.

Da tabela conclui-se naturalmente que a média da complexidade de um dado nódulo depende não só do grupo que o produz mas também da sua posição estrutural dentro da sentença como um todo. À primeira vista este fato sugeriria o uso de uma análise da variância em duas dimensões, mas uma rápida inspeção desta tabela revela que as relações entre as categorias são totalmente diferentes do modelo aditivo que subjaz a essa técnica. Por exemplo, o VEC dos

Mobralenses tem uma diferença de $4,1 - 1,6 = 2,5$ entre as primeiras duas categorias de ISP, enquanto que para os jornais essa mesma diferença é de $15,3 - 4,7 = 10,6$. Dado esse fato, desistimos de aplicar a análise da variância multi-dimensional ao caso.

3.6.2 - Medidas obtidas

Vide Tabelas 14 a 21, apresentadas a seguir.

TABELA Nº 14

MEDIDAS DE COMPLEXIDADE DO SINTAGMA NOMINAL

	DENS. MÉDIA*	DENS. MAX.	VEC. MÉDIA*	VEC. MÁXIMA	INTERVALO DE CONFIANÇA DE 95% PARA A MÉDIA DE DENSIDADE	NÚMERO DE CASOS
ELISABETH	0,5	1,9	2,9	26		135/*89
GEORGE F.	0,5	3,3	3,2	52		122/*76
MARLENE	0,4	3,6	1,8	48		324/*244
VANILDA	0,3	1,7	1,1	23		187/*151
PAULO	0,4	2,0	1,5	22		113/*94
MADALENA	0,3	1,6	1,6	14		143/*111
GEORGE C.	0,5	4,0	4,2	93		172/*111
HENRIQUE	0,4	3,5	2,1	48		292/*203
J. CARLOS	0,4	2,5	1,8	26		143/*104
ILZENI	0,4	2,0	2,2	28		137/*102
FAUSTINA	0,5	2,8	2,2	67		162/*120
LIDIO	0,4	3,2	2,1	57		173/*127
CANTIDIO	0,6	4,0	3,6	63		145/*92
HILDA	0,4	3,8	2,5	60		159/*111
SIDNEI	0,5	2,5	2,5	32		160/*112
WALDECIO	0,5	2,4	2,8	23		154/*99
GEORGE M.	0,3	2,4	1,4	21		178/*145
UBIRAJARA	0,5	2,9	3,5	49		215/*128
SONIA	0,4	2,6	2,5	46		188/*128
ELVIRA	0,5	2,8	2,9	34		178/*122
TODOS OS MOBRALENSES	0,4	4,0	2,3	93	0,40 a 0,44	3480/*2469
HIST. QUADRINHOS	0,4	1,5	1,2	12	0,29 a 0,43	104/*90
FOTONOVelas	0,5	2,0	2,7	32		539/*285
LIT. NACIONAL	0,5	3,2	3,0	80		1706/*996
FOTONOV. & LIT. NAC.	0,5	3,2	3,0	80	0,48 a 0,52	2245/*1281
JORNAIS "A"	0,8	6,4	9,0	195		2030/*700
JORNAIS "B"	0,8	7,0	9,6	744		2065/*670
JORNAIS "A" & "B"	0,8	7,0	9,3	744	0,75 a 0,82	4095/*1370

* ONDE T = 0

TABELA Nº 15

MEDIDAS DE COMPLEXIDADE DO SINTAGMA VERBAL

	DENS. MÉDIA*	DENS. MÁXIMA	VEC. MÉDIA*	VEC. MÁXIMA	INTERVALO DE CONFIANÇA DE 95% PARA A MÉDIA DE DENSIDADE	NÚMERO DE CASOS
ELISABETH	1,1	2,7	9,9	36		107/*47
GEORGE F.	1,0	3,9	10,6	74		121/*47
MARLENE	1,0	5,8	10,3	144		282/*120
VANILDA	1,1	3,2	11,9	55		152/*58
PAULO	0,9	3,4	7,7	45		113/*52
MADALENA	1,1	4,7	10,0	98		105/*55
GEORGE C.	1,1	3,8	12,2	75		108/*53
HENRIQUE	1,0	4,6	9,2	72		239/*117
J. CARLOS	0,8	3,5	6,1	38		118/*64
ILZENI	1,3	5,8	14,7	112		135/*51
FAUSTINA	1,3	3,7	12,1	69		123/*56
LIDIO	1,0	3,6	7,4	58		132/*85
CANTIDIO	1,4	4,2	12,8	64		86/*51
HILDA	1,4	4,4	17,6	63		147/*51
SIDNEI	1,1	3,4	9,6	60		137/*69
WALDECIO	1,3	4,4	13,1	80		114/*56
GEORGE M.	1,1	4,4	13,8	100		170/*68
UBIRAJARA	1,3	4,6	24,0	205		198/*52
SONIA	1,4	5,8	17,8	92		139/*56
ELVIRA	1,3	5,3	16,2	197		128/*55
TODOS MOBRALENSES	1,1	5,8	11,9	205	1,08 a 1,18	2854/*1263
HIST. QUADRINHOS	0,9	2,9	5,3	41	0,72 a 1,00	101/*63
FOTONOVelas	1,1	5,9	11,0	115		314/*170
LIT. NACIONAL	0,9	6,1	9,5	248		1199/*583
FOTONOV. & LIT.NAC.	1,0	6,1	9,8	248	0,94 a 1,04	1513/*753
JORNAIS "A"	1,3	6,7	24,6	363		719/*318
JORNAIS "B"	1,2	6,8	23,9	746		786/*305
JORNAIS "A" & "B"	1,3	6,8	24,2	746	1,19 a 1,34	1505/*623

