

PERCEPÇÃO LINGÜÍSTICA E FATIAMENTO

Mário A. Perini
Lúcia Fulgêncio
Maria Bernardete Rehfeld
Universidade Federal de Minas
Gerais

1. INTRODUÇÃO

Pretendemos, neste trabalho, examinar alguns aspectos de uma noção fundamental dos atuais modelos de leitura, assim como de compreensão em geral: a noção de "fatia". Todos os modelos atuais admitem que um dos processos básicos, dentro do processo geral da compreensão de material lingüístico é o que agrupa elementos apresentados linearmente em grupos maiores, que constituem unidades de processamento ao nível imediatamente superior — assim, agrupam-se letras em palavras e palavras em sintagmas.

Essa maneira de falar, na verdade, deixa de exprimir uma realidade muito importante: ao dizermos que um leitor agrupa letras em palavras, podemos dar a impressão de que ele reconhece as letras primeiramente, e a partir dessa informação "monta" as palavras; mas não parece ser isso o que ocorre. A evidência existente não deixa dúvidas de que o reconhecimento de uma palavra pode preceder, ou mesmo condicionar, o reconhecimento das letras que a compõem. Assim, as letras individuais da seqüência JANELA podem ser reconhecidas a maior distância, ou mais rapidamente, ou com iluminação pior, do que as letras NLEAAJ, isso porque a primeira seqüência, mas não a segunda, constitui uma palavra.

Não se pode, portanto, dizer simplesmente que as fatias (palavras, no caso) são montagens a partir de elementos de nível imediatamente inferior. O leitor lança mão, já desde o primeiro momento, de conhecimentos armazenados na memória de longo termo (MLT), como por exemplo um léxico que inclui as representações visuais das palavras da língua. Ao tentar reconhecer uma palavra (ou uma letra dentro de uma palavra),

o leitor precisa jogar com a informação visual disponível e ao mesmo tempo fazer hipóteses quanto à identificação da palavra ou letra, baseado em informação presente em sua MLT, mais o léxico visual (lista das representações visuais das palavras) e /ou regras de redundância. Por "regras de redundância" entendemos o conhecimento que o leitor tem das regras ortográficas, sintáticas e semânticas da Língua. As regras ortográficas restringem as possibilidades de ocorrência das letras umas em relação às outras; as regras sintáticas especificam as combinações possíveis de palavras dentro de sintagmas e sentenças e as regras semânticas determinam a interpretação de uma seqüência através de informação fornecida pelo contexto lingüístico ou extra-lingüístico. Como é óbvio, no caso da identificação de palavras, tanto o recurso ao léxico como o recurso às regras de redundância só são úteis no caso de a seqüência a identificar ser uma palavra da língua. Se a seqüência for NLEAAJ, o leitor dependerá exclusivamente da informação visual para o reconhecimento da seqüência como um todo; isto é, ele será forçado a reconhecer a seqüência letra por letra, e necessitará de mais tempo, assim como de melhores condições de visibilidade.

Pode-se dizer, portanto, que a seqüência JANELA pode constituir uma fatia de informação na leitura corrente porque constitui uma palavra da língua: está de acordo com regras que estabelecem as co-ocorrências permissíveis de letras em português e, além disso, existe, na MLT dos falantes do português, uma representação visual dessa seqüência, rotulada como parte do item léxico 'janela'. O resultado final é que é mais fácil reconhecer a seqüência JANELA, ou mesmo as letras que a compõem (digamos, "a primeira letra") do que a seqüência NLEAAJ. JANELA, em suma, tem condições de ser uma fatia, e portanto de ser processada como um todo, ao contrário de NLEAAJ, que provavelmente só pode ser processada como seis elementos em seqüência.

É essa noção de "fatia" que está no centro da discussão que desenvolveremos neste trabalho. Parece-nos que a idéia de fatia que subjaz à maioria das discussões sobre compreensão de material lingüístico é insuficientemente clara, e envolve na verdade mais de uma noção. No que se segue, apresentaremos um modelo do processo de compreensão na leitura, proposto por Massaro (1975) e seus colaboradores, modelo esse que aceitamos como adequado em suas linhas essenciais. Discutiremos depois a noção de fatia tal como se depreende da literatura especializada, criticando-a e refinando-a (na verdade, substituindo-a por duas noções distintas).

Procuraremos então integrar a nova concepção de fatia dentro do modelo de leitura que aceitamos.

2. UM MODELO DE LEITURA

Massaro (1975) apresenta uma proposta de um modelo de leitura que inclui os seguintes componentes: um registro sensorial (*preperceptual visual storage*); um processo de reconhecimento primário (*primary recognition process*); uma memória visual (*synthesized visual memory*); um processo de reconhecimento secundário (*secondary recognition process*); uma memória abstrata (*generated abstract memory*); e um processo de ensaio e recodificação (*rehearsal and recoding process*). Cada processo trabalha sobre o material contido na memória imediatamente anterior (considerando-se o registro sensorial também como um tipo de memória), recodificando a informação em termos da memória imediatamente posterior. O resultado do último processo é a abstração do significado de sintagmas e sentenças, e presumivelmente a sua transferência para a MLT. Esmiuçamos a seguir alguns dos pontos mais importantes do modelo:

O reconhecimento primário (R-1) tem como função o reconhecimento gráfico das palavras:

"The primary recognition process operates on a number of letters simultaneously (in parallel). Ten visual features read out at each spatial location define a set of possible letters for that position. The recognition process chooses from this candidate set the letter alternative that not only has a high correspondence in terms of visual features but also is probable in that particular context. The primary recognition process is therefore dependent on both the visual information in preperceptual storage and knowledge about the probabilities of letter strings held in long-term memory. The interaction of these two sources of information is a critical issue in the analysis of word recognition.

The primary recognition process transmits a sequence of recognized letters to synthesized visual memory."

(Massaro, 1975, p. 241-2)

Quanto ao reconhecimento secundário (R-2), sua função é a de processar as palavras, identificando-as e conectando-as com seus itens léxicos — incluindo as representações semânticas:

"[...] synthesized visual memory holds a sequence of letters that are operated on by secondary recognition process, which tries to close off the letter string into a meaningful word. The secondary recognition process makes this transformation by finding the best match between the letter string and a word in the LTM lexicon."

(p. 242)

Ou, mais adiante:

"The secondary recognition process operates to transform the visual information into meaning. The visual information made available by the primary recognition process and syntactic and semantic expectancies operate hand in hand in the abstraction of meaning. Our model assumes that syntactic and semantic redundancy operates at this point in time to facilitate word recognition."

(p. 286)

Para ilustrar a interação entre as duas fontes de informação (informação visual e conhecimento das probabilidades de ocorrência de letras) na identificação de letras e no fechamento de palavras, tomemos o exemplo da figura 1 abaixo:

sentou na rua

— Figura 1 —

A forma gráfica 'u' aparece aí quatro vezes; mas duas vezes ela será normalmente lida como 'n' e duas como 'u'. Em cada ocorrência o R-1 capta a informação visual disponível e interpreta-a não só com base na sua forma, mas também com base nas probabilidades de ocorrência de determinada letra naquele contexto e na existência ou não de um item léxico com aquela composição; isso implica, naturalmente, em um tipo de "colaboração" entre o R-1 e o R-2, que não devem assim ser concebidos como necessariamente ordenados temporariamente um após o outro. Voltando ao exemplo, considerando apenas a informação visual, mais o conhecimento da forma das letras, mas sem utilizar a redundância, o R-1 poderia ler *seuton na rua*, ou *seuton na rua*, ou *senton na rua* etc. Mas o conhecimento que tem o leitor das regras de sua língua e de seu léxico exclui todas essas leituras; a única possibilidade é *sentou na rua*. Em consequência, o leitor "vé" um n em na e um u em rua.

Como vimos acima, o resultado da operação do R-2, que dá entrada na memória abstrata, parece ser material composto basicamente de palavras identificadas lexicamente, o que inclui suas representações semânticas. Falta ainda, evidentemente, uma importante função de reconhecimento, que é a de integrar as sequências de itens léxicos em frases sintático-semânticas. Essa tarefa fica a cargo do processo de ensaio e recodificação (E&R). Aqui os autores não são tão explícitos quanto no caso de R-1 e R-2, mas acreditamos que a interpretação dada acima é correta. As seguintes passagens podem ser mencionadas:

"The recoding process operates on the string of words in generated abstract memory to derive meaning from the entire phrase or sentence."

(p. 16)

"The recoding process tries to close off the words held in generated abstract memory into a meaningful form at the phrase or sentence levels."

(p. 19)

A "recodificação" nesse caso não implica em mudança da natureza da informação (ao contrário do que acontece, por exemplo, com a recodificação da informação da memória visual para a memória abstrata: na primeira, a informação é codificada em forma gráfica; na segunda, em forma abstrata, ou seja, sintática e semântica:

"This operation [o processo de recodificação efetuado pelo E&R] does not change the nature of the information, allowing it to be recirculated through generated abstract memory."

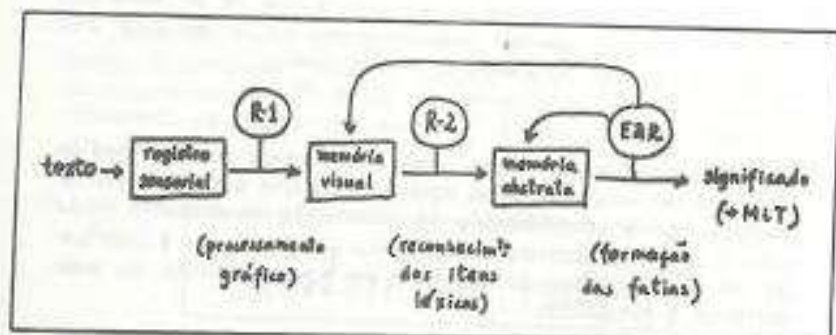
O processo de recodificação, naturalmente, tem acesso ao léxico e à gramática (regras sintáticas e semânticas), assim como a todo o conhecimento disponível ao sistema. Finalmente,

"The recoding process can also operate in reverse. Given an abstract idea, it can transform this idea into a sequence of words in generated abstract memory or a perceptual representation in synthesized auditory or visual memory. Finally, the processing at this stage may be a simple regeneration or repetition of information, in which case the operation is called *rehearsal*."

