

## **AS CATEGORIAS DE COESÃO TEXTUAL E A COMPLEXIDADE DO TEXTO**

**Marisa Porto do Amaral**  
Mestre em Letras pela PUCRS

### **1 — INTRODUÇÃO**

Em leitura, a complexidade lingüística, principalmente no que se refere ao léxico e à sintaxe, vem sendo apontada como um dos fatores responsáveis pela dificuldade de compreensão de um texto.

No entanto, a complexidade não pode ser medida apenas em termos de léxico e de sintaxe, uma vez que o texto constitui uma macroestrutura e apresenta propriedades diferentes das propriedades individuais de cada frase. Assim, a complexidade macroestrutural do texto aponta para novos rumos de pesquisa (Poersch, 1981:165-6).

Lingüistas do texto constataram a necessidade de se estudar, tanto na estrutura profunda lógico-semântica quanto na estrutura superficial, os fenômenos referentes à coerência e à coesão textuais, responsáveis pelas relações mais globais do texto e pelas relações que determinam as condições de encaideamento entre as frases.

Como tais fenômenos têm sido alvo de investigação na produção de textos, mais especificamente nas redações escolares\*, resolveu-se, neste trabalho, investigar alguns desses fenômenos no processo de recepção de textos.

O objetivo da pesquisa foi verificar, num nível transfrasal, quais entre os índices ou categorias responsáveis pelas relações intra e interfrasais, aqueles que apresentam maior dificuldade na compreensão de leitura de textos.

---

\* Veja-se: MUTTI, Regina M. V. Investigação sobre a coerência de textos dissertativos: análise transfrasal. *Letras de Hoje*, Porto Alegre, PUCRS, 14 (44): 40-57, jun. 1981; e SILVEIRA, Maria Izabel da. Análise da coerência em redações escolares. *Letras de Hoje*, Porto Alegre, PUCRS, 15 (48): 87-106, jun. 1982.

Goodman (1976a: 490) afirma que "essencialmente o único objetivo em leitura é a compreensão. Tudo o mais são habilidades usadas para a realização dessa compreensão."

No entanto, como salientam Wiener & Cromer (apud Ross, 1979: 222-3), para que a compreensão possa ocorrer, o leitor precisa decifrar a mensagem escrita, e isto, por sua vez, pressupõe que os símbolos dos quais a mensagem é composta sejam discriminados. Logo, receber o estímulo visual escrito, decifrá-lo em seu equivalente verbal e compreender o significado da mensagem contida na escrita são habilidades relacionadas e decisivas para a leitura.

No processo ativo de leitura, em que o leitor interage com o texto, o significado é o seu alvo constante. Ele testa continuamente suas escolhas e constrói hipóteses para ver se o que lê faz sentido. Para Goodman (1976a: 483), o processo não requer que o leitor perceba e identifique todas as pistas, mas que verifique suas hipóteses, de modo que possa reconhecer seus erros e juntar outras pistas, voltando ao texto todas as vezes que se fizer necessário.

Assim, Goodman (1976b: 498) vê a leitura como "um jogo psicolinguístico de adivinhação". E a eficiência dessa leitura não resulta da exata percepção e identificação de todos os elementos, mas da habilidade em adivinhar as palavras, selecionando as pistas dadas pelo texto.

Reconhece-se, entretanto, que o estudo de tal habilidade é muito complexo e constitui uma área difícil para a pesquisa, já que não é possível a observação direta dos processos mentais por meio dos quais uma pessoa extrai o significado de informações. Para se inferir que houve compreensão, é necessário que se verifique o comportamento manifestado pelo leitor.

Para um bom desempenho, ele deverá valer-se de estratégias e habilidades tais como: concentrar-se no contexto adjacente ou imediato, fazer predições quanto às relações gramaticais entre os itens lexicais (grupos de sentido), quanto aos mecanismos de estruturação frasal (ordem das palavras), ou valer-se do caráter redundante da língua, pois devido a essa redundância é possível compreender frases inteiras pela identificação de suas partes (Lado, 1964: 131). Com base no seu conhecimento anterior, na sua perspicácia, nas suas experiências, ele poderá fazer suas escolhas provisórias, testá-las no texto para confirmá-las, rejeitá-las ou refiná-las, de forma a conseguir reconstruir a idéia do autor.