* ONDE $T' = 0$

TABELA Nº 16

MEDIDAS DE COMPLEXIDADE DE SENTENÇA ENCAIXADA

	DENS. MÉDIA*	DENS. MÁXIMA	VEC. MÉDIA*	VEC. MÁXIMA	INTERVALO DE CONFIANÇA DE 95% PARA A MÉDIA DE DENSIDADE	NÚMERO DE CASOS
ELISABETH	1,3	2,7	12,3	41		79/*43
GEORGE F.	1,2	4,5	13,0	74		/*40
MARLENE	1,4	5,4	17,9	93		275/*91
VANILDA	1,4	4,3	16,5	60		150/*49
PAULO	1,5	4,7	14,4	61		75/*36
MADALENA	1,5	4,7	16,4	102		99/*43
GEORGE C.	1,6	4,5	19,7	76		101/*41
HENRIQUE	1,7	4,7	22,8	128		265/*80
J. CARLOS	1,5	5,0	16,3	104		97/*39
ILZENI	1,8	5,4	22,2	159		101/*42
FAUSTINA	1,9	6,0	28,8	199		109/*35
LIDIO	1,6	5,6	16,0	72		124/*61
CANTIDIO	2,0	6,7	25,3	148		66/*31
HILDA	1,9	6,2	26,6	124		122/*41
SIDNEI	1,5	6,2	16,5	87		103/*48
WALDECIO	1,9	7,4	22,6	155		94/*40
GEORGE M.	2,2	6,0	35,7	245		170/*44
UBIRAJARA	2,4	6,3	48,7	189		139/*38
SONIA	2,5	6,4	44,2	186		128/*37
ELVIRA	2,1	6,4	36,9	243		118/*37
TODOS OS MOBRALENSES	1,7	7,4	22,9	245	1,62 a 1,77	2509/*916
HIST. QUADRINHOS	0,8	2,9	5,6	27	0,59 a 1,01	64/*44
FOTONOVelas	1,3	7,4	15,2	356		207/*127
LIT. NACIONAL	1,2	6,5	14,5	248		752/*403
FOTONOV. & LIT.NAC.	1,2	7,4	14,7	356	1,15 a 1,30	959/*530
JORNAIS "A"	1,9	10,0	45,3	369		506/*204
JORNAIS "B"	1,8	5,9	35,7	317		470/*225
JORNAIS "A" & "B"	1,8	10,0	40,3	368	1,74 a 1,96	976/*429

TABELA Nº 17

MEDIDAS DE COMPLEXIDADE DA SENTENÇA PRINCIPAL

	DENS. MÉDIA	DENS. MÁXIMA	VEC. MÉDIA	VEC. MÁXIMA	INTERVALO DE CONFIANÇA DE 95% PARA A MÉDIA DE DENSIDADE	NÚMERO DE CASOS
ELISABETH	1,9	3,9	28,2	66		30
GEORGE F.	2,1	5,3	27,7	74		31
MARLENE	2,1	5,8	52,8	144		43
VANILDA	2,2	7,2	46,0	173		30
PAULO	2,3	4,5	32,2	112		30
MADALENA	2,3	5,3	38,1	121		30
GEORGE C.	2,4	5,9	45,0	171		30
HENRIQUE	2,4	8,3	58,3	256		43
J. CARLOS	2,4	7,4	40,5	111		30
ILZENI	2,5	4,7	46,6	186		30
FAUSTINA	2,5	5,9	50,2	200		30
LIDIO	2,5	6,7	45,0	134		30
CANTIDIO	2,6	5,5	44,2	149		30
HILDA	2,6	4,8	48,0	130		30
SIDNEI	2,8	7,4	51,8	184		30
WALDECIO	2,8	6,5	47,2	156		30
GEORGE M.	2,9	5,7	63,9	246		30
UBIRAJARA	2,9	6,1	75,6	206		30
SONIA	3,1	5,5	65,4	187		30
ELVIRA	3,3	6,2	72,4	257		30
TODOS OS MOBRALENSES	2,5	8,3	49,2	257	2,42 a 2,62	627
HIST. QUADRINHOS	1,2	4,6	9,4	73	0,94 a 1,39	60
FOTONOVELAS	1,5	7,3	23,8	356		139
LIT. NACIONAL	1,6	6,6	25,8	248		400
FOTONOV. & LIT.NAC.	1,5	7,3	25,3	356	1,45 a 1,63	539
JORNAIS "A"	2,4	11,7	78,8	539		198
JORNAIS "B"	2,4	6,8	80,2	746		200
JORNAIS "A" & "B"	2,4	11,7	79,5	746	2,27 a 2,54	398