(p. 25)

Parece-nos que se deve concluir desta última passagem que a formação de fatias não implica em ensaio. Se for assim, podemos aceitar o modelo de Massaro & al. como uma aproximação aceitável, suficientemente fundamentada dado o atual estado do nosso conhecimento do assunto.

Em resumo, a informação visual colhida do texto passa através dos estágios de processamento ilustrado na figura 2.



3. Tanque de Reserva

Segundo Bouma & de Voogd (1974), apoiados por Sheblake (1975), o mecanismo da leitura precisa comportar, também, um tanque de reserva (buffer), que fornece as unidades básicas para o processamento central. Bouma & de Voogd procuraram simular experimentalmente a sequência de estímulos resultantes dos saltos oculares (saccades) e fixações na leitura normal. Assim descreve Sheblake o experimento:

"Subjects fixated the center of a small window that displayed 60 character positions on each simulated fixation, which lasted for some duration, T, after which the text disappeared and then reappeared after 25 msec. [...] Utilizing a line-step presentation, each new display in the window was shifted horizontally by some number of character positions, n. In one condition, the text remained on for the whole duration T; in another it went off after 300 msec, leaving the window dark. Usually T remained constant within each trial while n started at a low value and increased by one letter space every 10 sec. Subjects started with oral reading until the speed got too high and continued with silent reading until they felt that they were no longer able to keep up."

(Sheblake, 1975, p. 308)

Os resultados mostraram que a leitura é possível mesmo quando o tempo de exposição (T) e o número de posições que o texto era movido (n) se afastavam bastante dos valores de T e n ocorrentes na leitura normal. Bouma & de Voogd concluíram que

"In ordinary reading there is no need for a precise programming of each next saccade on the basis of preceding text recognition and [...] central recognition process are largely insensitive to parameters of eye saccades."

(Bouma & de Voogd, 1974, p. 380)

Desse modo, propõem eles que o material que dá entrada no sistema fica reservado em um tanque à espera de processamento. Esse mecanismo permite o que eles chamam "lax control" dos saltos oculares por parte do mecanismo de reconhecimento:

"[lax control ensures] only that on the average the advancement of the eye would keep pace with ongoing text recognition."

(ap. Sheblake, 1975, p. 308)

Assim descrevem eles o funcionamento do tanque:

"For efficient operation, the proceeding of the eyes over the text should then be under the control of the content of the buffer; an empty buffer should lead to increase of eye speed whereas a filled buffer should slow down the eyes."

(ap. Sheblake, p. 307)

O "lax control" advogado por Bouma & de Voogd, se opõe ao "strict control",

"according to which each successive fixation is programmed in connection with the recognition of text during the preceding fixation."

(p. 506)

Shebilske aceita a proposta como basicamente correta, embora reservando a possibilidade de que o controle escrito funcione em certos casos. Por exemplo, se um leitor encontra a expressão **Universidade Federal de Minas Gerais** a certa altura, e novamente algumas linhas mais adiante ele poderá utilizar informação proveniente da visão periférica para economizar uma fixação, que de outro modo seria gasta com informação altamente previsível. Para a maioria dos casos, entretanto, Shebilske aceita a teoria de Bouma & de Voogd como correta: o controle do processo de reconhecimento sobre a entrada de informação através de fixações é lax, e depende do conteúdo do tanque.

Shebilske aponta dois pontos de certa importância que não ficaram, a seu ver, suficientemente claros, a saber, a natureza das unidades contidas no tanque e o posicionamento do tanque dentro do modelo sumariado na figura 2 acima. Quanto ao primeiro ponto, a natureza das unidades com que o tanque trabalha, Bouma & de Voogd especulam que seriam ao nível das palavras ou dos morfemas:

"On this account, the eyes would move so as to keep a relatively constant number of words available for higher-order processing."

(Shebilske, 1975, p. 308)

Entretanto, Shebilske dá a entender que essa é apenas uma "especulação" de Bouma & de Voogd; interpretamos essa expressão como querendo dizer que aqueles autores não apresentam evidência em favor dessa teoria. Shebilske considera possível que as unidades do tanque sejam ao nível das letras, ou dos grupos de centro vocálico propostos por Hansen & Rogers (1968), mas remete a questão a pesquisas futuras, o que sugere que tampouco ele tem em mãos argumentos relevantes para a solução do problema.

Parece-nos, entretanto, que não há motivo para essa hesitação. Se o conteúdo do tanque for composto de unidades significativas (tais como palavras ou morfemas), um fenômeno importante relacionado com o controle da entrada de informação — o retardamento da leitura provocado pela dificuldade conceptual ou sintática do texto — ficará automaticamente expli-

cado, ao passo que se seguirmos a sugestão de Shebilske de que o tanque contém unidades não significativas (letras, grupos de centro vocálico) esse fenômeno ficará sem explicação. Isso tentaremos mostrar a seguir.

Vamos considerar primeiro o papel do tanque como regulador da entrada de informação no sistema. Segundo Bouma & Voogd, o material, ainda em forma de peças (sejam elas perceptos visuais, sejam representações abstratas), iria sendo acumulado no tanque; o processo relevante ficaria "vigiando" o tanque, e assim que este contivesse material suficiente para o fechamento de uma unidade, o processo (seja ele o R-2, seja o E&R) formaria a unidade, o que liberaria o tanque para receber mais material. Bouma & Voogd afirmam que o conteúdo do tanque é que controla a velocidade da leitura (isto é, a frequência de fixações por n palavras); mas na verdade o que controla a velocidade da leitura é o processo de fechamento de unidades. Quando, por exemplo, o conteúdo do tanque é tal que dificulta o trabalho do processo, atrasando-o, o processo manda uma mensagem ao controle dos olhos, solicitando que diminua a velocidade de entrada de novos itens. Os olhos reagem aumentando as fixações ou regredindo, dessa forma retardando a velocidade geral da leitura. Isso é bastante claro no texto, como se pode ver pela citação feita na página do trabalho de Shebilske, e reproduzida páginas atrás.

Mas voltando ao problema do tipo de unidades que o tanque contém: existe um fenômeno, dissemos, que poderá ser explicado a partir da proposta de que o tanque contém unidades significativas. Esse fenômeno é o retardamento que se observa como consequência da dificuldade conceptual ou sintática de um texto. Esse é um fato suficientemente conhecido para que não seja necessário argumentar em favor de sua ocorrência. Ora, consideremos, por exemplo, um texto composto de palavras familiares, mas difícil (para certo leitor) em virtude de seu conteúdo; o que se observa é que esse leitor precisa lê-lo mais lentamente a fim de compreendê-lo. Parece-nos que a explicação mais natural para esse fato é a de que está funcionando aí precisamente o mecanismo descrito no parágrafo precedente: a dificuldade em que se encontra o processo para integrar os diferentes itens do tanque em unidades de ordem superior causa um represamento de itens no tanque, que se vai enchendo. Consequentemente, o processo envia mensagem ao controle dos olhos para que diminua a velocidade de entrada da informação. Mas essa explicação só se sustenta se admitirmos que o que está no tanque são unidades dotadas de significado, e que o processo em questão é o que forma sintagmas

e frases (sintáticas e semânticas) a partir de palavras (ou, melhor provavelmente, de morfemas). Como as palavras, individualmente, são familiares ao leitor, não haveria razão para que o processo de fechamento de palavras a partir de informação não significativa (isto é, o R-2) ficasse assoberbado, causando o retardamento observado. Ou seja, se o tanque contivesse unidades não significativas (como letras, por exemplo), e se o processo que age sobre ele fosse o R-2, seria de esperar um retardamento apenas no caso de o texto conter palavras difíceis, nunca no caso de ele conter palavras familiares mas ser sintática ou semanticamente complexo. Desse modo, concluímos que é forçoso aceitar que o conteúdo do tanque se compõe de unidades significativas; digamos, palavras.

Resolvido o primeiro ponto duvidoso apontado por Shebilske, cremos que o segundo fica também resolvido: é claro que se o tanque contém unidades significativas (palavras), então o processo que atua na sua dependência é o E&R; por conseguinte, o tanque é parte da memória abstrata, ou talvez se identifique com ela.

Acerca deste último ponto a posição de Bouma & de Voogd não fica muito clara no texto de Shebilske. A discussão sugere que o tanque poderia ser um componente separado da memória (ainda que presumivelmente inserido na memória visual ou na abstrata), mas outras razões nos levam a preferir identificar o tanque com a memória abstrata. Aliás, não é impossível que essa idéia reflita a posição de Bouma & de Voogd, ainda que isso não esteja bem explícito. Por exemplo, ao introduzirem a noção de tanque (buffer), dizem eles

"For efficient reading the point of fixation should proceed on the average just as fast as recognition, and for bridging momentary differences between the eye and the brain, a buffer function has to be assumed."

(Bouma & de Voogd, 1974, p. 231, grifo nosso)

Se eles falam de uma "função de tanque", é provável que tenham em mente que um dos componentes da memória assumia essa função, não havendo pois necessidade de propor um novo componente.

