

## 2 Gêneros textuais: pressupostos teóricos

Escrever é uma tarefa árdua e complexa. Toda vez que nos colocamos à frente de uma folha em branco para redigir um texto, mobilizamos uma série de estratégias e competências sem sequer nos darmos conta disso.

Identificar e descrever habilidades de escrita, um papel frequentemente atribuído a estudiosos de Linguística Textual, pode contribuir muito para o estudo e o ensino de produção de textos. Isso ocorre na medida em que o aluno é levado a compreender como determinado gênero textual se estrutura e a lançar mão de uma série de competências linguísticas envolvidas na produção daquele gênero. Para ilustrar: um alfaiate, ao compor uma roupa, precisa dominar técnicas de costura – pregar botões, cortar tecidos, tomar medidas etc. Da mesma forma, quem escreve precisa conhecer e saber aplicar, ainda que inconscientemente, determinadas técnicas discursivas para redigir bem um texto.

Ainda dentro dessa analogia, podemos afirmar que o conhecimento exato de cada tipo de roupa – suas características, o material mais usado, o público a que se destina etc – é o que habilita o alfaiate a produzir, com sucesso, uma variedade de artigos para atender a finalidades distintas. Do mesmo modo, o conhecimento sobre gêneros textuais tem se mostrado fundamental para o domínio de técnicas de produção textual, na medida em que somos levados a refletir sobre uma série de características comuns aos textos que configuram um determinado gênero. Essa aplicação didática do estudo de gêneros discursivos é reconhecida por Lopes-Rossi (2008, p. 62):

Um dos méritos do trabalho pedagógico com gêneros discursivos, de acordo com os pesquisadores do Grupo de Genebra, é o fato de proporcionar o desenvolvimento da autonomia do aluno no processo de leitura e produção textual como uma consequência do domínio do funcionamento da linguagem em situações de comunicação, uma vez que é por meio dos gêneros discursivos que as práticas de linguagem incorporam-se nas atividades dos alunos.

A postura pedagógica do uso de gêneros textuais em sala de aula esboçada acima encontra forte ressonância na definição de gêneros do discurso concebida por Bakhtin (2006, p. 262), definição essa que adotamos neste trabalho, a saber, “tipos relativamente estáveis de enunciados”, que resguardam entre si algumas semelhanças e cumprem determinada função social na esfera das relações humanas. Dentre as principais semelhanças partilhadas por textos relativos a um mesmo gênero, poderíamos citar como exemplo:

i) a *distribuição de conteúdo linguístico no papel* – normalmente conseguimos diferenciar uma notícia de jornal de uma bula de remédio simplesmente por observar como as palavras, ou mancha gráfica, estão organizadas no espaço que ocupam;

ii) a *estruturação composicional* – artigos de opinião geralmente obedecem a uma organização dissertativo-argumentativa, do tipo “apresentação da tese / desenvolvimento de argumentos e contra-argumentos / conclusão”;

iii) o *estilo* – piadas tendem a se caracterizar por linguagem coloquial e um ar de informalidade, ao passo que leis e decretos primam por maior apuro formal;

iv) as *escolhas linguísticas* – processos jurídicos costumam empregar grande quantidade de termos técnicos, nominalizações, inversões sintáticas etc; em contraste, relatórios de experimentos se caracterizam por uma linguagem predominantemente descritiva, com muitos verbos de ação e enumerações;

v) o *público-alvo* – receitas de bolo são, em geral, voltadas para pessoas interessadas em cozinhar, assim como trabalhos acadêmicos têm por público-alvo pesquisadores da área a que pertencem.

Sabemos que muitos alunos têm dificuldades de lidar com textos técnicos, como resenha, resumo e relatório, ao ingressarem na universidade. Grande parte dessa dificuldade parece vir do desconhecimento acerca de padrões de estruturação, uso da língua e formatação dos textos acadêmicos, o que está intimamente relacionado ao estudo de gêneros. Esse problema ressalta a necessidade premente de uma boa formação do professor quanto ao reconhecimento, caracterização e reflexão acerca dos gêneros textuais, de maneira a suprir as dificuldades do aluno. O trabalho com gêneros em sala de aula, portanto, permitiria ao aluno uma reflexão autônoma sobre o que está envolvido na produção de uma multiplicidade de textos cotidianos, desde uma notícia de jornal a um relatório científico. Este, por sua vez, seria levado a construir suas próprias percepções dos traços comuns a determinado gênero, partindo de um modelo básico que, aos poucos, seria aperfeiçoado com base na experiência e progresso do aluno em produção textual.