TABELA Nº 18
MEDIDAS DE PROFUNDIDADE DE ENCAIXE

	MÉDIAS SENTENCIAIS				MÁXIMAS SENTENCIAIS				INTERVALO DE CONFIANÇA 95% PARA AS MÉDIAS SENTENCIAIS MÉDIAS DE ISP	NÚMERO DE CASOS
	M É D I A S		M Á X I M A S		M É D I A S		M Á X I M A S			
	ISP	FLOOR	ISP	FLOOR	ISP	FLOOR	ISP	FLOOR		
ELISABETH	0,6	4,1	1,6	6,9	3,0	9,1	6	15		30
GEORGE F.	0,9	4,7	2,6	9,3	3,4	10,0	8	18		31
MARLENE	0,8	5,0	3,2	9,5	4,3	10,9	11	18		43
VANILDA	0,9	5,3	2,3	12,5	3,8	11,2	11	21		30
PAULO	0,9	3,7	3,0	5,8	3,5	8,0	9	13		30
MADALENA	0,9	4,3	2,7	7,1	3,8	9,4	10	16		30
GEORGE C.	0,9	4,4	3,4	8,1	3,9	9,3	12	15		30
HENRIQUE	1,0	4,9	6,1	9,3	4,3	10,6	13	22		43
J. CARLOS	1,0	3,9	6,6	7,9	3,7	8,1	13	14		30
ILZENI	0,9	4,6	2,7	8,7	3,6	9,8	8	18		30
FAUSTINA	1,0	4,4	3,7	9,3	4,3	9,6	11	18		30
LIDIO	1,0	3,8	4,5	7,6	4,2	8,5	12	16		30
CANTIDIO	1,1	3,7	3,6	6,0	4,0	7,7	11	13		30
HILDA	1,1	5,3	2,6	10,5	4,5	10,7	10	20		30
SIDNEI	1,2	4,1	3,0	7,2	4,8	8,8	11	13		30
WALDECIO	1,1	4,0	3,2	9,1	4,3	8,4	13	18		30
GEORGE M.	1,3	5,1	3,5	8,5	5,6	11,0	10	21		30
UBIRAJARA	1,4	6,6	4,0	11,1	5,6	13,6	11	24		30
SONIA	1,4	5,1	3,2	11,5	5,7	11,4	12	23		30
ELVIRA	1,6	4,7	4,5	8,5	5,5	10,5	12	19		30
TODOS OS MOBRALENSES	1,0	4,6	6,6	12,5	4,3	9,9	13	24	0,98 a 1,12	624
HIST. QUADRINHOS	0,3	2,2	2,3	5,2	1,1	4,8	8	11	0,16 a 0,38	60
FOTONOVelas	0,4	3,1	5,5	11,3	2,0	6,7	20	22		139
LIT. NACIONAL	0,5	3,4	5,2	11,0	2,0	7,2	13	23		400
FOTONOV. & LIT. NAC.	0,4	3,3	5,5	11,3	2,0	7,1	20	23	0,39 a 0,50	539
JORNAIS "A"	0,9	5,3	6,9	20,7	4,3	11,7	23	42		198
JORNAIS "B"	0,9	5,1	4,8	20,6	4,5	11,3	17	32		200
JORNAIS "A" & "B"	0,9	5,2	6,9	20,7	4,4	11,5	23	42	0,81 a 0,99	398

RELAÇÃO ENTRE COMPLEXIDADE E GRAU DE ENCAIXE

PARA SINTAGMAS NOMINAIS *

I S P		MOBRALEN SES	HISTÓRIAS EM QUADRINHO	LITERATURA & FOTONOVELA	JORNAIS "A" & "B"	TODOS OS DADOS
0	VEC	4,1	1,3	4,3	15,3	7,4
	DENS	0,6	0,4	0,6	1,0	0,7
	N	769	52	641	612	2.074
1 a 3	VEC	1,6	1,0	1,5	4,7	2,4
	DENS	0,3	0,3	0,4	0,6	0,4
	N	1304	34	583	694	2.615
4 a 6	VEC	1,3	1,0	1,9	1,8	1,4
	DENS	0,3	0,3	0,5	0,5	0,4
	N	329	4	48	51	432
≥ 7	VEC	0,9		2,6	2,8	1,3
	DENS	0,3		0,4	0,5	0,3
	N	67		9	13	89