Antes da lingüística do texto, as investigações das teorias não ultrapassavam os limites da frase. Mas, a partir do pressuposto de que a comunicação lingüística se realiza por meio de textos e não de frases sucessivas, o objeto de investigação passa ser o texto. Este é visto como um todo semanticamente organizado.

Uma vez que o texto é mais do que uma simples seqüência de enunciados, tanto sua compreensão quanto sua formação derivam de uma competência específica do falante, a **competência textual**, que faz com que cada falante de uma língua produza e compreenda enunciados gramaticalmente corretos. Ele é capaz de parafrasear um texto, de resumir-lo, de descobrir a complexidade do seu significado, de ver se ele está completo ou incompleto, de dar-lhe um título ou de produzi-lo a partir de um título dado.

Brown & Yule (1984: 190) definem o texto como "o registro verbal de uma ocorrência comunicativa", e destacam o papel relacional da coesão.

Para que um texto seja considerado bem-formado, é necessário que exista uma relação lógica entre os elementos, nas frases e entre frases, nos parágrafos e entre parágrafos, formando um todo significativo e coerente.

Crê-se que, para ser coerente, o texto precisa ser também coesivo. Pois, se, como diz Hasan (1984: 181), é através da coesão que se estabelecem as bases da coerência, pode-se deduzir que apenas uma destas propriedades do texto não é suficiente para estabelecer a textualidade ou textura.

Halliday & Hasan (1976: 4) propõem a propriedade de coesão como determinante parcial de textura, conceito que expressa a propriedade de "ser um texto". Para eles, o conceito de coesão é semântico; refere-se às relações de sentido que existem dentro do texto e que o definem como tal. As relações coesivas ocorrem quando a interpretação de algum elemento no discurso depende da interpretação de outro. Um pressupõe o outro, no sentido de que não pode ser decodificado sem recorrer ao outro. Quando isto acontece, a relação de coesão é estabelecida e os dois elementos estão integrados na estruturação do texto.

No sentido de Beaugrande & Dressler (1981: 3), coesão refere-se aos meios pelos quais os componentes da estrutura

superficial do texto, isto é, as palavras que ouvimos ou lemos, estão mutuamente relacionadas na seqüência.

O conhecimento das regras sintáticas de concordância nominal, de concordância verbal, de lógica do sistema verbo-temporal, de conjunções ou advérbios que fazem prever determinados elementos a seguir, bem como a apreensão de elementos textuais — articulações lógicas e retóricas do discurso, palavras-chave, redundâncias, anafóricos e outros — colaboram na construção do sentido do texto (Neis, 1982: 45-6).

Os elementos tratados na presente pesquisa podem ser facilmente recuperados no contexto frasal, pois muitos são repetidores ou substituidores. Além desses, outros componentes, também responsáveis pelas relações intra e interfrasais, foram trabalhados.

#### 4 — AS CATEGORIAS TEXTUAIS

Existe uma série de fenômenos não explicáveis numa gramática limitada só ao domínio da frase, mas que, referindo-se à estrutura textual, podem ser analisados em dois níveis de organização do texto; um, macroestrutural, e outro, microestrutural. O primeiro refere-se à estrutura profunda lógico-semântica do texto e dá conta das relações mais globais, caracterizando o texto inteiro. O segundo refere-se à estrutura superficial e diz respeito às condições de encadeamento entre as frases.

Segundo Van Dijk (apud Schmidt, 1978: 176), é apenas a hipótese da estrutura profunda do texto que pode esclarecer o porquê da possibilidade de existirem relações entre todas as frases de um texto, relações estas que ultrapassam as restrições microestruturais. A conexão das frases para a formação de textos e o encadeamento dos constituintes textuais pressupõem um princípio semântico mais amplo do que a frase, situando-se ao nível do texto global, ou seja, ao nível da macroestrutura. Mas, para que um texto seja macroestruturalmente coerente, é preciso que haja coesão no seu aspecto microestrutural. É este segundo aspecto que interessa no presente estudo.