Considerando esse contexto, apresentamos a seguir uma reflexão pautada no gênero textual conhecido como resumo escolar / acadêmico. Fornecemos uma descrição dos traços linguísticos, discursivos e contextuais comuns a textos resumitivos para, em seguida, apresentar o que está envolvido no processo básico de redução da informação lexical, conhecido como sumarização, levando em conta habilidades de leitura e escrita e conceitos como retextualização e macroestruturas. Por fim, encerramos o capítulo introduzindo uma discussão sobre o papel da leitura na produção de resumos acadêmicos.

## 2.1

### O resumo acadêmico: características fundamentais

Em seu artigo “Revisitando o conceito de resumos”, Anna Rachel Machado (2007) faz um estudo pormenorizado do resumo enquanto gênero textual, descrevendo algumas características envolvidas na produção desse construto. A autora busca recuperar algumas posições em Linguística Textual de autores como Dolz & Schneuwly (1998) e Bronckart (1997) dando a elas uma nova perspectiva com base numa reflexão bakhtiniana.

Uma primeira questão abordada por Machado tem a ver com a confusão terminológica criada em torno da palavra “resumo” enquanto gênero textual. Conforme a autora mostra, as pessoas costumam classificar sob esse nome uma série de textos diferentes. Chamam de “resumo” textos técnicos com uma avaliação crítica (resenhas), apresentações resumidas de filmes e livros (sinopses), blocos de texto introdutórios numa pesquisa acadêmica (os *abstracts*) e até mesmo o processo de produção do resumo em si (a sumarização). Isso ocasiona dois problemas para o ensino e reconhecimento de gêneros. O primeiro é que a nomeação nem sempre é sistemática e homogênea: um mesmo gênero pode ser encontrado sob nomes diferentes em veículos comunicativos diferentes, como é o caso da sinopse que, em alguns periódicos, aparece com o nome de “resumo”. O segundo é o obscurecimento da relação entre o gênero e os textos que o materializam, como é o caso do resumo enquanto síntese de um texto acadêmico e o resumo enquanto apresentação de um trabalho científico, ambos identificados por muitos como sendo relativos ao mesmo gênero. Entretanto, conforme aponta Machado (2007, p. 148), o resumo introdutório de monografias, dissertações e teses já tem caminhado na direção de uma independência; portanto, deve ser reconhecido como um gênero à parte, nomeado por alguns de *abstract*.

O *abstract*, por sua vez, vem recebendo muita atenção nos últimos anos, sendo considerado uma espécie de subgênero do resumo, comumente encontrado em monografias, trabalhos de conclusão de curso, dissertações de mestrado e teses de doutorado. Sua função comunicativa básica costuma ser fornecer uma descrição sucinta do conteúdo do trabalho, poupando tempo ao leitor e ajudando-o a determinar o grau de interesse que aquela pesquisa assume para ele em si. Os trabalhos de Samraj (2002), Yakhontova (2002), Zanella & Heberle (2005), Araújo (2009) e Motta-Roth & Hendges (2009) demonstram que a composição de um *abstract* varia de acordo com fatores de ordem cultural, linguística e disciplinar. Esses fatores determinam a predileção do pesquisador de determinada área por certas escolhas linguístico-discursivas, numa quantidade e posicionamento específicos no texto. Essa percepção é reforçada mediante um estudo desse construto textual sob a ótica do modelo CARS (acrônimo de “*Create A Research Space*”) proposto por Swales (1989, e mais tarde reelaborado por Swales em 1990), cujo princípio básico é o estudo de gêneros através da identificação dos chamados movimentos retóricos – ações tomadas pelo escritor durante a produção textual, traduzidas em escolhas linguístico-discursivas específicas, facilmente identificáveis no texto. Parece ser razoável concluir, portanto, que um dos fatores fundamentais envolvidos na produção de um gênero textual é o conhecimento partilhado com o público-alvo e o veículo comunicativo em que o texto se insere.

Vejam agora alguns dos principais traços característicos do resumo escolar / acadêmico mais especificamente. Nunes (2010, p. 13) o descreve como fazendo parte de uma atividade primariamente de reescritura, na qual se resguardam a coerência e coesão em relação ao texto fonte, constituindo-se, porém, uma nova unidade que pode ser lida independentemente. Abasse (2008, p. 34) corrobora essa noção, ao afirmar que um resumo tem por função “apresentar de forma concisa as ideias básicas de um texto como forma de avaliação de aprendizagem no conteúdo” e é composto por sequências narrativas que preservam, ou refletem, estruturas linguísticas do texto original, mantendo seu modelo global relativamente intacto. Assim, por se tratar de uma simples apresentação sintática do conteúdo e organização do texto fonte, esse gênero não costuma conter comentários ou julgamentos relativos ao texto resumido, papel que normalmente cabe à resenha. Há que se notar também que o resumo serve como um instrumento valioso tanto para o educador, como para o educando: permite ao professor avaliar até que ponto o aluno compreendeu o

texto sumarizado e auxilia um aluno na hora de estudar determinado conteúdo ou organizar suas pesquisas.