PROB. F VEC 0,88
 DENS 0,88

* SINTAGMAS NOMINAIS COM T = 0

T A B E L A N º 2 0

RELAÇÃO ENTRE COMPLEXIDADE E GRAU DE ENCAIXE

PARA SINTAGMAS VERBAIS *

I S P		MOBRALEN SES	HISTÓRIAS EM QUADRINHO	LITERATURA & FOTONOVELA	JORNAIS "A" & "B"	TODOS OS DADOS
0	VEC	14,7	5,8	10,9	29,1	16,6
	DENS	1,3	0,9	1,1	1,4	1,2
	N	898	54	634	483	2.069
1 a 3	VEC	5,4	2,1	4,4	7,2	5,6
	DENS	0,8	0,5	0,6	0,8	0,8
	N	318	9	103	125	555
4 a 6	VEC	3,8		3,3	9,5	4,7
	DENS	0,7		0,6	0,9	0,7
	N	41		16	13	70
<u>≥ 7</u>	VEC	0,2			7,0	1,9
	DENS	0,1			1,3	0,4
	N	6			2	8

PROB. F	VEC	0,52	0,01
	DENS	0,29	

* SINTAGMAS VERBAIS COM T = 0

T A B E L A N º 2 1

RELAÇÃO ENTRE COMPLEXIDADE E GRAU DE ENCAIXE

PARA SENTENÇAS ENCAIXADAS *

I S P		MOBRALEN SES	HISTÓRIAS EM QUADRINHO	LITERATURA & FOTONOVELA	JORNAIS "A" & "B"	TODOS OS DADOS
0	VEC	26,2	5,7	15,1	42,6	26,1
	DENS	1,8	0,8	1,2	1,9	1,6
	N	755	41	499	386	1.681
1 a 3	VEC	7,7	3,7	8,5	21,2	10,0
	DENS	1,1	0,4	1,0	1,5	1,1
	N	152	3	29	36	220
4 a 6	VEC	4,0		10	11,7	7,9
	DENS	0,9		1,5	1,1	1,0
	N	8		2	7	17
<u>≥ 7</u>	VEC	2,0				2,0
	DENS	0,7				0,7
	N	1				1

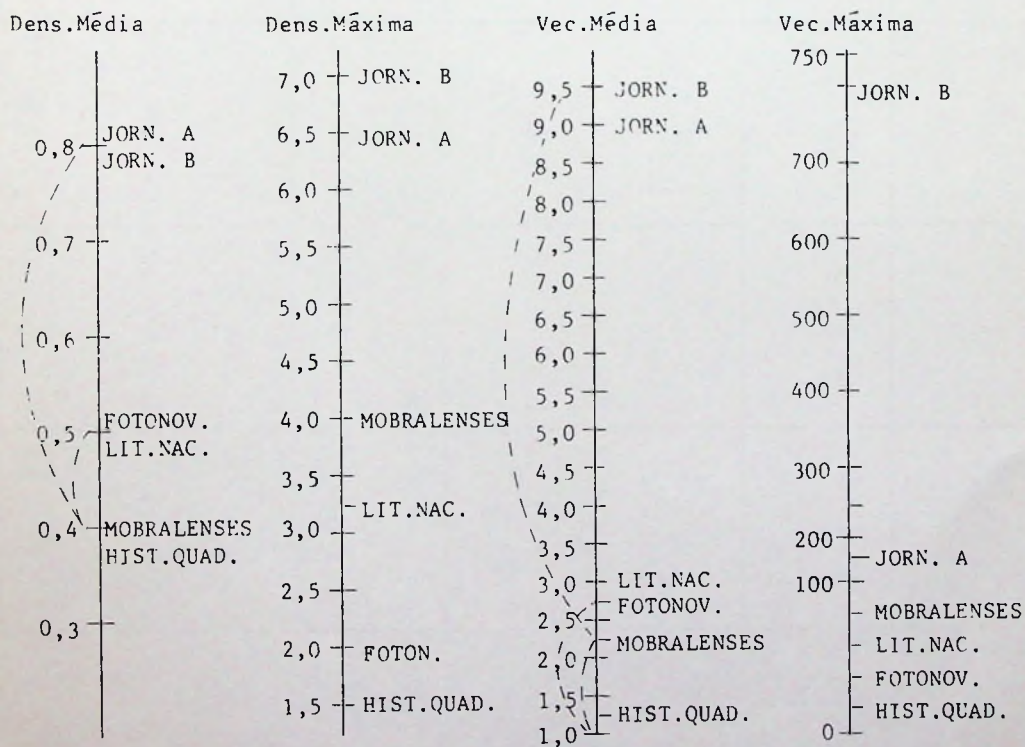
PROB. F	VEC	0,96	0,57	0,04
	DENS	0,80	0,50	0,08