Se essa é ou não a opinião de Bouma & de Voogd não precisa nos preocupar aqui. O fato é que a MCT, entre suas atribuições tradicionalmente reconhecidas, uma que nos parece semelhante à "função de tanque" que vimos discutindo (ver Atkinson & Shiffrin, 1968, entre outros). A maioria dos autores (embora discordem quanto aos detalhes) admite que a MCT,

ou algum subcomponente dela, desempenha a função de conservar por um breve espaço de tempo alguns itens à espera de processamento. Reitman (1970) chama a isso "sala de espera" (waiting room), e coloca-a como um componente separado da MCT. Sem querer discutir aqui a proposta de Reitman, podemos observar que, em termos do modelo aqui adotado, não há inconveniente em identificar o tanque com a memória abstrata. Em resumo, adotamos da proposta de Bouma & de Voogd a necessidade de que a memória abstrata funcione como tanque de reserva de itens que esperam ser processados pelo E&R. Continuaremos a falar de "tanque" para nos referirmos à memória abstrata quando considerada nessa função específica.

4. Fatias: tipo A e tipo B

Massaro & Schuller (1975), referindo-se a experimentos de reconhecimento de estímulos formados por palavras, letras, algarismos e coisas parecidas, dizem:

"In terms of the perceptual processing time relevant to reading, it is interesting to note that these studies have often resulted in an estimate of around 200 or 250 msec, or $\frac{1}{4}$ sec, which is also the time that the eye remains fixated at a given point in the text during reading. [...] It may be, then, that the reason these fixation periods extend for 200 msec is that this is the length of time necessary to synthesize the information available within one fixation."

(Massaro & Schuller, 1975, p. 237)

A coincidência apontada é realmente sugestiva, e a conclusão parece plausível. Mas é preciso ter em mente que eles se referem ao reconhecimento de elementos presentes literalmente na MLT do leitor. É com tais elementos que lidam os experimentos feitos para averiguar a rapidez do processamento. Por exemplo, uma letra ou uma palavra correspondem a um "signo" na MLT. O leitor reconhece, em um sentido bastante literal, o estímulo como sendo uma realização daquele signo. O processo consiste em comparar o estímulo com o conjunto de traços presente no signo, e caso haja suficiente semelhança, decidir que se trata de uma ocorrência daquele signo.

1. Vale a pena repetir: como não podemos consultar o artigo original de Bouma & de Voogd, não podemos avaliar com exatidão a extensão em que essas idéias coincidem com as deles. Por ora, pois, assumiremos a responsabilidade de identificar o tanque com a memória abstrata.

Nesse contexto, o "signo" corresponde a uma fatia de informação, tal como o termo "fatia" é tomado em trabalhos como o de Simon (1974), por exemplo. Entretanto, parece-nos que essa noção de fatia tem sido indevidamente generalizada para o contexto da leitura (e da compreensão em geral, onde, na verdade, o mais das vezes não se trata de entidades literalmente presentes na MLT do leitor. Vamos, abaixo, examinar o uso do termo "fatia" na literatura corrente, e tentaremos mostrar que há na verdade duas noções bastante diferentes, cuja distinção nem sempre é respeitada.

O primeiro a usar o termo "fatia" (chunk) foi Miller (1956), quando estabeleceu que a capacidade da MCT poderia ser traduzida no "número mágico 7, mais ou menos 2". Embora Miller tenha admitido que não lhe era possível ser muito definido com relação ao que constitui uma fatia, ao mesmo tempo relacionou e interligou essa noção exclusivamente a conceitos já presentes como unidades inteiras dentro da MLT:

"For example, the memory span of five words [...] might just as appropriately have been called a memory span of 15 phonemes, since each word has about three phonemes in it. Intuitively, it is clear that the subjects were recalling five words, not 15 phonemes, but the logical distinction is not immediately apparent. We are dealing with a process of organizing or grouping the input into familiar

(Miller, 1956, grifo nosso)

Miller afirmou, além disso, que a capacidade da memória mantém-se sempre constante. Quando em vez de 7 letras conseguimos repetir 21, como por exemplo em 7 palavras de 3 letras cada uma, o que aumentou não foi a capacidade da memória, mas o tamanho daquilo que se toma como unidade. Ao construirmos fatias maiores estamos, por conseguinte, aumentando o total de informação presente na MCT, mas mantemos inalterado o número limite de fatias que a MCT pode conter:

"Since the memory span is a fixed number of chunks, we can increase the number of bits information that it contains simply by building larger chunks, each containing more information than before."

(Miller, 1956)

Para testar a hipótese de que a capacidade da memória se mantém constante em torno do número de 7 fatias, Simon (1974) desenvolveu testes empíricos, obtendo resultados que não confirmavam inteiramente a hipótese sugerida por Miller.

Seus resultados evidenciaram que o material presente na MCT poderia ser de 7 palavras de uma sílaba, 7 palavras de 2 sílabas, 6 palavras de 3 sílabas, mas somente 4 SNs de 2 palavras (como por exemplo *differential calculus, criminal lawyer*) e 3 sintagmas maiores (como por exemplo *to be or not to be, that is the question, in the beginning was the word*). Esses resultados levariam à conclusão de que as unidades da memória são fatias (e não, por exemplo, letras ou sílabas) mas, na verdade, a capacidade da memória não é tão constante como foi proposto por Miller, já que ao se aumentar o tamanho da presumível fatia diminui-se paralelamente o número de fatias que a MCT pode conter. Conclui Simon que

"... the 'constant capacity (chunks) hypothesis is a rough first approximation of the true state of affairs..."

(Simon, 1974, p. 482)

Restam portanto as perguntas: Qual foi o problema com a testagem feita por Simon? Por que seus resultados empíricos foram contrários à tese original da constância da capacidade da MCT? Na tentativa de responder essas perguntas, examinaremos mais atentamente a formulação da noção de fatia.

É importante notar, em primeiro lugar, que o fatiamento se utiliza da informação presente na memória de longo termo, como acontece, por exemplo, com a identificação das letras e das palavras. Portanto, o fatiamento envolve a mediação da MLT.

O ponto crucial da nossa crítica está no tipo de material presente na memória de longo termo. Um dos tipos de material armazenado nesse depósito são informações em sua forma literal, digamos assim, "concreta": letras, palavras, algarismos etc. Segundo a formulação original de Miller, e segundo ainda os exemplos apresentados que apóiam sua proposta, as fatias são formadas unicamente de material "familiar". Isso poderia ser interpretado como significando que as fatias são compostas exatamente do mesmo material presente na MLT, isto é, como se todo o trabalho na construção de fatias fosse o de buscar na MLT as informações aí armazenadas como unidades, trazendo-as para o nível da MCT. Dessa forma teríamos o direito de dizer que **qualquer** unidade presente na MLT tem a possibilidade de vir a constituir uma fatia.

Foi exatamente essa noção de fatia que Simon pressupôs ao realizar seus experimentos e tirar suas conclusões. Observemos que Simon usou sílabas, palavras e sintagmas do tipo "chavão", que estavam presentes como um inteiro na MLT.

Podemos deduzir, a partir do material testado por Simon, que todo o seu experimento se baseava na noção subjacente de que as fatias se compõem de elementos presentes literalmente na MLT:

"The chunking hypothesis does not assert that the word will always be the unit. Units much larger than words may be highly familiar, hence may serve as chunks."

(Simon, 1974, p. 483)

Poderíamos ir ainda mais longe do que Simon, dizendo que até mesmo poemas inteiros, ou qualquer item, de qualquer tamanho, que tenha sido decorado pelo indivíduo (estando portanto em sua MLT) poderia, em princípio, formar uma única fatia. No entanto, poderia ser impossível a testagem dessa hipótese de que qualquer material, independentemente de seu tamanho, possa formar uma fatia única. Vejamos porquê.

Como sabemos, a MCT tem uma duração breve, e o material nela contido pode ser esquecido devido a dois fatores: passagem de tempo e interferência de outros estímulos. Sabemos também que, por definição, qualquer material "familiar", ou seja, presente na MLT, pode constituir uma fatia única. Portanto, seria teoricamente possível que uma pessoa repetisse sem problemas, logo após uma apresentação, uma série de 6 letras não relacionadas mais um dado poema que o sujeito soubesse de cor previamente. Esses elementos somados totalizariam o número de 7 fatias, rigorosamente dentro do limite da capacidade da memória. No entanto, esse teste seria impraticável na realidade, já que enquanto o sujeito estivesse recebendo do experimentador o texto do poema para armazená-lo na MCT, os dois fatores de esquecimento estariam interferindo na posterior repetição das demais 6 letras. Queremos dizer que tanto a interferência das palavras do poema quanto o tempo necessário para pronunciá-lo estariam prejudicando o desempenho do indivíduo na lembrança das 6 primeiras letras. Em outras palavras, enquanto o indivíduo estivesse recitando o poema, o tempo estaria passando, esgotando o limite temporal de lembrança na MCT. Além disso, possivelmente as palavras do poema estariam interferindo na lembrança das outras fatias.

A experiência inversa (em primeiro lugar o poema e em seguida as 6 letras) seria igualmente impossível de ser testada, se vinculermos a devolução do material apresentado a uma repetição seriada, isto é, na mesma ordem em que foi apresentada. Enquanto o sujeito repetindo o poema, os dois fatores

de esquecimento da informação contida na MCT estariam agindo, causando a perda das 6 letras. No entanto, se não fícermos fechados dentro da exigência de uma repetição seriada, obteremos resultados reveladores. Se apresentarmos a um indivíduo em primeiro lugar o poema decorado e posteriormente uma série de 6 letras, e se aceitarmos que ele repita em primeiro lugar as seis letras e em seguida o poema, aí então verificaremos que, se isso é possível de ser realizado por um indivíduo, logo todas as letras e o poema estariam contidos na sua MCT.

Poderia ser argumentado que, depois de dizer as seis letras, o indivíduo iria buscar o poema na MLT, isso quer dizer que o que a MCT teria guardado não seria propriamente o poema inteiro, mas, por exemplo, o seu título, ou a frase inicial, que lhe permitiria buscar na MLT o poema. A MCT conteria então algo como

< C + P + W + U + M + T + "Os Lusíadas" >

Não negamos que essa hipótese seja razoável, mas não vamos maneira de se decidir seguramente se é isso o que acontece ou não, pelo menos a partir dos dados da recuperação. E note-se que a mesma objeção pode ser levantada contra os experimentos de Simon: segundo esse ponto de vista, o que a MCT teria guardado não seriam propriamente "frases", mas "rótulos", o que possibilitaria a recuperação das frases, exatamente como no caso do poema. Ou seja, parece-nos que tanto no experimento de Simon quanto no que aqui propomos, temos o mesmo tipo de material (elementos presentes literalmente na MLT), e o mesmo fenômeno.