De modo geral, o resumo é considerado satisfatório “se for claro, breve e informacionalmente fiel, se utilizar palavras próprias, bem como se enfatizar a ideia central do assunto” (Nunes, 2010, p. 13). Machado, Lousada & Abreu-Tardelli (2004, p. 17) contribuem para solidificar essa percepção ao estabelecer alguns parâmetros para avaliação de um bom resumo. Entre os elementos que, na visão das autoras, devem fazer parte do novo texto produzido, podemos destacar:

- correção gramatical e léxico adequado à situação escolar / acadêmica;
- seleção das informações colocadas como as mais importantes no texto original;
- indicação de dados sobre o texto resumido, no mínimo autor e título;
- organização de tal modo que o leitor (nesse caso mais específico, o professor) possa avaliar a compreensão do texto lido, incluindo a compreensão global, o desenvolvimento das ideias do texto e a articulação entre elas;
- apresentação das ideias principais do texto original e de suas relações;
- possibilidade de compreensão do resumo final por si mesmo.

Tendo em mente o padrão consensual descrito acima, relativo às principais características de um resumo, precisamos analisar agora o complexo processo envolvido na elaboração desse gênero textual. Examinemos, portanto, o que está por trás da atividade conhecida como redução da informação semântica, ou sumarização.

## **2.2**

### **A atividade de sumarização**

A sumarização é uma competência textual imprescindível à vida acadêmica. Na elaboração de resumos dos mais diversos tipos estão envolvidas várias habilidades básicas de escrita que deverão ser assimiladas por alunos de graduação para qualquer carreira que decidam escolher. Além disso, um bom domínio de técnicas inerentes à redução de informação semântica, como separação entre dados relevantes e acessórios, bem como estratégias de reformulação de enunciados, constituem a base para o domínio de gêneros acadêmicos de maior complexidade.

A atividade de sumarização está fortemente associada ao conceito de “macroestrutura textual”, conforme introduzido por van Dijk (1977, 2004), que diz respeito a uma “reconstrução teórica de noções intuitivas como a de tópico ou a de tema de um discurso” (2004, p. 51). Em suas pesquisas, van Dijk aponta para um fato interessante, inerente ao processo de compreensão e interpretação de um texto: como os usuários da língua conseguem resumir textos relativamente complexos, responder perguntas sobre estes e memorizá-los sem terem acesso imediato a toda e cada sentença individual que compõem o construto como um todo? A solução proposta pelo teórico para esse questionamento veio com o advento de um modelo de compreensão leitora, cuja origem se deu a partir da observação de estruturas narrativas, que costumam resguardar uma semelhança no que diz respeito a seu esquema global, grosseiramente expresso numa sequência do tipo “situação inicial, conflito, resolução”. Considerando essa semelhança, van Dijk elaborou os conceitos de macro e microestrutura.

A macroestrutura corresponde a um elemento abstrato inerente, que expressa as relações lógico-semânticas de um determinado texto, possibilitando ao leitor determinar o que é mais relevante e atribuir um sentido global ao texto. Em outras palavras, a coerência textual vai sendo estabelecida à medida que o leitor vai aplicando, inconsciente e intuitivamente, uma série de macrorregras semânticas sobre o texto, construindo uma espécie de síntese mental, em que as informações básicas são retidas e as acessórias são eliminadas, chegando-se, ao final desse processo, à significação básica do texto. A microestrutura, por outro lado, diz respeito às relações sintáticas estabelecidas entre os períodos que constituem um texto, bem como à carga semântica de seus elementos. Podemos afirmar, portanto, que a macroestrutura, cuja operação se dá no nível global de um texto, é o resultado da aplicação de uma série de regras macro-semânticas sobre a microestrutura de um texto, a qual atua no nível sentencial (CORACINI, 2009, pp. 25-6).

As macrorregras têm a ver com o processamento cognitivo de um texto, e são responsáveis pela organização e significativa redução de informações do construto original, conforme afirma Nunes (2010, p. 14):

À medida que o leitor avança na leitura do texto-fonte, ele o reduz e organiza sua micro-estrutura, colocando-o dentro de uma macro-estrutura por uma série de transformações conhecidas como macro-regras. A aplicação dessas regras de mapeamento, de forma cíclica e recursiva, permite que o escritor/leitor elabore uma síntese textual coerente e compreensível. Isso ocorre, porque, num processo ascendente, as proposições foram conectadas, organizadas, transformadas e condensadas em idéia principal. E mais, sua máxima adequação semântica fica

atrelada à compreensão constituída, refletida, metalinguística e quase crítica do conteúdo original.