* SENTENÇAS ENCAIXADAS COM T = 0

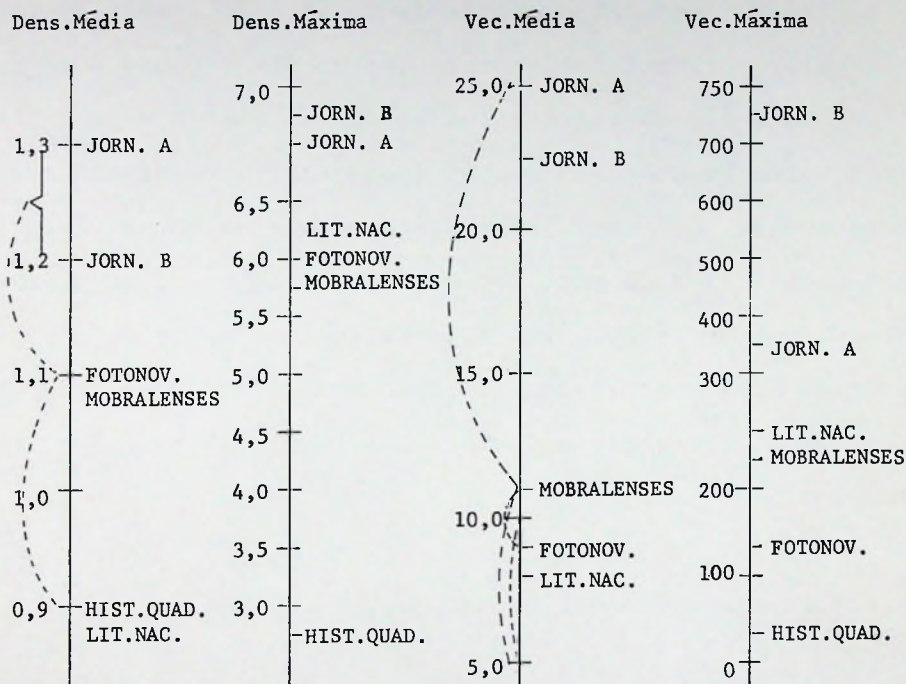
3.6.2.1 - Os sintagmas menores

Tomemos as tabelas 14, 15 e 16, que dão as medidas de complexidade dos Sintagmas Nominais, Sintagmas Verbais e Sentenças Encaixadas de cada variedade e coloquemos os resultados de cada uma das medidas ali apresentadas numa escala. As linhas pontilhadas indicam os pares de resultados que foram submetidos ao teste estatístico de Student, e nos quais o teste mostrou serem significativas as diferenças constatadas.

