

O Teste de Cloze na Avaliação da Compreensão em Leitura

Acácia A. Angeli dos Santos¹

Ricardo Primi

Fernanda de O. S. Taxa

Claudette M. M. Vendramini

Universidade São Francisco, São Paulo

Resumo

Este trabalho investigou as propriedades psicométricas de um instrumento de avaliação da compreensão em leitura utilizando a técnica de Cloze. Participaram 612 estudantes ingressantes de quatro cursos de uma universidade particular, cuja tarefa era preencher a lacuna do texto com a palavra que julgassem mais apropriada. Observou-se um bom ajuste ao modelo de dois parâmetros da Teoria de Resposta ao Item, sendo que a escala formada pela soma do número de lacunas corretamente completadas provou possuir consistência interna alta ($KR-20 = 0,84$). Os itens formaram dois grupos em razão da dificuldade: um grupo com itens muito fáceis e outro com itens muito difíceis. Parte da variabilidade dos índices de dificuldade pôde ser explicada pela classe gramatical da palavra a ser descoberta ($r=0,62$, $p < 0,001$). Do maior ao menor nível de complexidade foram observados: adjetivos, verbos, advérbios, substantivos, pronomes, preposição e artigo. Analisando-se os processos cognitivos que poderiam explicar as diferenças de dificuldade correlacionadas com as classes gramaticais e aliando-se a aplicação da Teoria de Resposta ao Item (TRI), discutiu-se o avanço das interpretações de proficiência em leitura referenciadas nos itens.

Palavras-chave: Compreensão em leitura; técnica de Cloze; Teoria de Resposta ao Item.

The Cloze Teste on the Reading Comprehension Assessment

Abstract

This research studied the psychometric properties of a reading comprehension test using the Cloze's technique. The participants were 612 freshmen undergraduate students whose task was to discover the omitted words. It was observed a good fit to a two parameter model of Item Response Theory. The total score composed by the sum of the correct completions achieved high level of reliability ($KR-20 = .84$). Considering the item complexities, two groups were formed: one composed of very easy items, and the other with very complex ones. A parcel of item complexity was explained by the grammatical category of the word to be discovered ($r = 0.62$, $p < 0.001$). From the highest to the lowest difficulty level: adjectives, verbs, adverbs, nouns, pronouns, prepositions, articles were observed. The cognitive processes that could explain the differences in the difficulties correlated with grammatical categories were analyzed. Through this analysis and through Item Response Theory (IRT) the advances of item referenced interpretation in describing reading comprehension proficiency were discussed.

Keywords: Reading comprehension; Cloze's Technique; Item Response Theory.

A leitura representa ferramenta fundamental para a formação social e cognitiva do sujeito, o que o qualifica para sua inserção na cultura. A habilidade de leitura ocupa papel importante na vida humana, em especial no sistema escolar, que tem como um de seus principais objetivos ensinar conceitos por meio de práticas que requerem habilidades de leitura. Embora o processo de alfabetização, durante as séries iniciais do ensino fundamental, tenha sido objeto de preocupação de pesquisadores de diferentes áreas,

há poucos estudos enfocando a prática da leitura e compreensão de estudantes do ensino médio e superior.

O ato de ler será aqui compreendido como um processo, no qual a interpretação do que é lido depende, não só do que está impresso, mas também das hipóteses do próprio leitor, formuladas com base no seu conhecimento prévio, e do estabelecimento de conexões intertextuais que permitem a leitura significativa (Smith, 1978, 1997).

A leitura na universidade é enfatizada por Witter (1990) como um dos caminhos que levam o aluno a ter acesso à produção científica e ressalta a importância de uma leitura crítica por parte do estudante, de modo a recuperar a informação acumulada historicamente, utilizando-a para uma prática profissional eficiente.

Além disso, Santos (1997) assinala que a Universidade quase sempre se constitui na última oportunidade formal de ensino que pode garantir ao aluno a remediação e o desenvolvimento do hábito de leitura e de compreensão

¹ Este artigo corresponde a parte dos dados obtidos por meio do projeto de pesquisa: "Avaliação das habilidades básicas de alunos ingressantes" subsidiado pelo programa institucional de apoio à pesquisa da Universidade São Francisco.