Isso nos leva à conclusão de que a "fatia" com que Simon lidou em seu experimento é de tamanho em princípio ilimitado, dado que mesmo um longo poema pode estar contido em uma só fatia: a única exigência seria a de que esse poema deveria estar presente como uma unidade na MLT do sujeito. Ou, para dizer o mesmo de outro ponto de vista, a hipótese de que uma fatia possa ter tamanho ilimitado é compatível com a noção de fatia expressa por Miller e retomada por Simon. Não nos parece que experimentos do tipo dos realizados até o momento possam invalidar essa hipótese — o que, sabemos bem, tira muito do interesse dessa noção de fatia para o estudo da compreensão.

A esse tipo de fatia, definido por Miller e testado por Simon, chamaremos "fatia tipo A". Podemos concluir que, nes-

te primeiro tipo de fatia, todo e qualquer item presente como um inteiro na MLT pode ser tomado como uma fatia, independentemente de seu tamanho.

Agora podemos passar a determinar o problema ocorrido com os testes de Simon, ou seja, por que ele encontrou discrepâncias significativas no número de fatias versus o tamanho das mesmas. Como já vimos, Simon encontrou uma relação inversamente proporcional entre o tamanho da fatia e o número de fatias contido na MCT:

" [...] something about the additional length of the material reduces the total number of imputed chunks that can be retained [...]"

(Simon, 1974, p. 453)

Mas gostaríamos de lembrar, por outro lado, que o tempo gasto na reprodução de cada fatia também mantém uma relação com o tamanho da fatia, sendo que desta vez a relação é diretamente proporcional. Vamos esquematizar as diversas relações que existem entre o número de fatias a ser reproduzido, o tamanho de cada fatia e o tempo necessário para reproduzir cada fatia, da seguinte maneira:

Seja N = número de fatias;
 T = tamanho de cada fatia;
 R = tempo gasto na reprodução de cada fatia.

Parece-nos que as seguintes relações são verdadeiras:

- (a) N é inversamente proporcional a T ;
- (b) R é diretamente proporcional a T ; e
- (c) N é inversamente proporcional a R ;

Em outras palavras, o tempo gasto na reprodução de cada fatia é inversamente proporcional ao número de fatias presentes na MCT, num determinado tempo constante x . Podemos concluir que quanto maior o tempo gasto na emissão de cada fatia menor é o número de fatias que será possível reproduzir. Isso advém do fato de o tempo ser o tempo dos responsáveis pelo esquecimento das informações contidas na MCT. Esse fator tempo, embora fundamental, parece ter sido ignorado nas análises de Simon, o que o levou a resultados que aparentemente contradizem a idéia da rigidez da capacidade da MCT; daí também advém a idéia (que, como vimos, é infundada) de que há um limite potencial ao tamanho de cada fatia.

É importante observar ainda que o experimento de Simon foi efetuado com o objetivo de averiguar o que pode estar contido na MCT, mas as conclusões que dele se podem tirar se referem ao que é possível recuperar da MCT, o que é muito diferente. É possível que a MCT contenha sete fatias do tipo *all is fair in love and war or to be or not to be, that is the question*, mas é impossível verificar isso através de testes de recuperabilidade. A razão, como vimos, é que o tempo gasto na devolução de duas ou três dessas fatias faz com que as outras se percam. Esse fato invalida os resultados de Simon em termos de evidência quanto à capacidade da MCT.

5. Montagem de fatias tipo B

Do ponto de vista da percepção linguística, o primeiro tipo de fatia considerado — o tipo A — não é de grande utilidade porque, ao lermos um texto, muito raramente estamos recodificando informações já recebidas anteriormente, codificadas de maneira idêntica à situação presente. Na maioria das vezes, a maior unidade linguística em termos de fatias tipo A seriam as palavras individuais ou, ocasionalmente, pequenos sintagmas frequentes na língua (como, em inglês, *Milky Way*, um dos exemplos de Simon, ou, em português, *Via Láctea, usos e costumes, dos quatro costados, desastre ecológico* etc.).

No entanto, temos evidência de que, à semelhança do processo de fatiamento proposto por Miller, o leitor junta as palavras em grupos naturais na construção de uma interpretação do significado do texto. Se o leitor se utilizasse exclusivamente de informações literalmente presentes na MLT, construindo fatias tipo A, muito provavelmente cada palavra constituiria isoladamente uma única fatia. Como se explicaria, então, que é perfeitamente possível repetir uma frase de bem mais de sete palavras (embora não seja possível repetir uma sequência de bem mais de sete palavras, mas que não forme uma frase)?

Concluimos de imediato, a partir dessa observação, que cada palavra não completa obrigatoriamente uma fatia. Logo, elas devem agrupar-se, formando unidades maiores. Essas fatias não devem ser definidas arbitrariamente, mas muito provavelmente reúnem grupos de palavras que formam um todo coerente e com sentido.

A explicação do fato de se poder formar fatias "novas" só é possível se admitirmos a existência na MLT de outro tipo de material, além das informações literais. Na MLT não se de-

ve encontrar apenas informações sobre entidades já encontradas previamente, mas também esqueletos sintáticos, semânticos e discursivos (ou, mais exatamente, sistemas de regras que descrevem esses esqueletos), aos quais as palavras têm de se ajustar quando se decodifica um enunciado.

Já que o fatiamento linguístico não é aleatório, poderíamos inferir daí que o leitor ou ouvinte poderia também utilizar-se desses esqueletos para fatiar as sentenças, encaixando nelas as palavras reconhecidas, de maneira a formar unidades de significado. No entanto, as fatias resultantes terão necessariamente um status diferente das fatias tipo A, já que seriam **construídas** pelo leitor ou ouvinte, num processo que se pode descrever como ativo, e não apenas comparadas com elementos já presentes na MLT. Esse tipo de fatia (tipo B) contrasta claramente com o tipo A anteriormente examinado.

Voltemos agora à observação de Massaro & Schuller (1975) de que

"It may be [...] that the reason these fixation periods extend for 250 msec [durante a leitura normal] is that this is the length of time necessary to synthesize the information available within one fixation."

(p. 277)

Ora, a coincidência notada por eles é entre o tempo necessário para uma fixação durante a leitura corrente e o tempo necessário para reconhecer elementos (palavras, letras, algarismos) em experimentos taquioscópicos. Mas é preciso observar que durante a leitura corrente o leitor faz uso de fatias tipo B, tipicamente, e nos experimentos citados os sujeitos reconheciam fatias tipo A. É este último tipo de fatia que, segundo Massaro & Schuller, é sintetizado (isto é, transferido para a memória visual) em 250 msec. Na verdade, há evidência (Neisser & Beller, 1965) de que nesse espaço de tempo é possível reconhecer também semanticamente uma palavra. Neisser & Beller apuraram que os sujeitos eram capazes de decidir em aproximadamente 200 msec se as palavras de uma lista pertenciam ou não a determinado campo semântico (por exemplo, "nomes de animais"). Pode-se supor, pois, com alguma base que uma fatia tipo A requer aproximadamente 200 a 250 msec para seu processamento.

Consideremos as tarefas que um leitor precisa executar ao ler um texto cuja forma não está literalmente presente em sua MLT (essa é, evidentemente, a situação usual de um leitor, assim como de um ouvinte). O processamento de uma sen-

tença envolve o reconhecimento de uma série de elementos, sem dúvida, por exemplo, as palavras individuais, e possivelmente alguma expressão maior já feita, e como tal presente literalmente na MLT, como *Via Láctea* etc. Mas o reconhecimento desses elementos não esgota o trabalho de percepção da sentença — e aqui está a grande diferença entre a tarefa de decodificar sentenças e a de reconhecer elementos tais como os estudados por Simon (1974). Se o leitor reconheceu os elementos (digamos, as palavras) de

(1) uma onça pintada comeu minha avó a semana passada

resta ainda a tarefa, essencial, de processar a estruturação sintático-semântica desses elementos: para lembrar um velho chevão linguístico, uma sentença é mais do que a soma de suas palavras. Se formos definir "fatias" na sentença (1), digamos (apenas como exemplo) algo como

(2) / uma onça pintada / comeu minha avó / a semana passada/

não poderemos identificar tais fatias com as fatias tipo A, porque existe uma diferença importante do ponto de vista do processamento: fatias como as de (2) (tipo B) **não estão** presentes literalmente na MLT do leitor. Um leitor não pode "reconhecer" *comeu minha avó*, sintagma que ele possivelmente nunca encontrou antes, no mesmo sentido que reconhece *Via Láctea*. O fatiamento de *Via Láctea*, por assim dizer, já vem pronto, ao passo que o leitor tem de trabalhar para conseguir uma fatia como *comeu minha avó*.

Em outras palavras, o processamento de uma fatia tipo A, como *Via Láctea*, implica, essencialmente, numa comparação com um signo (provavelmente um conjunto criterial de traços) presente da MLT do leitor. Mas o processamento de uma fatia tipo B, como *comeu minha avó*, implica em algo mais do que isso. A própria determinação de *comeu minha avó* como uma fatia (e não, digamos, *pintada comeu*, ou *comeu minha*) demanda o reconhecimento dessa sequência como uma das realizações possíveis (dentro um número extremamente grande) da entidade abstrata SV; reconhece-se *comeu* como V, *minha* como Possessivo, *avó* como N, *minha avó* como SN, e finalmente *comeu minha avó* como V + SN = SV. É bastante plausível, poi, que o processamento de uma fatia tipo B seja uma tarefa diferente, e mais complexa, do que o processamento de uma fatia tipo A.