As relações de encadeamento nas frases e entre frases consecutivas de um texto pertencem ao nível microestrutural. Nele, dá-se conta de certos fatores que tratam dos aspectos co-textuais, ou seja, é permitido constatar que tais relações determinam a estrutura superficial do texto.

Desses elementos, abordar-se-ão a pronominalização — através de relações endofóricas — a definitivização, a substituição lexical, a nominalização e os articuladores.

##### 4.1 — Pronominalização

Irena Bellert (1981: 150) mostra que uma condição necessária para a coerência de um texto consiste na repetição. E uma das formas de se fazer tal repetição é através do uso de pronomes que referem, a uma certa distância, um sintagma, uma frase ou uma seqüência de frases.

Dentro da referência pronominal, foram tratadas as relações endofóricas, isto é, as que formam relações coesivas dentro do texto. Assim, quando se volta atrás no texto para interpretar determinada entidade, a relação é anafórica; quando se segue adiante, a relação é catafórica.

Os pronomes anafóricos formam a categoria que, para Harweg (1979), ocupa um papel decisivo na constituição do texto.

Ocorre **anáfora**, quando o item em questão refere algo que o precede no texto. Exemplos:

"**João** chegou tarde em casa e **ele** estava muito cansado".

"...**era tratado como gente** e se orgulhava **disso**."

nos quais "ele" e "disso" referem seus antecedentes "João" e "era tratado como gente", respectivamente.

Ocorre **catáfora**, quando o pronome antecipa seu referente.

Exemplo:

"Só penso **nisso**: **que brevemente o verei**".  
onde o pronome "nisso" refere, por antecipação, a frase "que brevemente o verei".

##### 4.2 — Definitivização

A definitivização, ou passagem de um artigo indefinido para o definido, está sujeita a certas regras, como:

— O referente indefinido só pode ser retomado por um referente definido. Exemplo:

"**Um** cidadão em Santa Teresa (...). Solteiro, **o** cidadão tinha(...).

— O referente definido também só pode ser retomado por um referente igualmente definido. Exemplo:

"(...) a casa que o Zanine(...). Meses depois, ao visitar o amigo, Pio já era quase um galo."

Afirma Weinrich (1981: 59-60) que a distribuição dos artigos num texto é um aspecto importante de sua estrutura signíca e textual, uma vez que são sinais que guiam o leitor ou o ouvinte a compreender os demais signos do texto e a sua interdependência.

O **definido** guia o leitor/ouvinte para a informação anterior (direção anafórica), enquanto o **indefinido** leva à informação posterior (cataforicamente), fazendo com que o decodificador espere novos determinantes no texto subsequente para compreender perfeitamente o grupo nominal ou **articulado** (como ele chama) introduzido por esse indefinido.

#### 4.3 — Substituição lexical

A substituição consiste na utilização de um item no lugar de outro(s) ou até de uma oração inteira. É uma relação dentro do texto. O princípio que a distingue da referência está claro em Halliday & Hasan (1976: 89): "Substituição é uma relação entre itens lingüísticos, tais como palavras ou frases; enquanto referência é uma relação entre significados". Esta pertence ao nível semântico, ao passo que aquela pertence ao nível léxico-gramatical, o nível da gramática e do vocabulário.

É regra geral que o item substitutivo deve ter a mesma função estrutural que o seu co-referente. Exemplos:

"(...) atirou a **bola** cá fora (...), ajeitou cuidadosamente a **pelota**, (...) e vai levando a **bichinha** no ar (...).

"Quando o padre **ajoelhou** (...), (...) **fizeram o mesmo**."

#### 4.4 — Nominalização

A nominalização é um substituto verbal que se refere a um verbo do co-texto ou do contexto. Primeiro aparece o verbo, depois sua forma nominalizada com pré-determinantes de diversos tipos. Exemplo:

"(...) **declarou** (...) . (...) esta nova **declaração** (...)"

O exemplo dado mostra que a nominalização de um verbo

do contexto tem uma função de ligação entre as frases de um mesmo parágrafo e entre os parágrafos de um mesmo texto. Parece que há no receptor-leitor memorização do sentido de contexto onde o verbo aparece, uma vez que lhe permite decodificar a significação desse mesmo verbo, através de um paralelo semântico-formal entre as duas proposições (Moirand, 1975: 65).