É importante ressaltar que o processo de sumarização está intimamente relacionado ao conceito de ação de linguagem, conforme introduzido por Bronckart (1997 *apud* MACHADO, 2007, p. 140):

Conjunto de representações, interiorizadas pelos agentes, de determinadas representações sociais sobre o mundo físico e sócio-subjetivo, que por eles são mobilizadas diante da necessidade de produzirem um texto oral ou escrito. Um primeiro conjunto de representações se constitui como o contexto da produção e um segundo constituirá o conteúdo temático.

van Dijk (1977) aponta para a existência de três macrorregras, ou estratégias, fundamentais à construção da macroestrutura textual. A primeira delas, conhecida como *apagamento*, ou *deleção*, é de ordem seletiva – informações desnecessárias e acessórias são identificadas e eliminadas na elaboração do resumo mental. Alves (2010, p. 41) elabora essa definição, ao afirmar que resumir é “uma atividade complexa que compreende a elisão daquelas partes que o leitor concebe como sendo informações adicionais, ou até elementos estilísticos, retóricos, colocados no texto pelo autor como complemento à estrutura inicialmente concebida”.

A segunda regra, chamada de *generalização*, consiste na substituição de uma sequência de nomes de seres, propriedades e ações por um termo de ordem mais genérica, ou superior, como um hiperônimo, por exemplo. Assim, uma sequência como “Ontem, no zoológico, vi gatos, leopardos e tigres”, poderia ser substituída por “Ontem, no zoológico, vi felinos”.

A terceira e última regra, *construção*, diz respeito ao enxugamento de uma sequência através da inferência de termos mais relevantes a uma determinada passagem, resultando na formação de uma sequência menor e mais genérica. Em outras palavras, poderíamos encurtar uma sequência formada por “n” proposições substituindo tal sequência por uma única proposição inferível a partir da associação de significados das proposições originais. Isso implica por vezes uma alteração na sintaxe e seleção vocabular. A título de exemplo, se tivermos uma sequência como “Pedro voltou do trabalho, entrou em casa, tomou um banho, fez a barba, vestiu-se, colocou perfume, dirigiu até a igreja, viu a noiva entrar etc.”, poderíamos facilmente substituí-la por uma proposição inferível a partir do contexto, como “Pedro foi a um casamento”.

É importante ressaltar ainda o caráter recorrente dessas estratégias: podemos aplicar a sumarização vez após vez sobre um enunciado complexo, de modo a torná-lo cada vez mais resumido, de acordo com as necessidades do texto a ser produzido. Também vale notar que a aplicação de tais estratégias está “condicionada ao objetivo da leitura, ao conjunto de conhecimentos prévios do leitor, ao tipo de situação em que se processa a leitura; enfim a uma série de fatores contextuais” (MACHADO, 2007, p. 141). Novamente, informações desse tipo estão ligadas ao estudo de gêneros, manifesto na análise dos traços recorrentes em textos representativos do que se convencionou chamar “resumo acadêmico”. Nunes (2010, p. 14), por sua vez, aponta para o fato de as operações cognitivas envolvidas na produção de um resumo mental possuírem uma correspondência textual nos níveis lexical e sintático.

Alves (2010, p. 43) reconhece ainda a alta demanda cognitiva imposta pela atividade de sumarização, traduzida em três etapas significativas que exigem muito da memória de trabalho: um processo de *compreensão do conteúdo do texto-fonte*, seguido de uma *seleção das ideias centrais*, etapa em que o escritor lança mão de estratégias de leitura e, por último, a *organização de um texto inteiramente novo*, que reflita as características do gênero resumo. A autora ainda fornece detalhes sobre o que se encontra envolvido nessas etapas:

Este tem que ler o texto-fonte, avaliar o conteúdo mensurando a importância de cada segmento (ou episódio, conforme van Dijk), sem deixar de prestar atenção aos aspectos linguísticos (vocabulário, organização sintática, etc.) e estruturais (tipos diferentes de fonte, destaques, imagens, etc.) do texto. Concomitantemente, precisa estabelecer relações entre o conteúdo do texto e seus conhecimentos prévios e entre as inferências feitas, monitorando de forma constante a manutenção do critério de fidelidade. Há, ainda, as demandas cognitivas relacionadas à construção de um novo texto, tais como escolha de vocabulário adequado, correção ortográfica e estrutural, correspondência fiel de conteúdo, etc. **Esses problemas podem ser agravados se o resumidor, por alguma razão, não puder dispor do texto-fonte no momento de escrever o resumo**, em razão da perda de informações devido à pouca capacidade de armazenamento da memória de trabalho (Alves, 2010, p. 44 - grifo nosso).