Se o processamento de uma fatia tipo B é realmente mais complexo do que o processamento de uma fatia tipo A, então será de esperar que leve inclusive mais tempo para ser completado. Parece-nos pouco plausível que a cada fixação corresponda a integração de uma nova fatia tipo B, pelo simples fato de que cada fixação dura o tempo necessário para processar uma fatia tipo A — uma conclusão que vai frontalmente de encontro à idéia de Massaro & Schmitter de que a coincidência de duração entre as fixações e o reconhecimento de um elemento (fatia tipo A, em nossa nomenclatura) revela que o tempo de fixação reflete o tempo de processamento do material colhido naquela fixação.

Acabamos de sugerir que não seria de se esperar que um leitor ou ouvinte conseguisse processar uma fatia tipo A no mesmo espaço de tempo que uma fatia tipo B, dado que a tarefa de processar a fatia tipo B é mais complexa, e deve presumivelmente levar mais tempo. Ora, suponhamos que (como propõem Bouma & de Voogd (1974) o tempo de fixação não seja reflexo direto do tempo de processamento de uma fatia; antes, trata-se apenas do tempo necessário para levar o material até o tanque. A montagem das fatias (tipo B) se iria processando paralelamente à colheita de novo material, sem sincronização necessária com a entrada desse material; ou seja, a entrada de novo material não precisa corresponder ao início de nova fatia. Essa nos parece uma consequência da posição de Bouma e de Voogd. Dessa maneira, o mecanismo de fatiamento poderia trabalhar com material proveniente de uma, duas ou mais fixações, material esse conservado no tanque para esse fim.

Lembremo-nos de que o R-2 tem como função interpretar o material léxico (talvez só, ou principalmente, as palavras de conteúdo; ver seção adiante), ao passo que o E&R "fecha" as fatias propriamente ditas. E que o tanque se coloca à altura da memória abstrata (rever a figura 1). Como consequência, o material jogado no tanque seria reconhecido perceptualmente e com os itens léxicos já interpretados.

Consideremos então o processamento de uma hipotética fatia / abcde /; digamos que ela corresponde ao material colhido em duas fixações sucessivas. Podemos então dividir o processamento dessa fatia em dois momentos, cada um dos quais correspondente a uma fixação, e com a duração aproximada de 200 a 250 msec. No primeiro momento, a fixação colhe /abc/, executa o processamento visual e léxico desse material e atira-o no tanque. No segundo momento, a fixação

colhe /de/ e igualmente o leva até o tanque, já percebido visualmente e interpretado lexicamente. Paralelamente, o E&R trabalha para fechar a fatia; esse trabalho pode se iniciar já no primeiro momento, com a formulação de hipóteses quanto à possível estrutura da fatia que se inicia; e se completa no segundo momento. O que é importante nessa maneira de descrever o processo é que a interpretação do material léxico é feita em 200-250 msec para cada fixação (o que está de acordo com os resultados de Neisser & Beller (1965); e, por outro lado, o E&R dispõe do dobro desse tempo, ou seja, a duração de suas fixações, para fazer a "montagem" da fatia tipo B. Esquematizando, teremos:

momento	fixação	R-2	no tanque	E&R	tempo
1	abc	a; b; c (itens)	abc	(hipóteses sobre a fatia)	200
2	de	d; e (itens)	abcde	/abcde/ (fatia)	200

— Figura 3 —

Parece-nos que esse modelo substitui com vantagens a concepção do processo subjacente à afirmação de Massaro & Schmitter, se é que com efeito eles admitem que o tempo de uma fixação é o tempo necessário para fechar uma fatia. Segundo essa concepção, inclusive, os resultados de Bouma & de Voogd são inteiramente inesperados, pois seria antes de crer que a cada fixação correspondesse uma fatia, e por conseguinte que a distribuição das fixações tivesse necessariamente de se subordinar a um critério linguístico. (Voltaremos adiante ao problema da montagem das fatias).

Até agora vimos que, embora tradicionalmente o termo "fatia" seja usado indiscriminadamente para qualquer tipo de material que constitua uma unidade dentro da MCT, essa noção engloba fatos distintos, que não se encontram diferenciados na literatura. Propusemos aqui a reformulação da postura tradicional, sugerindo a subdivisão da noção de "fatia" em dois tipos diferentes, o tipo A e o tipo B.

O que aqui chamamos "fatia tipo A" é a noção utilizada por Miller e testada por Simon. Segundo a definição original, as fatias são copiadas do material presente na MLT, que é levado de forma literal ao nível da MCT. Nesse tipo de fatia-

mento o processo é relativamente passivo, pois as fatias vêm já "prontas" da MLT. Argumentamos que, segundo essa concepção de fatia, qualquer material presente na MLT, e identificado como um todo, pode constituir uma fatia, independentemente de seu comprimento.

As fatias tipo B, características do processamento linguístico, são, ao contrário das fatias tipo A, "construídas" pelo leitor que, manipulando regras sintáticas ou semânticas guardadas na MLT, elaboram unidades sintática e semanticamente coerentes que, como tais, não estão elas próprias presentes previamente na MLT. A construção dessas unidades é relativamente "ativa", e consome, presumivelmente, mais tempo do que o reconhecimento de fatias tipo A.

6. Sobre as estratégias de previsão

A discussão precedente deixa bem clara, a nosso ver, a importância crucial da formação de fatias tipo B para a percepção linguística. É evidente que as fatias tipo A também são essenciais, pois são fatias tipo A as palavras, assim como as expressões do tipo *Via Láctea* etc. Mas o processamento das fatias tipo B apresenta problemas particulares, em geral ainda sem solução. Por exemplo, há o problema das hipóteses levantadas a respeito das fatias para nortear o seu fechamento (por exemplo, aquilo que o E&R desempenha no 1.º momento da figura 3). A formulação de tais hipóteses é certamente um ingrediente crucial do processo, pois só assim poderá o leitor evitar uma busca de todas as estruturas que podem ser geradas pelas regras que ele possui na MLT. O número dessas estruturas, com efeito, é elevadíssimo, havendo quem fale de um número ilimitado.

A existência de estratégias de pré-análise que permitam reduzir o campo de busca a apenas algumas dessas estruturas possíveis tem sido, portanto, reconhecida como essencial na literatura psicolinguística. Por exemplo, para citar um caso anatómico, Fodor, Bever & Garret (1974) propuseram uma estratégia do "sentóide canônico" (*canonical sentoid strategy*), segundo a qual os falantes, ao encontrarem uma sequência composta de $SN + V \times SN$, imediatamente tentarão pré-analisá-la como sendo o sujeito, o verbo e o objeto de uma sentença profunda. O que nos interessa aqui é a necessidade de tais estratégias, que tomam pistas dentro os elementos explicitamente presentes no enunciado, e a partir de tais pistas formulam hipóteses quanto à estruturação desses enunciados.

Os detalhes dessas estratégias são importantes, mas não são bem conhecidos, a despeito de alguns trabalhos já realizados. Clark & Clark (1977), por exemplo, mencionam um bom número de possíveis estratégias, mas a evidência citada em favor delas é certamente insuficiente. Sem pretender chegar ainda à validação experimental de alguma dessas estratégias, gostaríamos de fazer, a respeito de um dos aspectos das estratégias de Clark & Clark, alguns comentários que nos parecem úteis como norteamento parcial de futuras pesquisas. Desejamos frisar o caráter apriorístico das notas que se seguem, e a necessidade urgente de se realizarem experimentos que venham trazer mais evidência empírica à escassa fundamentação das hipóteses concernentes às estratégias de pré-análise.

Dentre as estratégias de pré-análise propostas por Clark & Clark selecionaremos dois grupos, a saber, o das estratégias baseadas no uso de palavras funcionais e o das baseadas no uso de palavras de conteúdo. A respeito do primeiro grupo, afirmam os autores que

"Function words — determiners, prepositions, conjunctions, pronouns, quantifiers and the like — may play a crucial role in the strategies for segmenting speech into constituents."

(Clark e Clark, 1977, p. 59)

De acordo com essa idéia, apresentam eles uma estratégia (originalmente proposta por Kimball (1973) nos seguintes termos:

"STRATEGY 1: Whenever you find function word, begin a new constituent larger than one word."

(p. 68)

Essa estratégia reflete a crença de Clark & Clark de que as palavras funcionais são pontos-chave na determinação dos constituintes de uma sentença. É preciso reconhecer que (em português tanto quanto em inglês) as palavras funcionais marcam, de fato, o início de constituintes maiores. Mas, evidentemente, isso não significa automaticamente que os leitores as utilizem da maneira proposta por Kimball; essa é ainda uma simples teoria, e teremos mais a dizer sobre sua plausibilidade mais adiante.

Os autores subdividem a Estratégia 1 em uma série de sub-estratégias, cuja função é especificar o uso dos diferentes tipos de palavras funcionais. Por exemplo,

"Strategy 1.1: Determiners and quantifiers
Whenever you find a determiner (a, an, the) or quantifier
(some, all, many, two, six, etc), begin a new noun phrase
(NP)."

(p. 39)

Para exemplificar o funcionamento dessa estratégia, tomemos um exemplo. Seja a frase

(3) as meninas estavam chegando quando nós saíamos

Vejamos como se processaria esse enunciado utilizando-se estratégias baseadas em palavras funcionais (deixamos claro que aqui se trata apenas de um exemplo, destinado a esclarecer melhor nossa concepção desse tipo de estratégia, e não uma proposta quanto ao que se passa literalmente na cabeça do leitor ao se defrontar com a frase (3). De acordo com a Estratégia 1 (e seus desdobramentos), o leitor ao encontrar um determinante deverá procurar um nome e tentar fechar um SN. Assim, em (3), ele procurará juntar *as* com *meninas* para formar o SN *as meninas*, que constituirá a primeira fatia de (3). É importante lembrar que o leitor partiu de *as*, percebido visualmente, e hipotetizou a existência de um nome imediatamente após esse determinante. Passando adiante, a identificação de um verbo auxiliar, *estavam*, permite ao leitor prever a ocorrência provável de um verbo principal, que fechará o SV, que será a segunda fatia. Por fim, ao encontrar uma conjunção subordinativa, *quando*, o leitor deve esperar uma oração imediatamente subsequente. Esse conjunto formará a terceira fatia, *quando nós saíamos*. Observe-se que esse tipo de processamento segue a Estratégia 1 de Clark & Clark, e se baseia no uso crucial das palavras funcionais na determinação das fatias.