Acontece, às vezes, que a unidade nominalizadora, embora apareça diferente de todos os verbos do enunciado, repete o mesmo "objeto de pensamento" que uma outra proposição do texto. Isto é, uma mesma "noção" é expressa primeiramente por uma unidade verbal e repetida mais adiante, no discurso, por um substituto verbal, formalmente diferente do verbo. Exemplo:

"O Gato Malhado refletiu e **compreendeu** então **que fugiam** dele (...). Foi uma triste **constação**."

Até aqui, viram-se os elementos de repetição e de substituição, responsáveis pelas relações coesivas no co-texto, podendo todos, com exceção da catáfora, ser recuperados anaforicamente. Os outros fatores trabalhados foram os articuladores.

#### 4.5 — Articuladores

Os articuladores diferem um tanto das relações vistas até aqui, por não se tratar simplesmente de uma relação anafórica. Eles são coesivos não por si próprios, mas indiretamente, em virtude das relações significativas específicas que se estabelecem entre as orações dentro do período, entre os períodos dentro de um parágrafo, entre os parágrafos no interior do texto. Essas relações são as responsáveis pela coesão e coerência textual.

Os principais tipos de elementos conjuntivos são: advérbios e locuções adverbiais; conjunções coordenativas e subordinativas, locuções conjuntivas; preposições e locuções prepositivas; itens continuativos como **daí**, **então**, **a seguir**, etc.; além de outras palavras que, segundo a Nomenclatura Gramatical Brasileira, recebem classificação à parte das outras classes, como: **até**, **mesmo**, **também**, **somente**, **aliás**, **ou melhor**, etc. Exemplos:

"Não havia prestado atenção nela e ela estava ali (...)"  
"Li os jornais de hoje; **allás**, leio os jornais todos os dias".

Para atingir o objetivo de determinar as categorias textuais que apresentavam maior dificuldade para os alunos na compreensão de leitura de textos, foi realizada a pesquisa, aferida pelo "procedimento cloze", uma técnica que, a partir de seus primeiros experimentos, tem-se demonstrado válida e fidedigna para medir a legibilidade de textos.

Ao construir um teste cloze, suprime-se sistematicamente palavras de um texto e avalia-se o sucesso do leitor na tentativa de repor as palavras suprimidas.

O "corpus" trabalhado constituiu-se de doze textos extraídos de livros didáticos e de leitura extraclasse para 1º e 2º graus, aplicados a 84 alunos, provenientes de três turmas de 8ª série da Escola Estadual Bibiano de Almeida, em Rio Grande, no período de agosto a outubro de 1983.

De cada texto organizou-se uma amostra de 250 palavras no mínimo e 300 no máximo, sendo que as amostras foram desdobradas em cinco versões, cada uma com cinquenta lacunas (cujo tamanho-padrão foi de 15 espaços datilografados), totalizando sessenta textos-amostra e três mil itens apagados.

Optou-se pelo tipo de cloze mais usado, ou seja, o que suprime a quinta palavra da amostra na versão-base, bem como a quarta, a terceira, a segunda e a primeira palavra das demais versões, fazendo com que, considerando-se o conjunto das cinco versões, todas as palavras fossem "apagadas". O tempo máximo estabelecido para cada teste foi o de uma hora-aula, correspondente a cinquenta minutos.

Como esta pesquisa segue o modelo de Goodman (1976a e 1976b), esperava-se dos alunos um procedimento semelhante ao exposto anteriormente; ou seja, na tentativa de adivinhar as palavras suprimidas no texto, eles deveriam testar suas hipóteses e fazer suas escolhas, levando em conta o contexto adjacente.

Na contagem dos escores, preferiu-se o critério da aceitação de sinônimo ou palavra apropriada ao contexto ("contextually appropriate scoring method"), já que o uso de equivalentes demonstra também que o aluno havia compreendido o texto.

Concluída a aplicação dos testes, procedeu-se à tabulação geral dos resultados referentes aos dados brutos, num total de 50.400 itens, distribuídos em 1.008 testes relativos aos doze textos com 50 lacunas respondidos pelos 84 alunos.

Primeiramente, efetuou-se o fichamento de cada uma das lacunas de cada versão em 60 quadros, correspondentes aos doze textos com as cinco versões.