Conforme destacado acima, a não-disponibilidade do texto fonte ao resumidor no momento da atividade de sumarização acarreta uma demanda cognitiva ainda maior à memória de trabalho. Isso sinaliza a possibilidade de o escritor gastar um tempo ainda maior na elaboração de enunciados no papel, dado que precisa fazer uso da memória para recuperar dados importantes do texto original, ao mesmo tempo que faz um julgamento sobre a relevância ou não de se incluir esses dados no texto produzido como resumo.

van Dijk (2004) ainda faz algumas considerações sobre os processos sintáticos envolvidos na atividade de sumarização. Recuperando os conceitos de macro e microestrutura, faz uma distinção entre o chamado episódio, conceito de base semântica, e o parágrafo em si, que nada mais é do que a manifestação superficial dessa base semântica. No discurso escrito, o episódio pode ser construído a partir da localização de determinadas “pistas” gramaticais presentes no texto, como paragrafação, tópicos frasais, marcadores de mudança temporal ou espacial e a “introdução de predicados que não possam ser subsumidos debaixo do mesmo (macro) predicado e/ou não combinem com o mesmo script ou frame” (ALVES, 2010, pp. 46-7).

Machado, Lousada & Abreu-Tardelli (2004, p. 22) destacam a importância de se levar em conta uma reflexão sobre fatores externos ao texto em si. Isso implicaria tornar o aluno capaz de reconhecer, numa situação concreta de produção escrita, elementos como a função social do autor (que lugar ocupa na sociedade e no contexto considerado), a imagem que o autor tem de seu destinatário (traços básicos inerentes à definição de um público-alvo), locais e/ou veículos onde possivelmente o texto circulará (para se determinar o tom, a escolha vocabular, a necessidade de se incluir ou não um dado específico), o possível momento da produção e o objetivo do autor do texto.

É digno de nota também que as autoras se preocupam em apresentar ao aluno em seu material didático uma série de estratégias concretas envolvidas na etapa de sumarização. Poderíamos agrupar tais estratégias em duas grandes subdivisões. Uma diz respeito a questões de leitura, o que envolveria, por exemplo, antecipar o conteúdo a ser lido e posteriormente resumido; reconhecer as marcas linguísticas que indicam o público-alvo e a época de produção do original; e saber determinar com precisão alguns dos movimentos retóricos utilizados pelo autor que apontariam para a macroestrutura argumentativa do texto, manifesta na questão discutida, posição assumida e posição rejeitada, argumentos de sustentação e conclusão (2004, pp. 23, 39). A segunda subdivisão tem a ver com estratégias de produção ativa, mais ligadas a questões de escrita. Dizem respeito a uma série de procedimentos de uso da língua que contribuiriam para excluir informações acessórias, reduzindo a carga semântica do original. As autoras as reconhecem conforme a seguir (2004, p. 26):

a) Apagamento de conteúdos facilmente inferíveis a partir de nosso conhecimento de mundo;

- b) Apagamento de sequências de expressões que indicam sinonímia ou explicação;
- c) Apagamento de exemplos;
- d) Apagamento das justificativas de uma afirmação;
- e) Apagamento de argumentos contra a posição do autor;
- f) Reformulação das informações, utilizando termos mais genéricos;
- g) Conservação de todas as informações, dado que elas não são resumíveis.

Tais estratégias constituem um exercício de sumarização em si, na medida em que estimulam o aluno a refletir sobre sua própria leitura do texto e reconhecer, de maneira prática e concreta, o que de fato pode ser eliminado do original, segundo o contexto de produção e o objetivo que se pretende atingir com o resumo.

As estratégias apresentadas acima resguardam uma semelhança, até certo ponto, com os processos empregados por sumarizadores eletrônicos na confecção de resumos automáticos. Examinemos brevemente alguns aspectos relevantes nesse respeito.

## 2.3

### **Sumarização eletrônica: algumas considerações**

As pesquisas em sumarização automática têm contribuído muito, em anos recentes, para o desenvolvimento de ferramentas mais eficazes de produção de resumos. Radev & McKeown (2002), em seu artigo *Introduction to the Special Issue on Summarization*, ressaltam a importância que o desenvolvimento de sumarizadores automáticos passou a ter nas pesquisas linguísticas. À medida que o volume de informações *online* cresce, é cada vez mais desejável a existência de uma ferramenta computacional sólida que possa fornecer ao leitor um resumo consistente e eficaz do bloco texto examinado, garantindo economia de tempo e otimização da pesquisa. É digna de nota a definição que Radev e McKeown fornecem para o gênero resumo, pois revelam um aspecto importante do processo de sumarização:

“A summary can be loosely defined as a text that is produced from one or more texts, that conveys important information in the original text(s), and that is no longer than half of the original text(s) and usually significantly less than that. *Text* here is used rather loosely and can refer to speech, multimedia documents, hypertext, etc” (2002, p. 399).