Desejamos frisar mais uma vez que estratégias desse tipo são instrumentos presumivelmente utilizados pelo leitor para estabelecer hipóteses quanto ao que se vai seguir no texto. Conforme vimos anteriormente, esse processo de formulação de hipóteses permite que a montagem de uma fatia (tipo B) seja feita durante o tempo de várias fixações, colocando à disposição do E&R o prazo necessário para montar a fatia (ver figura 3). Sem esse processo de formulação de hipóteses, a montagem de uma fatia tipo B teria de ser feita toda no momento em que entra no tanque o material suficiente para fechá-la — em termos do esquema da figura 3, no 2.º momento. Consequentemente, teríamos de admitir que a percepção de uma fatia tipo B se faz toda (exceto algumas palavras individuais,

abc na figura 3) em aproximadamente 200 msec, o mesmo tempo necessário para uma fatia tipo A. Isso, como vimos, é pouco provável.

Essa é uma função importante do mecanismo de que fazem parte estratégias do tipo da Estratégia 1 acima, mas não é a única função. A existência de hipóteses prévias é também um fator essencial para a percepção do material visual subsequente, diminuindo a amplitude das possibilidades de ocorrência dos diversos itens e assim facilitando seu reconhecimento. É só dessa maneira que um leitor consegue ler 300, 400 ou 500 palavras por minuto, quando a velocidade máxima de reconhecimento de palavras isoladas (sem pistas prévias) é de cerca de 200 a 250 por minuto.

Ora, é evidente que se pode conceber uma estratégia baseada no uso não de palavras funcionais, mas de palavras de conteúdo. Clark & Clark também propõem estratégias desse tipo, sem comentários acerca da importância relativa dos dois tipos; parece que a preocupação desses autores é mais com a apresentação de hipóteses razoavelmente plausíveis, e não com a defesa de uma ou outra das hipóteses com base em dados empíricos. Na nossa opinião, entretanto, é importante investigar a importância relativa das diversas estratégias (seu "peso" no processo geral da percepção linguística). Deve haver uma diferença considerável entre as estratégias favoritas, aquelas que são usadas com maior frequência, e cuja aplicação o leitor provavelmente tenta inicialmente e, de outro lado, estratégias mais especializadas para casos relativamente raros. É de se supor que a aplicação das primeiras estratégias implique em menor dispêndio de esforço e menos tempo de processamento.

Vejamos agora um exemplo de estratégia baseada no uso de palavras de conteúdo:

"Strategy 8: Using content words alone, build propositions that make and parse the sentence into constituents accordingly."

(Clark & Clark, 1977, p. 75)

Nosso objetivo aqui é tentar mostrar que a percepção de um texto se beneficia de um uso extenso de estratégias baseadas em palavras de conteúdo; em outras palavras, desejamos propor que muito provavelmente as estratégias favoritas (usadas com maior frequência e tentadas em primeiro lugar) são predominantemente estratégias baseadas no uso de palavras de conteúdo. Isso se originaria do fato de que os leitores

"sabem"² que tais estratégias lhes poupam esforço perceptual e tempo, porto olegre, a a aprendizagem da leitura corrente poderá incluir a aprendizagem de um uso cada vez maior de estratégias desse tipo.

Para compreender a razão pela qual o uso de palavras de conteúdo como pistas básicas economiza esforço perceptual, observemos que nas línguas naturais as palavras de conteúdo são incomparavelmente mais numerosas do que as palavras funcionais. Veremos que dessa fato decorre que a ocorrência de uma determinada palavra de conteúdo em determinado contexto é muito menos previsível do que a ocorrência de uma determinada palavra funcional; portanto, um leitor dependerá de relativamente pouca informação visual quando utiliza estratégias baseadas em palavras de conteúdo. E, dependendo de pouca informação visual a extrair do texto, ele poderá ler mais eficientemente (por exemplo, mais rapidamente).

Tomemos, como um exemplo, os nomes e as preposições. Enquanto que, em português, há alguns milhares de nomes, o número das preposições é de no máximo duas ou três dezenas. Ora, se considerarmos, com Smith (1978), que a leitura é um processo de redução de incerteza, e que a magnitude da incerteza é uma função do número de alternativas entre as quais o leitor deve decidir, fica fácil ver como a decisão entre uma ou outra preposição é uma tarefa muito mais fácil do que a decisão entre um ou outro nome. Em outras palavras, quanto menor for o número das alternativas possíveis, mais rapidamente e com maior segurança poderá o leitor fazer a sua decisão. Esse é um fato suficientemente comprovado experimentalmente. Por exemplo, Gibson & Levin (1975) afirmam que, tanto para a compreensão oral quanto para a leitura, uma pessoa

"can identify words more accurately if he knows the list from which the words will be chosen, and the shorter the list, the more intelligible the words will be."

(Gibson & Levin, 1975, p. 214)

Vejamos agora como seria o processamento da sentença (3) segundo a hipótese do uso de estratégias baseadas em palavras de conteúdo.

(3) as meninas estavam chegando quando nós saíamos

2. Referimo-nos, é claro, a um conhecimento implícito, do qual os falantes não têm a menor consciência.

Vamos manter a suposição de que (3) se segmenta em três fatias: A primeira fatia seria

(3a) as meninas

Segundo a presente hipótese, o leitor reconhecerá (visualmente) a palavra de conteúdo *meninas*, e sua tarefa será prever a palavra funcional que a acompanha.

A primeira vista essa tarefa poderia parecer exatamente tão difícil de executar quanto a tarefa inversa — a de reconhecer palavras funcionais e a partir delas prever palavras de conteúdo — porque alguém poderia alegar que junto a um nome como *meninas* pode vir um artigo, um pronome, um quantificador etc. O leitor teria, então várias alternativas possíveis dentre as quais precisaria selecionar a correta.

Mas digamos que o leitor "sabe" que há uma probabilidade muito maior de aparecer o nome *meninas* precedido de artigo do que um artigo aparecer seguido do nome *meninas* (e não qualquer outro nome). Nesse caso, a incerteza a respeito da palavra funcional que pode estar acompanhando *meninas* seria reduzida a essa alternativa; a palavra funcional é um artigo, com toda a probabilidade³. Considerando-se esse fato, da maior frequência do artigo, e também o provável uso de informação visual periférica, que pode fornecer por exemplo o comprimento da palavra (duas letras), a incerteza do leitor sobre ela seria quase nula. Das alternativas possíveis restam apenas os artigos definidos *os*, *as* dos quais o primeiro é excluído em virtude do gênero de *meninas*. Como se vê, na verdade as alternativas foram significativamente reduzidas neste caso.

Tomemos, por outro lado, a estratégia inversa, baseada em palavras funcionais. Depois de perceber visualmente a palavra funcional *as*, o leitor deverá prever a palavra de conteúdo. Isso é impossível na prática, pois a incerteza é muito grande: como vimos, há milhares de nomes em português. A única maneira de se decidir pela alternativa correta da palavra (nome) que se segue ao artigo é utilizar informação visual relativamente detalhada. Em outras palavras, ao contrário do que

3. Esta afirmação pressupõe uma capacidade dos leitores de utilizar uma "sensibilidade à frequência" das diversas formas linguísticas para orientar hipóteses e decisões. Hasher & Chromiak (1977) apresentam evidência de que essa "sensibilidade" é um fato nos seres humanos.

parece à primeira vista, a tarefa de prever uma palavra funcional a partir de uma palavra de conteúdo é bem mais fácil do que a tarefa inversa.

Continuando a exemplificação, no caso das outras fatias que compõem (3) o raciocínio é semelhante. No constituinte

(3b) estavam chegando

depois de reconhecer o verbo principal **chegando**, o leitor terá boas chances de acertar na previsão de que o verbo auxiliar é **estavam**, pois esse auxiliar é a única alternativa sintática e semanticamente plausível nesse ambiente, depois de um (presumível) sujeito **as meninas** e antes de **chegando**, verbo que dificilmente aceitará outro auxiliar:

(4a)* eram chegando

(4b)* haviam chegando etc.

Já se admitíssemos a estratégia inversa, evidentemente, após o sujeito **as meninas** e o auxiliar **estavam** haveria uma variedade enorme de verbos admissíveis, como

(5a) estavam brincando

(5b) estavam estudando etc.

Por fim, tomemos o terceiro constituinte,

(3) quando nós saímos

e apliquemos primeiro a estratégia baseada em palavras de conteúdo. Após reconhecer o verbo **saiamos**, o leitor terá relativamente pouca dificuldade em prever **quando**, pois as alternativas, aqui também, são muito poucas. Só teríamos as seguintes possibilidades: **depois que**, **logo que**, e, **mas** e talvez mais uma ou duas.

Já se o leitor começa por reconhecer a conjunção **quando**, permanece uma incerteza muito grande quanto ao verbo da oração subsequente⁴, pois pode ser qualquer um de uma lista de milhares:

4. Neste exemplo específico desprezamos o caso do sujeito *nós*, porque ele não precisará ser reconhecido separadamente: é totalmente previsível a partir da forma verbal *saiamos*. Naturalmente, em outros casos a identificação do sujeito é uma etapa importante do processo.

(6a) quando nós dançávamos

(6b) quando nós cantávamos

(6c) quando nós dormíamos etc.