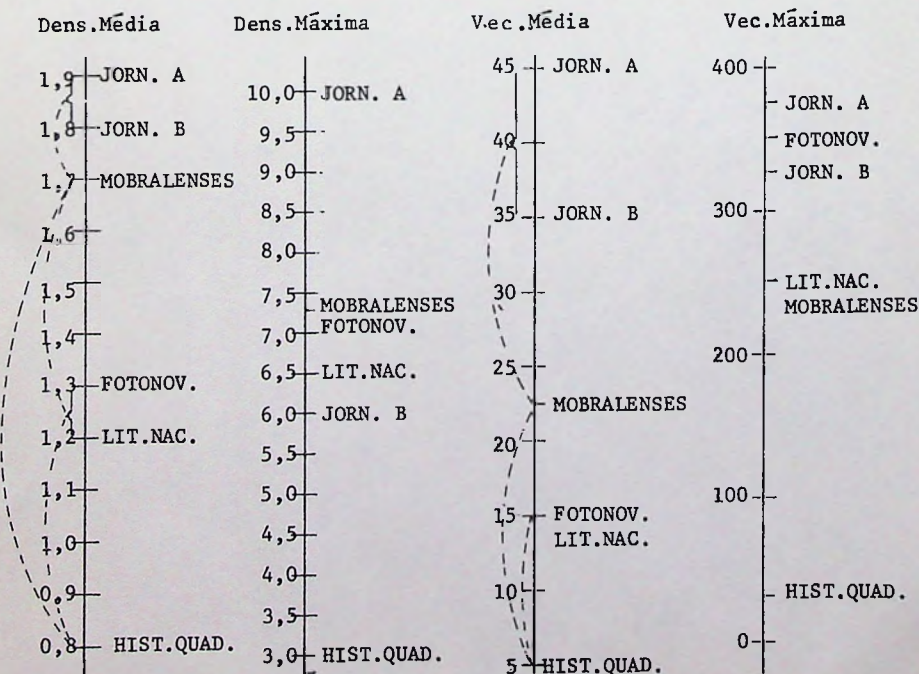
SINTAGMA NOMINAL



SINTAGMA VERBAL



SENTENÇA ENCAIXADA

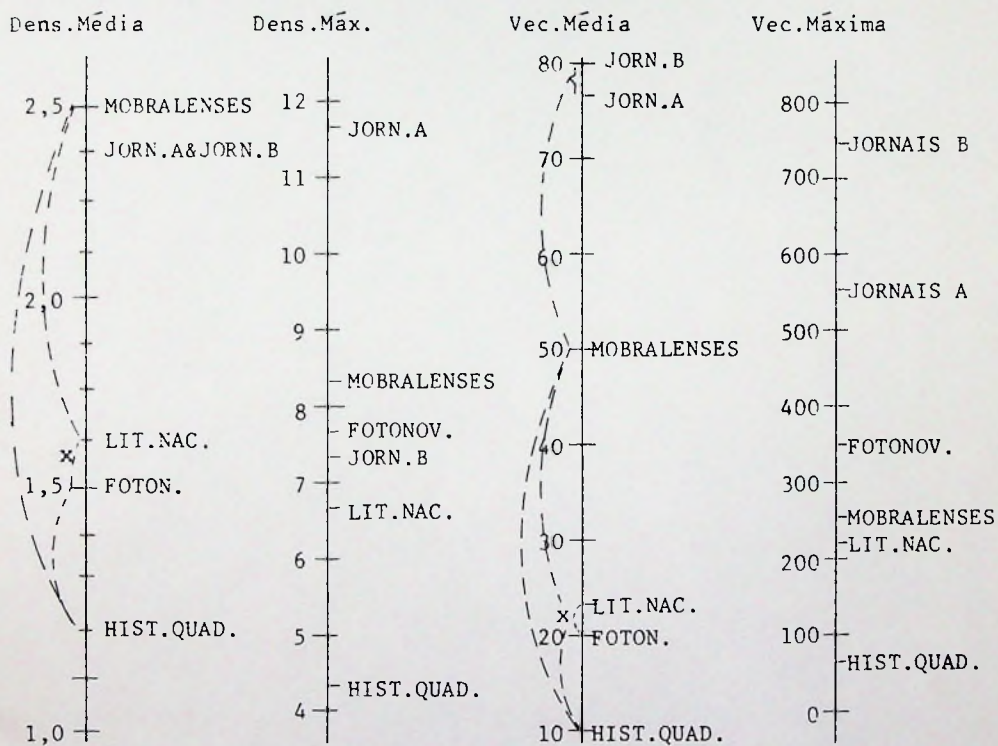


Pode-se ver que as medidas de complexidade constatadas para os Mobralenses, comparadas com as da amostra de língua escrita, só estão regular e significativamente abaixo das medidas dos jornais.

Quanto à língua escrita, vê-se que a amostra de literatura escolhida é menos complexa que a dos jornais, regulando com a das fotonovelas. Os resultados obtidos para os jornais não confirmaram a estratificação suposta. As histórias em quadrinhos ficaram consistentemente no nível mais baixo de complexidade em todas as medidas. Tais resultados justificam as médias conjuntas entre fotonovelas & literatura e jornais A & jornais B, apresentadas nas tabelas.

3.6.2.2 - As sentenças principais

Da tabela 17, derivamos os gráficos abaixo.

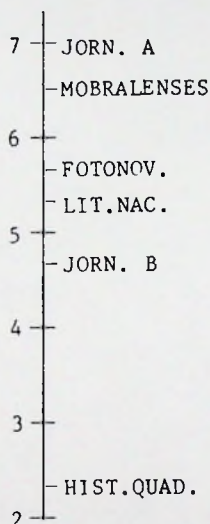


NOTA: Os arcos com x indicam a falta de significância estatística da diferença.