Seguindo essa definição, podemos estabelecer alguns pressupostos fundamentais: a) um resumo está sempre atrelado a um texto (ou grupo de textos) fonte; b) em princípio, o objetivo geral de produção de um resumo é a apresentação das características mais relevantes do(s) texto(s) fonte; c) o resumo deve, obviamente, ser substancialmente menor que o material a partir do qual foi produzido; d) a definição de texto dada parece se referir a algo bem específico, a saber, um grupo de enunciados em sua forma concreta (mídia impressa ou eletrônica). Fica claro, conforme os autores salientam, que a grande dificuldade da sumarização automática reside justamente em como identificar e apresentar determinados fragmentos do material fonte como sendo de maior relevância em detrimento do conteúdo restante. O estabelecimento de critérios para tal processo, bem como a avaliação das ferramentas produzidas parece ser um ponto de particular interesse nos estudos em Linguística Computacional.

Uma informação bastante útil presente no texto de Radev e McKeown a esse respeito é a tentativa de classificação dos diferentes resumos produzidos. Tal classificação nos proporciona maior visibilidade sobre a função comunicativa que o resumo deve apresentar, o que facilita a elaboração e comprovação de critérios para seleção de informações relevantes. Diversos estudos feitos nesse campo contribuíram para a identificação dos seguintes tipos de resumo:

- a) *indicativos*: responsáveis por fornecer uma visão geral do texto, sem se prenderem a informações específicas;
- b) *informativos*: espécie de versão reduzida do original, concentrada nas ideias que compõem o conteúdo como um todo;
- c) *centrados em um tópico*: conforme o nome já indica, são resumos produzidos a partir de um ou mais tópicos de interesse do leitor, dentro do universo do texto, ou grupo de textos, fonte;
- d) *genéricos*: aqueles que refletem o ponto de vista do autor do material fonte.

Tendo lançado essa nomenclatura simples e funcional, os autores passam a explicar dois dos principais conceitos em sumarização automática, a saber: extração e abstração. Nesta pesquisa, vamos nos deter apenas no primeiro processo, por ser o que resguarda mais semelhanças com a sumarização realizada por seres humanos. Segundo Radev e McKeown, *extração* é o

processo de criação de um resumo a partir de porções reutilizadas de um *input* textual, como frases e palavras, ao passo que a *abstração* envolve a criação de um novo material a partir dos fragmentos coletados no processo de extração. Esses dois conceitos fundamentais podem se combinar a outras etapas importantes, compondo uma estrutura mais complexa. Teríamos então a confecção automática de um resumo em etapas bem definidas, iniciando-se com a identificação e separação pura e simples de elementos textuais (extração), seguida da reformulação desse novo *input* (abstração), que pode ocorrer através de uma  *fusão*, ou combinação de porções selecionadas, ou de uma *compressão*, que, como o nome já indica, envolve a eliminação de informações consideradas desnecessárias para o tipo de resumo produzido.

O método de extração tem sido o mais utilizado na elaboração de ferramentas computacionais de produção de resumos, e o processo todo funciona basicamente com a aplicação de critérios gerais de identificação das passagens mais importantes de um texto. Alguns destes são: *posição da informação* (dependendo do parágrafo ou do posicionamento que ocupam em relação a outras porções, algumas sentenças podem ter maior ou menor relevância); *frequência* (número de vezes que determinada palavra ou expressão ocorre no texto, sendo, em princípio, diretamente proporcional a seu grau de relevância); e *uso de expressões-chave* (encadeamentos vocabulares que teoricamente introduzem passagens relevantes, como “é interessante notar que”, “vale ressaltar que”, “é importante” etc). Uma das técnicas de sumarização citada por Radev & McKeown – e talvez a mais interessante do ponto de vista linguístico – tem por base a análise de linguagem natural, sobretudo no que tange à relação entre palavras e à estrutura discursiva. Os autores explicam que determinadas pesquisas utilizam um critério que se baseia no grau de correlação lexical entre passagens em potencial e o restante do texto, o que implicaria avaliar o número de palavras, sinônimos e anáforas em comum, por exemplo. Uma variação dessa técnica seria a extração de passagens que incluíssem elementos linguísticos em comum com um *input* verbal produzido dentro de determinado tópico de interesse do usuário.

Jones (1998, p.1), ao prover uma definição para resumo, novamente resgata dois conceitos fundamentais: o de seleção de informações para extração e o de generalização do material separado como mais relevante. Acrescenta ainda a noção de que a produção de um resumo automático em geral obedece a um processo de três etapas: *interpretação* do texto fonte, que por sua vez resultará em uma representação; *transformação* da representação obtida na

etapa anterior, que conduzirá a uma nova representação, desta vez a de um texto resumido; e *geração* de um resumo final, com base na representação do texto resumido. Esse esquema, conforme a própria autora defende, pode parecer óbvio, mas facilita muito a compreensão da lógica de funcionamento adotada por sistemas mais específicos, justamente por fornecer uma base comum que permita identificar as premissas em que se fundamentam esses sistemas, permitindo a comparação com outros.