O resultado dos dados brutos está resumido na tabela abaixo:

#### DISTRIBUIÇÃO DO CLOZE NO TOTAL DE TEXTOS

Cloze	Total
itens lacunados	50.400
respostas adequadas	23.404
erros	26.996

Num segundo momento, procedeu-se à tabulação dos dados que realmente interessavam à pesquisa: as lacunas referentes às categorias textuais. A visão global desses dados encontra-se na tabela a seguir:

#### DISTRIBUIÇÃO DOS ESCORES DE ACERTOS POR CATEGORIA NO TOTAL DOS TEXTOS

Categoria textual	Itens lacunados	Acertos	%
anáfora	3.300	1.530	46,36
catáfora	83	41	49,39
definitivização	924	623	67,42
substituição lexical	420	141	33,57
nominalização	133	33	24,81
articuladores	3.734	1.665	44,59
Total	8.594	4.033	46,92

Para verificar a hipótese geral de que "entre as diversas categorias textuais em estudo existem algumas que influem mais significativamente na legibilidade dos textos de leitura", proce-

deu-se ao tratamento estatístico, achando a frequência esperada dos acertos e erros por categoria. Para comparar as categorias, utilizou-se o teste Qui Quadrado ( $X^2$ ), cujo valor encontrado (219,36) permitiu rejeitar a hipótese nula em nível de significância inferior a 0,01. Este resultado permitiu a comprovação da hipótese geral.

Consultando-se a tabela acima, pode-se verificar que a categoria das nominalizações foi a que obteve menor grau de acertos: 24,81%, seguida da categoria das substituições lexicais: 33,57%, o que comprova parte da hipótese. Esta não pôde ser totalmente confirmada, porque o processo catafórico não está entre as categorias de maior dificuldade, já que obteve o segundo maior grau de acertos: 49,39%. Assim, a hipótese de que "exercem maior influência na dificuldade de um texto as nominalizações, as substituições lexicais e o processo catafórico" foi parcialmente confirmada.

A definitivização obteve um expressivo percentual de acertos em relação às outras categorias: 67,42%, o que significa 623 respostas para 924 itens de **cloze**. Com isso, confirmou-se a hipótese de que "a definitivização é a categoria que exerce menor influência na dificuldade de um texto".

Analisados os resultados quanto às categorias textuais referentes à coesão, pôde-se fazer algumas constatações:

A pronominalização foi a categoria que, depois da definitivização, apresentou mais acertos. Seu emprego torna-se um recurso variado para assegurar a repetição. Entretanto, convém não exagerar no seu uso em um texto, a fim de não comprometer a quantidade de informação, pois é preciso atenção e memória para poder reaver o termo referido no momento apropriado. A distância entre o referente e o referido tem alguma significação, já que a anáfora relacional imediata (com pronome relativo) obteve mais acertos — 62,13% do que a anáfora bem mais distante, causadora de uma maior dificuldade nos acertos.

A definitivização foi a categoria que obteve um percentual mais alto de acertos. Como o artigo definido indica que o item em questão é específico e identificável, como, em algum lugar, a informação necessária para identificá-lo é recuperável, e como ele refere principalmente o que vem antes, foi de mais fácil reconhecimento para os alunos.

Quanto às substituições lexicais, viu-se que elas não são muito usadas, já que a tendência é repetir simplesmente o item em questão.

As nominalizações quase não aparecem nos textos completados pelos alunos. Seu emprego é muito difícil, já que requer do usuário da língua a capacidade de condensar uma determinada idéia numa única palavra, principalmente quando, na frase matriz, o verbo a ser nominalizado não se faz presente, mas fica subentendido pelo contexto.

Os articuladores foram outros recursos coesivos que apresentaram problemas de compreensão. O uso indevido de conectivos desvirtuou o sentido do texto, em alguns casos.

A falta de coesão ocorreu pela não observância de certos itens que pressupõem a presença de outros componentes no texto; ou ainda, pela impossibilidade de se perceber que é necessária uma adequação semântica entre termos, expressões e partes desse todo.

## 6 — CONCLUSÕES

As constatações feitas pelos dados obtidos e analisados na pesquisa parecem indicar que esses elementos, responsáveis pela relações internas do texto, podem ser considerados índices de complexidade.