Mais uma vez o número de alternativas faz com que a tarefa do leitor se torne especialmente difícil.

Sumariando, afirmamos que há uma probabilidade alta de que bons leitores se utilizem preferencialmente de estratégias de pré-análise baseadas na identificação visual de palavras de conteúdo, com as palavras funcionais identificadas com o auxílio de previsões (o que permite que sejam apenas parcialmente processadas visualmente, ou mesmo que não sejam processadas visualmente). Já as estratégias que utilizam palavras funcionais como pistas básicas desempenham um papel secundário.

A partir das considerações feitas acima, propomos que o processo de fatiamento dos enunciados se faça em três estágios:

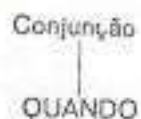
- identificação da pista (em geral, palavra de conteúdo);
- previsão esquelética a partir das pistas disponíveis;
- preenchimento do esqueleto com material sintetizado:

parcialmente percebido ou construído com ajuda de conhecimento armazenado.

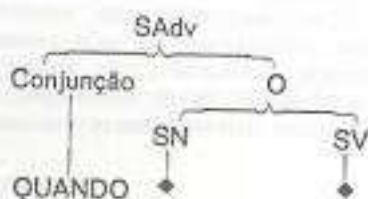
A distinção do processo de formação de fatias em etapas não é usual. Entretanto, parece-nos difícil escapar de uma noção como essa, em que inclusive é possível a identificação da pista preceda temporalmente a previsão esquelética, e esta o preenchimento. A ausência da distinção em etapas na literatura que conhecemos deve ser, acreditamos, mais resultado de omissão do que recusa em admitir o caráter complexo do fatiamento.

Voltando ao esquema (a — c) acima, é oportuno lembrar que ambas as hipóteses vistas anteriormente (uso de palavras de conteúdo ou uso de palavras funcionais como pistas principais) terão de incluir essa subdivisão do processo em etapas. A diferença entre as duas posições está na natureza das pistas utilizadas na etapa (a), e consequentemente no tipo de previsão efetuada na etapa (b). Esquemáticamente, o processo de fatiamento funcionará, segundo cada hipótese, como está nas figuras 4 e 5 abaixo.

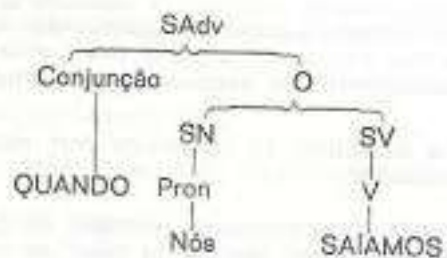
a) identificação da pista:



b) previsão esquelética:



c) preenchimento:



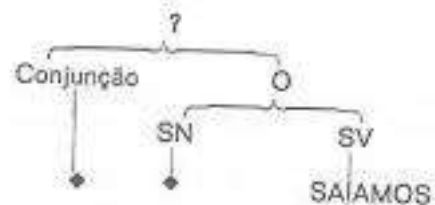
— Figura 4 —

O processo de fatiamento, segundo a hipótese do uso de palavras funcionais como pistas

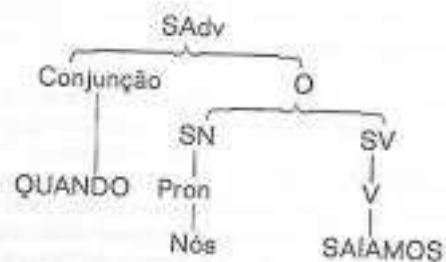
a) identificação da pista:



b) previsão esquelética:



c) preenchimento:



— Figura 5 —

O processo de fatiamento, segundo a hipótese do uso de palavras de conteúdo como pistas

As figuras acima deixam de exprimir um aspecto da maior importância, a saber, o fato de que preencher o lugar da conjunção a partir da previsão esquelética feita no segundo estágio (como está na figura 5) envolve menos trabalho do que preencher o lugar do verbo a partir do esqueleto que inclui apenas a conjunção (figura 4). A estratégia esquematizada na figura 5 possibilita ao leitor um uso mais eficiente da informação não visual (seu conhecimento da língua), consequentemente aliviando a carga perceptual necessária ao processamento da sentença. Para ser mais específicos: o preenchimento (estágio (c) no caso da figura 4 exige a obtenção de muito mais informação visual do que o preenchimento ilustrado na figura 5.

Das duas hipóteses que acabamos de discutir, argumentamos que aquela que sustenta que as pistas principais utilizadas durante a leitura são as palavras de conteúdo é a mais plausível, "a priori", por envolver menor esforço do mecanismo de processamento e, portanto, possibilitar um uso mais eficiente da informação não visual disponível ao leitor. Essa hipótese é, acreditamos, passível de comprovação empírica; e, sem pretensões de liquidar o problema, gostaríamos de apontar uma previsão decorrente dessa hipótese que nos parece relativamente fácil de levar a testagem experimental.

Consideremos dois fenômenos corriqueiros em situações de leitura: erros de leitura oral e não-percepção de erros de imprensa. As duas hipóteses em discussão fazem previsões opostas quanto ao modo pelo qual se manifestam esses dois tipos de erros. Segundo a hipótese do uso preferencial das palavras funcionais (rejeitada nas páginas precedentes), seria de esperar que a maior parte dos erros de leitura oral e de não-percepção de erros de imprensa envolvessem as palavras de conteúdo. Isso porque as palavras funcionais, segundo essa hipótese, são necessariamente percebidas com maior perfeição; as palavras de conteúdo serão, em parte, supridas através do uso da informação não-visual (redundância nos diversos níveis linguísticos e não-linguísticos)⁵. Já segundo a hipótese do uso preferencial das palavras de conteúdo como pistas, será de esperar que os erros se concentrem nas palavras funcionais (que são supridas em parte pelo uso da redundância). Em outras palavras, segundo a primeira hipótese um sintagma como

⁵ Observe-se aliás, que neste caso é implausível a hipótese de que o leitor procede inteiramente de informação visual. Ele terá que lançar mão, pelo menos, de informação visual incompletamente captada.

(7) uma casa no lago

poderia ser ocasionalmente lido como

(8) uma caça no lago

Mas a segunda hipótese preveria antes um erro como

(9) a casa do lago

Acreditamos que os fatos favoreçam a segunda hipótese, e certamente deve ser possível planejar um experimento para comprovar essa nossa impressão. Aliás, a seguinte passagem de Wanat (1976) parece indicar que é mais do que uma simples impressão:

"A clear indication that proficient readers attend to less than the available information in word recognition and in the reading of sentences is the phenomenon of 'proofreader's errors'. Very proficient readers can read materials without detecting errors such as letter substitutions, transpositions, omissions, or additions. [...] Function words can be repeated (e.g., 'the the') or omitted entirely without the reader noticing the error."

(Wanat, 1976, p. 110; grifo nosso)

Não é nosso objetivo aqui resolver o problema de uma vez por todas; bastará dizer que temos razões (acima mencionadas) para esperar que a verificação empírica corroborará a hipótese de que as pistas principais utilizadas pelo leitor durante a leitura são as palavras de conteúdo — ficando a estratégia baseada em palavras funcionais, na melhor das hipóteses; como um recurso auxiliar a ser empregado em certos casos.

7. Palavras funcionais e palavras de conteúdo

Antes de passar adiante, porém, parece-nos importante fazer algumas observações acerca das noções de "palavra de conteúdo" e "palavra funcional". Os autores que utilizam essas categorias não se têm preocupado com sua definição rigorosa, contentando-se com uma noção intuitiva e algo vaga, remanescente talvez da ideia tradicional de palavras "com" e "sem" conteúdo, ou com conteúdo "concreto" versus "gramatical, relacional". Acreditamos, entretanto, que é possível refinar essa noção, que como vimos é importante para o estudo do fenômeno da compreensão.

Os termos "funcional" e "de conteúdo" são enganadores, pois dão a impressão de que, por um lado, a dicotomia se baseia em um critério semântico mais ou menos puro e, por outro lado, de que se trata de uma divisão nítida e discreta entre duas classes (como é, por exemplo, a divisão entre substantivos masculinos e femininos). Mas a discussão feita nas páginas precedentes, das estratégias de construção de frases, deve ter deixado claro que a verdadeira oposição se prende a uma dimensão de **previsibilidade**. Para voltar a um exemplo anterior, vimos que, dado o pequeno número de preposições existentes em português, é tarefa relativamente fácil identificar uma preposição determinada, ao passo que é muito mais difícil identificar um substantivo determinado (isso em um contexto que nos permita prever qual o tipo de palavra que deve ocorrer). Dissemos então que a magnitude da incerteza (e, podemos acrescentar, a dificuldade da tarefa de reduzi-la) é uma função do número de alternativas entre as quais o leitor deve decidir.

Não excluimos a possibilidade de haver uma relação entre a previsibilidade de uma palavra em determinado contexto e algo que possa ser chamado de "quantidade de significado". Mas esta última noção é, no mínimo, vaga demais para ser útil por ora, ou mesmo para que se possa investigar melhor seu relacionamento com a previsibilidade. Preferimos, portanto, trabalhar com a noção que nos parece mais clara.

Previsibilidade é um fator que depende do contexto e das circunstâncias. Temos condições de prever a ocorrência frequente do nome de D. Pedro I em um texto sobre a Independência, e de Getúlio Vargas em um texto sobre a Revolução de 1930, mas não vice-versa: a previsibilidade desses nomes nesses contextos é função não apenas da semântica dos textos em que se inserem mas também, é claro, dos conhecimentos históricos do leitor. Além disso, a previsibilidade é uma questão de grau: não há palavras previsíveis em oposição a imprevisíveis em um contexto dado, mas apenas palavras mais previsíveis e menos previsíveis. Voltando ao exemplo de um texto sobre a Independência do Brasil, pode-se dizer que a ocorrência de D. Pedro I é mais de se esperar (portanto, "mais previsível") do que a de D. João VI, mas a deste é mais de se esperar do que a de D. Carlota Joaquina, e a de qualquer desses é mais de se esperar do que a de Ronald Reagan. Novamente, essa gradação depende dos conhecimentos extra-linguísticos do leitor. Um leitor inteiramente desinformado sobre a história da Independência pode muito bem não distinguir entre D. Pedro I e Ronald Reagan em termos de elaboração de previsões.