Em seguida, Jones procura detalhar uma metodologia específica que deseja propor, concentrando-se em fatores contextuais. Existe uma aproximação muito forte entre o ponto de vista da autora e questões tipicamente ligadas a gêneros textuais, o que fica evidente quando ela afirma que “effective summarising requires an explicit, and detailed, analysis of context factors, as is apparent when we recognise that what summaries should be like is defined by what they are wanted for, as well as by what their sources are like” (1998, p. 5).

A partir dessa percepção, Jones constrói uma nomenclatura para classificação dos fatores contextuais a que se refere. Os três grandes grupos nesse sentido são fatores relacionados ao *input*, *output* e objetivo. Embora tente detalhá-los, a autora reconhece a dificuldade de definir de maneira precisa cada um dos fatores contextuais de maneira que componham uma ferramenta sólida, aplicável a todo tipo de texto. Vejamos como cada um dos três fatores é descrito.

Em primeiro lugar, temos o fator relativo ao *input*, ou seja, o material fonte. Dentro dessa categoria, Jones reconhece a existência de três subdivisões, a saber: *forma* (ligada a traços como macroestrutura textual, padrões retóricos, tamanho do texto e gênero a que pertence); *tipo de assunto* (domínio semântico e restrição lexical são questões pertinentes); e *unidade* (se o resumo será produzido a partir de um único documento ou de vários). Jones defende, com muita propriedade, que a produção do resumo precisa ser moldada a partir das características do material fonte, o que se torna mais fácil se adicionarmos todos os critérios acima à equação.

Em segundo lugar, a autora descreve o fator que considera mais importante, a saber, o objetivo. Dentro desse grande campo, estabelece novamente três subcategorias: *situação* (contexto comunicativo no qual o resumo será utilizado); *público-alvo* (como o próprio nome indica, a classe de leitores à qual o resumo é destinado); e *uso* (novamente auto-explicativo: com que finalidade o resumo será utilizado? Recuperar informações do texto original? Obter uma visão geral do material fonte?).

Por último, o fator relacionado ao *output*. Aqui, Jones reconhece como subdivisões o *material* (até que ponto o resumo se destina a capturar todo – ou parte de – o conteúdo relevante do texto fonte); *formato* (organização dos elementos para-textuais e distribuição do texto na página); e *estilo* (direção comunicativa que o resumo assume, como informar, apresentar dados, oferecer uma crítica, agregar vários pontos de vista etc).

Concluimos que a sumarização automática obedece a um processo bem definido, que perpassa basicamente dois caminhos: extração de informações relevantes, detectadas a partir de uma série de critérios linguísticos previamente fixados e geração de novas informações a partir dos dados obtidos na extração. Além disso, a inclusão de fatores contextuais, relativos a gêneros discursivos, tem se mostrado uma ferramenta eficaz na elaboração de processos de sumarização automática, na medida em que torna mais precisos os critérios de extração de informações e assegura uma maior fidelidade do resumo ao material fonte.

## 2.4 O papel da leitura

Examinemos, agora, algumas questões importantes no que tange à leitura e sua relação com a elaboração de resumos. Coscarelli (2002, p. 2) propõe que o complexo processo da leitura, ou decodificação de signos, seja dividido em duas partes fundamentais. A primeira diz respeito ao *processamento da forma*, que pode ser subdividido em uma etapa de decodificação lexical e outra de decodificação sintática (*parsing*). A segunda parte tem a ver com o *processamento do significado*, que pode ainda ser desmembrado em três etapas, levando-se em conta a maneira como ocorre a construção da coerência, a saber: local (análise do significado das frases e a relação entre elas, resultando na formação das proposições que constituirão a macroestrutura textual), temática (organização das proposições e formação de um construto superior, privilegiando-se determinados aspectos estruturais em detrimento de outros) e externa (recuperação de conhecimento prévio do leitor com fins de interpretação das informações do texto e sua pertinência para o objetivo da leitura).

É importante reconhecer que cada uma dessas etapas de processamento pode ser facilitada ou dificultada – o que resultaria, por exemplo, numa variação de tempo gasto na leitura – de acordo com uma série de fatores associados ao

perfil do leitor e à própria língua (Coscarelli, 2002, pp. 3-15). O processamento da forma, por exemplo, pode ser mais rápido ou mais lento, tanto no nível lexical quanto no sintático, dependendo de aspectos como a complexidade silábica de uma palavra, seu grau de familiaridade para o leitor, a frequência com que ocorre na língua, a ambiguidade sintática, a presença ou não de sentenças do tipo *garden path* – sentenças “labirinto”, que admitem mais de uma concatenação entre componentes sintáticos e que, portanto, dificultariam o processamento, como no exemplo: “Alguém atirou contra o empregado da atriz que estava na varanda” etc. No caso do processamento do significado, diríamos que a construção da coerência local seria facilitada ou dificultada segundo o grau de familiaridade que o leitor apresenta com o assunto lido, por exemplo. Já a coerência temática poderia ser estabelecida mais rápida ou lentamente de acordo com fatores como o conhecimento que o leitor partilha sobre aquele determinado gênero lido, a canonicidade ou não da organização textual (se o texto encontra-se estruturado segundo o consenso) e a capacidade do leitor de identificar as ideias mais relevantes de acordo com seu propósito. Por último, alguns fatores que afetariam a construção da coerência temática são: a capacidade do leitor de fazer julgamentos, generalizações, analogias etc. e aspectos relacionados à quantidade e velocidade de recuperação das informações na memória de trabalho.