Quanto ao procedimento **cloze**, verificou-se que não é aconselhável para qualquer categoria textual, já que umas são mais previsíveis do que as outras. Categorias como a nominalização e a substituição lexical, por sua baixa previsibilidade, dificultam a "descoberta" pelos alunos e não devem ser testadas por este método.

Apesar da limitação do trabalho — pois os resultados obtidos são uma amostra do desempenho de um grupo representativo de alunos concluintes do 1º grau — novas pesquisas poderão ser realizadas, como: medir o texto mais fácil e o mais difícil, através de categorias textuais; ou verificar se a diferença das médias de acertos por versão é significativa.

Com a realização desta pesquisa, além da descoberta de novos índices de complexidade da mensagem, pretendeu-se apresentar subsídios para uma melhor adequação do material de leitura ao nível dos alunos, quer se trate de textos para leitura recreativa, quer de textos para treinamentos ou de textos de informações. Aconselha-se, em exercícios de leitura, um treinamento mais específico com as categorias textuais que maior dificuldade oferecem. A atenção a esses fatores pode facilitar a legibilidade dos textos e, conseqüentemente, melhorar o nível dos alunos em relação à leitura.

REFERENCIAS BIBLIOGRÁFICAS

- 1 — BELLERT, Irena. Una condizione della coerenza dei testi. In: CONTE, M. E. (org.). *La linguistica testuale*. 2. ed. Milano, Feltrinelli, 1981. p. 148-80.
- 2 — BEAUGRANDE, Robert de & DRESSLER, Wolfgang. *Introduction to text Linguistics*. London, Longman, 1981.
- 3 — BROWN, Gillian & YULE, George. *Discourse analysis*. Cambridge, Cambridge University Press, 1984.
- 4 — GOODMAN, Kenneth. Behind the eye: what happens in reading. In: SINGER, H. & RUDDELL, R. B. (ed.). *Theoretical models and processes of reading*. 2. ed. Newark, International Reading Association, 1976a. p. 470-96.
- 5 — ————. Reading: a psycholinguistic guessing game. In: SINGER, H. & RUDDELL, R. B. (ed.). *Theoretical models and processes of reading*. 2. ed. Newark, International Reading Association, 1976b. p. 497-508.
- 6 — HALLIDAY, M. A. K. & HASAN, R. *Cohesion in English*. London, Longman, 1978.
- 7 — HARWEG, Roland. Substitutional text linguistics. In: DRESSLER W. V. (ed.). *Current trends in textlinguistics*. New York, Walter de Gruyter, 1979. p. 247-60.
- 8 — HASAN, R. Coherence and cohesive harmony. In: FLOOD, James (ed.). *Understanding reading comprehension: cognition, language and the structure of prose*. Newark, International Reading Association, 1984. p. 181-219.
- 9 — LADO, R. *Language teaching: a scientific approach*. s. l., McGraw-Hill, 1964. p. 131-42.
- 10 — MOIRAND, Sophie. Le rôle anaphorique de la nominalisation dans la prose écrite. *Langue Française*, Paris, Larousse (28): 60-78, déc. 1975.
- 11 — NEIS, Ignacio A. A competência de leitura. *Letras de Hoje*. Porto Alegre, Pontifícia Universidade Católica do Rio Grande do Sul, 15 (48): 43-57, jun. 1982.
- 12 — POERSCH, José Marcelino. Interdependência dos membros da relação maturidade lingüística X complexidade lingüística. *Letras de Hoje*. Porto Alegre, Pontifícia Universidade Católica do Rio Grande do Sul, 14 (44): 160-87, jun. 1981.
- 13 — ROSS, Alan O. *Aspectos psicológicos dos distúrbios da aprendizagem e dificuldades na leitura*. São Paulo, McGraw-Hill do Brasil, 1979.
- 14 — SCHMIDT, Siegfried J. *Lingüística e teoria de texto*. São Paulo, Pioneira, 1978.
- 15 — WEINRICH, Harald. Sintassi testuale dell'articolo francese. In: CONTE, M. E. *La linguistica testuale*. 2. ed. Milano, Feltrinelli, 1981. p. 53-65.