A ordenação das medidas da densidade média apresenta uma inversão notável em relação às tabelas anteriores: os mobralenses atingiram, em suas sentenças de maior complexidade, valores médios que alcançam, e até superam ligeiramente, a média de densidades dos jornais. Esse fato deve ser atribuído ao método de amostragem das sentenças: aleatório para a língua escrita, porém não para a língua falada, onde intencionalmente escolhemos sentenças longas e contendo grandes ramificações para a esquerda, configuração essa que resulta em medidas de densidade elevadas. A tabela 18, onde são comparados os níveis de encaixe atingidos nas diversas variedades, confirma isso: a tabela das médias de Índice de Posição

Do mesmo modo, estão semelhantemente ordenados os resultados das médias de ISP máximo (tabela 18, 3a. coluna) e os das densidades máximas (tabela 17, 2a. coluna):

Médias Sentenciais Máximas de ISP (derivada da tabela 18, 3a. coluna)



A pergunta que nos propuséramos ao deliberarmos escolher não-aleatoriamente as sentenças para a amostra da fala dos mobralenses foi: atingirão as sentenças mais complexas dos mobralenses os índices de complexidade média dos jornais? A tabela de densidade da sentença principal mostra que a resposta a essa pergunta é afirmativa. Quer dizer: com unidades constituintes menos complexas, em média, que as dos jornais, através da superposição de encaixes em ramificações à esquerda, as sentenças mais complexas desses falantes ligeiramente superam em complexidade a média de complexidade da amostra dos jornais.

3.6.2.3 - Relação entre complexidade dos sintagmas e seu nível de encaixe

As tabelas 14, 15 e 16 nos mostraram que os sintagmas nominais, verbais e sentenças encaixadas dos mobaralenses são menos complexos que os dos jornais. As tabelas 19, 20 e 21 nos mostram que essa constatação é verdadeira para todos os níveis de ISP. Lidas no sentido horizontal, essas tabelas nos informam, em cada nível de ISP, e para cada tipo de sintagma, qual a hierarquia de complexidade das variedades lingüísticas comparadas.

Lidas no sentido vertical, demonstram essas mesmas tabelas que existe uma relação inversa entre a profundidade de encaixe (ISP) de um sintagma e a sua complexidade. Quanto mais profundamente encaixados, menos complexos os sintagmas tendem a ser. Estes resultados proporcionam evidência empírica consistente com a concepção do processamento mental das sentenças exposta em 3.1.2. e do método de ponderação da complexidade sintática aqui proposto.

Vemos, assim, que a maneira de conceber a forma de processamento mental da estrutura sintática tornou possível a elaboração de um método de mensuração que por sua vez tornou possível registrar medidas que vêm a ser evidências favoráveis à concepção da qual decorrem. Há circularidade, mas não tautologia, nesse andamento, pois os resultados obtidos pelo método não o foram por necessidade lógica.

3.6.3 - Conclusões práticas

- 1) Em termos da competência lingüística de que já dispõe pelo domínio da língua falada, o adulto alfabetizado não deverá ter maiores dificuldades em tornar-se um leitor de variedades da língua escrita equivalentes em complexidade às das histórias em quadrinhos, fotonovelas e (boa parte d) a literatura nacional.
- 2) Na leitura dos jornais, encontrará uma percentagem pequena de sentenças de complexidade maior do que aquelas que estão dentro dos seus limiares superiores de uso. Por exemplo, a informante Elisabeth, cuja densidade máxima de S_0 foi de 3,88, encontrará 15% das sentenças de jornais A acima dessa medida. Em literatura nacional, 5%, em histórias em quadrinhos, 2%, em fotonovelas 1%.
A informante Ilzeni, cuja densidade máxima é 4,73, encontrará, nos jornais A, 6% das sentenças com valores superiores a esse. Em literatura nacional, 2%, histórias em quadrinhos, 0%, fotonovelas 1%.
A informante Elvira, cuja densidade máxima é de 6,18, encontrará nos jornais A 1% das sentenças acima dessa medida. Na literatura nacional, 0,5%.
- 3) Um programa planejado de treinamento de leitura deveria tomar em consideração a relação inversa existente entre a complexidade de um sintagma e a sua posição estrutural, e graduar os textos em termos dessa constatação. Assim, poderíamos ter o seguinte tipo de gradação de dificuldades.

- 1^a fase: São textos contendo constituintes bem simples.
- 2^a fase: Textos contendo constituintes gradativamente mais complexos em posições estruturais de índice zero.
- 3^a fase: Textos contendo constituintes gradativamente mais complexos em posições estruturais de índice 1 a 3.
- 4^a fase: Textos contendo constituintes gradativamente mais complexos em posições estruturais de índice igual ou superior a 4.

3.6.4 - Seguimentos desejáveis da pesquisa

- 1 - Acrescentar um estudo da complexidade sintática da língua falada por falantes da classe média e de alto nível de escolaridade.
- 2 - Construir testes psicolinguísticos para verificar correlações entre desempenho linguístico e as medidas de complexidade acima descritas. Por exemplo, testes de memorização de sintagmas de igual extensão mas diferentes densidades.
- 3 - Verificar as gradações de complexidade sintática na fala de crianças de diversas idades.
- 4 - Verificar se existem correlações entre os níveis de complexidade sintática atingidas pelos falantes e os níveis de desempenho em outros tipos de medidas da capacidade mental: capacidade de memória em geral, Q. I., níveis piagetianos de operações mentais, rendimento escolar, etc.