Alves (2010) complementa essa perspectiva ao demonstrar a necessidade de se dominar o significado das palavras dentro de uma dada relação sintagmática, se temos de ter uma compreensão efetiva do texto que estamos lendo. Isso significa fazer sentido das palavras enquanto conjunto, não de maneira isolada no texto, demonstrando capacidade para reconhecer as relações hierárquicas e semânticas que assumem em determinado contexto. Para tal, segundo Alves, o leitor faz uso de uma série de estratégias que regulam o processo de compreensão, ajudando-o a formular determinadas hipóteses sobre o conteúdo do texto que está lendo e aos poucos ir confirmando, abandonando ou reformulando tais hipóteses, num constante movimento progressivo e regressivo. Essas estratégias de compreensão estão ligadas a uma variedade de atividades cognitivas pertinentes ao processamento da leitura, como

compreender os propósitos da leitura; ativar os conhecimentos prévios sobre o assunto; saber manter o foco no essencial; avaliar a consistência interna do texto (em termos de coesão e coerência); controlar a manutenção e a coerência da

compreensão e, por fim, conferir se o sentido global do texto coincide com as hipóteses do início da leitura. (SOLÉ, 1998 *apud* ALVES, 2010, p. 20)

Garcez (2004, p. 24) apoia esse ponto quando afirma que o leitor precisa “compreender simultaneamente o vocabulário e a organização das frases; identificar o tipo de texto e o gênero; ativar as informações antigas e novas sobre o assunto; perceber os implícitos, as ironias, as relações estabelecidas com o nosso mundo real”.

É de interesse considerar também o que leva uma pessoa a ler determinado texto em particular. Alves (2010, pp. 21, 22) afirma que uma leitura guiada por determinados objetivos, ou expectativas, pode fazer com que o leitor se movimente com maior liberdade no contexto em busca de informações específicas que lhe interessam. Esses objetivos podem incluir compreender de que trata o texto, ter uma visão geral do conteúdo e buscar informações específicas, por exemplo.

Por fim, Alves (2010, p. 25) aponta para uma série de conhecimentos que o leitor deve ter, de modo a interpretar de maneira eficiente determinado conteúdo. Esses conhecimentos, fortemente associados à noção de gêneros textuais, dizem respeito à percepção da situação comunicativa em que o texto se insere, conhecimento paralinguístico presente no texto – legendas, figuras, relevos etc –, conhecimento de ordem linguística (elementos morfológicos, sintáticos e semânticos), conhecimento textual, ou discursivo, que diz respeito às relações entre os níveis estruturais do texto (referência, coesão etc.) e conhecimento de mundo, ou enciclopédico.

Garcez (2004, pp. 26-7) amplia essa ideia ao citar as diversas imposições simultâneas que o processo de leitura faz ao cérebro, compreendendo diversas habilidades sofisticadas a serem dominadas ao longo da formação do leitor, como antecipação de informações, agrupamento de palavras em blocos conceituais, seleção e hierarquização de ideias, focalização da atenção e reorientação dos procedimentos mentais. Essa autora afirma que faz parte do processamento da leitura lidar “com a capacidade simbólica e com a habilidade de interação mediada pela palavra”. Isso pressupõe a manipulação de diversos elementos, como signos, frases, sentenças e argumentos, abrangendo não só a competência linguística, mas também a experiência de vida acumulada pelo indivíduo.

Coracini (2009, p. 30) faz referência ainda a um modelo de construção/integração proposto por Kintsch, que divide o processo de representação mental de um texto em duas fases, uma de construção, na qual o

leitor forma um modelo mental aproximado do texto com base em sua leitura e conhecimento de mundo, e outra de integração, em que tal modelo mental sofre um processo de estabilização, descartando-se informações irrelevantes e redundantes. As proposições formadas são ativadas e avaliadas no que tange às relações que estabelecem umas com as outras. Ligações mais fracas são descartadas, resultando no fortalecimento de uma representação mental específica, que é incorporada à memória de trabalho.

Examinaremos, no capítulo seguinte, alguns pressupostos teóricos relativos ao processamento da escrita, sobretudo modelos processuais, destacando sua relevância no estudo de resumos acadêmicos.