As considerações feitas no parágrafo anterior parecem deixar em um impasse qualquer tentativa de distinguir classes de palavras "altamente previsíveis" frente a outras não tão previsíveis. Se a previsibilidade de uma palavra depende do texto em que se insere e dos conhecimentos prévios do leitor, como se poderia tomar uma palavra (digamos, uma preposição) e dizer que ela é "mais previsível" do que outra (digamos, um substantivo), isso sem considerar seu contexto e a pessoa que a lê? Argumentaremos, entretanto, que esse é um impasse apenas aparente, e que é perfeitamente possível distinguir graus de previsibilidade das palavras independentemente de contexto e leitor.

O tipo de previsibilidade de que falamos logo acima, como o que se verifica no caso de um texto sobre a Independência, mas já não se verifica em um texto sobre a Revolução de 1930, é totalmente acidental. Nesses casos, a palavra *x* é altamente previsível apenas por se encontrar no contexto *y* e por ser lida pelo leitor *z*, que possui determinados conhecimentos extra-linguísticos. Ninguém teria a idéia de colocar D. Pedro I ao lado do artigo *a*, ou da preposição *em*, como um dos itens mais altamente previsíveis da língua portuguesa. Ora, é preciso então perguntar o que é que *a*, *em*, têm que D. Pedro I não tem.

A diferença, obviamente, é que *a* e *em* são altamente previsíveis em **qualquer** texto da língua portuguesa, lido por **qualquer** leitor que domine essa língua (independentemente de seus conhecimentos sobre o assunto do texto). Exemplificando, em um texto sobre a Independência lido por um leitor informado do assunto, D. Pedro I é altamente previsível no contexto *João Bonifácio correu a avisar...* mas em outro texto, sobre assunto, nada permite ao leitor prever a ocorrência de D. Pedro I logo após, digamos, *em minha cidade tem uma avenida chamada Avenida...* Por outro lado, no contexto *a capital — Portugal é Lisboa* a preposição *de* é altamente previsível qualquer que seja o assunto do texto, e quaisquer que sejam os conhecimentos extra-linguísticos do leitor (desde, é claro, que ele conheça suficientemente bem a língua portuguesa). Ou seja, a alta previsibilidade da ocorrência da preposição *de*, ou de outra preposição qualquer, no contexto *x* não é acidental, decorrente do assunto do texto ou dos conhecimentos do leitor; é na verdade uma decorrência das próprias estruturas da língua, mais o conhecimento implícito que tem qualquer leitor da plausibilidade de ocorrência de cada uma dessas estruturas em qualquer texto. No caso de *a capital — Portugal* poderia, em princípio,

ocorrer em, por exemplo mas um leitor "sabe" que essa possibilidade é muito remota, de maneira que prefere arriscar um *de*, e normalmente acertará em sua previsão.

Há muito o que dizer sobre esse uso que o leitor faz de seu conhecimento linguístico para prever a ocorrência desta ou daquela palavra. Aqui não será possível discutir o assunto em detalhe, mas gostaríamos de deixar claro que não estamos na ilusão de que se trata de um processo simples e transparente. Por exemplo, mesmo com todo o seu conhecimento das estruturas da língua e de sua frequência relativa, um leitor pode encontrar casos em que tal conhecimento é insuficiente para decidir qual é a preposição que deve existir em determinado local; um livro sobre a mesa ou um livro sob a mesa são, em princípio, igualmente admissíveis. Ai o leitor, certamente, põe em jogo seus conhecimentos extra-linguísticos para resolver o problema. Ele pode considerar (implicitamente) que é raro a gente colocar um livro *sob* a mesa, e que portanto a preposição deve ser *sobre*, desse modo utilizando um procedimento que é mais frequente identificação de substantivos do que na identificação de preposições. Queremos dizer apenas que o uso de informação estritamente linguística (estrutural) é típico do processo de identificação de preposições, e o uso de informação extra-linguística é típico do processo de identificação de substantivos.

8. Evidência em favor do uso de palavras de conteúdo

Retornemos agora aos exemplos hipotéticos (7) e (9). Dissemos acima que um erro típico de leitura oral seria ler (9) no lugar de (7):

- (7) uma casa no largo
- (9) a casa do largo

isto é, com certa frequência o leitor substituiria uma palavra funcional por outra, mantendo inalteradas as palavras de conteúdo. Admitamos que os fatos são realmente esses, e consideremos a representação dos mesmos em termos do modelo de leitura aqui adotado.

A montagem das fatias, conforme já vimos, é função do componente a que Massaro (1975) chama *recoding and rehearsal* — aqui, *ensaio e recodificação*, ou E&R. Para compreender melhor o papel do E&R na geração de erros como o de ler (9) em vez de (7), teremos que propor um esboço da estrutura interna desse processo.

A ocorrência do erro ilustrado em (7) e (9) é evidente de que nesse caso as palavras funcionais não foram percebidas, isto é, não sofreram a aplicação do R-2 e não deram entrada no tanque. As palavras *a* e *do* de (9) devem ter sido, portanto, sintetizadas pelo E&R durante a montagem da fatia, isso com base na informação sintática e semântica disponível a partir das palavras de conteúdo, de sua ordenação, e do contexto anterior. Enfocando esse processo segundo o esquema ilustrado na figura 5, diríamos que o preenchimento a partir de informação visual (ainda que parcial) pode ser, em muitos casos, inteiramente dispensado. Nesses casos, o sistema introduz itens hipotéticos, completando a montagem da fatia sem necessidade de recurso à informação visual.

Tal é possível, naturalmente, em razão da redundância da expressão linguística. Por exemplo, existe frequentemente mais de uma fonte disponível de informação para se decidir a respeito da identidade de uma palavra funcional: por um lado, o sinal gráfico; e, por outro lado, o nosso conhecimento das palavras de conteúdo circunvizinhas, mas as regras sintáticas da língua, mais, se for o caso, o contexto (semântico) anterior. Ilustrando, se o leitor tiver captado visualmente o seguinte material:

- (10) — cachorro — vovó

não lhe será difícil prever que o sintagma é

- (11) o cachorro da vovó

dado que as alternativas são muito poucas. Na leitura corrente, a inserção da fatia resultante (ou melhor, de sua face semântica) no contexto geral fornece um modo de confirmar a previsão feita.

Isso significa que o E&R deverá incluir, entre outros, os seguintes subcomponentes: (a) Um mecanismo de previsão (MP), cuja função é construir a previsão esquelética a partir da informação visual disponível, mais a informação não-visual existente na MLT (regras gramaticais da língua, contexto anterior). (b) Um mecanismo tapa-buraco (MTB), que em certos casos preenche nódulos terminais do esqueleto com itens léxicos (o mais das vezes palavras funcionais) sintetizadas pelo sistema, sem nenhum ou quase nenhum recurso à informação visual.

Como se vê, deixamos aberta a possibilidade de o MTB colocar também palavras de conteúdo. Isso deve ser menos frequente em leitores experientes (naturalmente, como decorren-

cepção do verbo colocar permite ao leitor prever a provável ocorrência de três SNs: um agente, um paciente e um lugar (como em a empregada colocou o queijo na geladeira). Como é evidente, o mais provável é que esta estratégia seja utilizada ao lado da estratégia baseada na silhueta gráfica das palavras como recursos complementares para a pré-identificação gráfica das palavras como recursos complementares para a pré-identificação de palavras funcionais.

REFERÊNCIAS BIBLIOGRÁFICAS

- ATEKINSON, R. L. & SHIFFRIN, R. M. (1968) "Human memory: a proposed system and its control processes", in K. & J. Spence, ed., *The Psychology of Learning and Motivation*, N. York, Academic Press.
- CLARK, H. H. & CLARK, E. V. (1977) *Psychology and Language*, N. York, Harcourt Brace Jovanovich.
- BOUMA, H. & DE VOGD, A. H. (1974) "On the control of eye saccades in reading" *Vision Research*, 14.
- FODOR, J. A., BEVER, T. G. & GARRET, M. P. (1974) *The Psychology of Reading*, N. York, McGraw-Hill.
- GIBSON, E. J. & LEVIN, H. (1975) *The Psychology of Reading*, Cambridge, MIT Press.
- GOODMAN, K. S. (1970) "Psycholinguistic universals in the reading process", *Journal of Typographic Research*, 4.
- HANSEN, D. & ROGERK T. (1968) "An exploration of psycholinguistic units in initial reading" in K. Goodman, ed., *The Psycholinguistic Nature of the Reading Process*, Detroit, Wayne St. Univ Press.
- KIMBALL, J. P. (1973) "Seven principles of surface structure parsing in natural language" *Cognition*, 2.
- MASSARO, D. W. (ed.) (1975) *Understanding Language*, N. York, Academic Press.
- MASSARO, D. W. & SCHMULLER, J. (1975) "Visual features, preperceptual storage, and processing time in reading" in Massaro, ed.
- MILLER, G. A. (1965) "The magical number seven, plus or minus two: some limits on our capacity for processing information" *Psychological Review*, 63.
- NEISSER, U. & BELLER, H. K. (1965) "Searching through word lists" *British Journal of Psychology*, 56.
- SHEELESKE, W. (1975) "Reading eye movements from an information-processing point of view" in Massaro, ed.
- SIMON, H. (1974) "How big is a chunk?" *Science*, 183.
- SMITH, P. (1978) *Understanding Reading*, N. York, Holt.
- WANAT, S. F. (1976) "Relations between language and visual processing" in *Theoretical Models and Processes of Reading*, Newark, International Reading Assoc.