- 5 - Verificar se existe alguma correlação entre os níveis de complexidade sintática atingidos pelos falantes e a sua aptidão para a leitura.
- 6 - Procurar mais evidências lingüísticas que validem empiricamente o método de mensuração proposto.

A B S T R A C T

Aiming at the improvement of teaching materials, the variety of Portuguese employed by Mobral students (illiterate people) from Rio de Janeiro was investigated in two ways:

The subject-verb agreement rule, variable in this dialect, was investigated with the intent of determining the conditioning factors for its application. The morphological, syntactic and semantic factors hypothesized as influencing the rule's application were found to have their effects hierarchically ordered along a scale of saliency.

Syntactic complexity was compared in samples from the following kinds of texts: Mobralenses' spoken language, comics, literature, newspapers. The basis for the comparison was a method of evaluation of complexity based on Yngve's depth hypothesis. An inverse relation was found between degree of embedding and constituent complexity. The hierarchy of complexity found in the corpora shows that the speakers can attain degrees of sentence complexity even higher than the mean values found for newspapers, although the mean values for their noun-phrase, verb-phrase and embedded sentence constituents is lower than the newspapers' mean values for those units.

R E F E R Ê N C I A S

1. Barbaud , Philippe. 1976. "Constructions Superlatives et Structures Apparentes".
Linguistic Analysis, 2, 2, pp. 125-174.
2. Blalock, H. M. 1960/72. Social Statistics. New York,
Mc Graw-Hill.
3. Cedergren, Henrietta e David Sankoff. 1974. "Variable rules:
performance as a statistical reflection of competence."
Language 50, pp. 335-355.
4. Chen, Matthew Y. e William S-Y. Wang. 1975. "Sound change:
actuation and implementation". Language 51, pp. 255-281.
5. Chomsky, Noam. 1970. "Remarks on Nominalizations". Readings in
English Transformational Grammar, editado por R. Jacobs e P.
Rosembaum. Waltham, Massachussetts: Ginn. Pp. 184-221.
6. Chomsky, Noam. 1976. "Conditions on Rules of Grammar".Linguistic
Analysis 2, 4.
7. Fiengo, Robert Wilson. 1974. Semantic Conditions on Surface
Structure. Tese doutoral, Massachussetts Institute of
Technology, Cambridge, Massachussetts.
8. Fodor, J.A., T.G. Bever e M. F. Garrett. 1974. The Psychology
of Language. New York: Mc Graw-Hill.
9. Freidin, Robert. 1976. "The analysis of passives". Language
51, 1, pp. 384-405.
10. Jackendoff, Ray. 1972. Semantic Interpretation in Generative
Grammar. Cambridge: MIT Press.

11. Kimball, John. 1973. "Seven Principles of Surface Structure Parsing in Natural Language". Cognition 2, pp. 15-48.
12. Labov, William. 1972. "Contraction, deletion and inherent variability of the English Copula". Language in the inner city, pp. 65-129. Philadelphia: University of Pennsylvania Press.
13. Lemle, Miriam. 1975. "Universais lingüísticos: uma reinterpretação". Comunicação apresentada ao Sétimo Congresso Brasileiro de Língua e Literatura da Sociedade Brasileira de Língua e Literatura, Rio de Janeiro, 14 a 18 de julho de 1975. Rio de Janeiro: Novacultura, pp. 53-70.
14. Lemle, Miriam e Anthony J. Naro. 1976. "Hierarquias de Complexidade e Regras Gramaticais". Comunicação apresentada ao Primeiro Congresso Nacional de Lingüística e Literatura da Sociedade Universitária Augusto Motta, Rio de Janeiro, janeiro de 1976.
15. Naro, Anthony J. 1971. "Syntactic change as a surface phenomenon". Comunicação apresentada no Winter Meeting da LSA. St. Louis, Missouri. Tradução portuguesa em Littera 16 (1976), pp.125-131.
16. Naro, Anthony J. 1975. "The SWAMINC program". Mimeografado, Departamento de Letras, PUC/RJ.
17. Naro, Anthony J. 1976. "The genesis of the reflexive in Portuguese: a study in syntactic change as a surface phenomenon". Language 52, pp. 779-810.
18. Ross, John R. 1967. "Constraints on variables in syntax". Massachusetts Institute of Technology. Tese doutoral. Mimeografado, Indiana University Linguistics Club.

19. Sankoff, David. 1975. VARBRUL2. Mimeografado, Université de Montréal.
20. Yngve, Victor. 1960. "A Model and an Hypothesis for Language Structure". Proceedings of the American Philosophical Society 104, pp. 446-466.