² Endereço para correspondência: Universidade São Francisco, Laboratório de Avaliação Psicológica e Educacional, LabAPE, Programa de Estudos Pós-Graduados em Psicologia, Rua Alexandre Rodrigues Barbosa, 45, 13251-900, Itatiba, SP. E-mail: acacia@saofrancisco.edu.br

de textos, indispensáveis ao profissional que ela se propõe a formar. Demonstra que esse é um trabalho possível ao desenvolver um estudo com 44 alunos ingressantes, divididos em grupo experimental ($n=22$) e de controle ($n=22$), que se submeteram a um programa de remediação em leitura e estratégias de estudo. Os resultados desse estudo não demonstraram superioridade do grupo experimental na compreensão em leitura, mas evidenciaram aumentos significativos no desempenho acadêmico e melhoras qualitativas nas estratégias de aprendizagem utilizadas.

Diante da importância da leitura, e independentemente da concepção de compreensão adotada, considera-se fundamental o diagnóstico da habilidade de leitura dos alunos para que se possa identificar seus limites, bem como seu potencial, posto que é principalmente por meio da leitura que ocorre o acesso ao conteúdo das diversas disciplinas.

O Teste de Cloze

Um dos procedimentos para avaliação da compreensão em leitura surgido das pesquisas de Taylor (1953) é denominado Técnica de Cloze. Consiste na seleção de um texto de aproximadamente 200 vocábulos, do qual, na proposta original do autor, omite-se o quinto vocábulo, como forma mais adequada para o diagnóstico da compreensão. Os examinandos devem preencher a lacuna com a palavra que julgarem ser a mais apropriada para a constituição de uma mensagem coerente e compreensiva. Os escores são obtidos somando-se os números de lacunas preenchidas corretamente.

Essa técnica tem se mostrado bastante eficaz, tanto do ponto de vista prático, tendo em vista a facilidade de elaboração, aplicação e correção, bem como do ponto de vista empírico, em função dos altos índices de correlação positiva de seus resultados com o desempenho acadêmico; isto é, alunos com maiores percentuais no teste apresentam melhores resultados nas médias das disciplinas (Marini, 1986; Santos, 1990).

Um dos aspectos mais importantes de qualquer instrumento de avaliação refere-se ao significado dos escores, isto é, à interpretação atribuída aos diferentes níveis de desempenho. Tradicionalmente, a inteligibilidade dos resultados pode ser conferida por três procedimentos nomeadamente: referência à norma, referência ao conteúdo, referência ao critério (Cronbach, 1996).

A referência à norma confere significado, comparando-se os escores obtidos por um sujeito com os escores obtidos por um grupo de referência (grupo normativo) e indicando a posição relativa desse escore frente ao grupo. A referência ao conteúdo é utilizada quando o conjunto de problemas presentes no instrumento pode ser considerado uma amostra

representativa do universo de problemas de um determinado conteúdo ou domínio. Nessas condições interpreta-se o escore nas tarefas (amostra) diretamente como uma estimativa do escore que o sujeito teria se respondesse a todos os problemas daquele determinado domínio (população). A referência ao critério confere significado ao escore relacionando-o a alguma outra medida que se deseja prever, chamada critério externo. Se existe uma correlação significativa entre as duas medidas pode-se conferir significado ao primeiro escore, indicando, para cada nível, qual a expectativa de desempenho no critério externo.

Tradicionalmente, o significado dos escores do teste de Cloze é conferido seguindo-se o último procedimento (referência ao critério). Seguindo-o, Bormuth (1968) elaborou parâmetros para analisar o desempenho dos sujeitos apresentando três níveis de leitura. O nível de frustração, correspondente ao percentual de acerto de até 44% do total do texto, indica que o leitor conseguiu retirar poucas informações da leitura e, conseqüentemente, obteve pouco êxito na compreensão. O nível instrucional, que corresponde a um percentual de acertos entre 44% a 57% do texto, mostra que a compreensão da leitura é suficiente, porém indica a necessidade de auxílio adicional externo (do professor, por exemplo). Por fim, o nível independente, que corresponde a um rendimento superior a 57% de acertos no texto, equivale a um nível de autonomia de compreensão do leitor.

Entretanto, como a proporção de acertos não depende simplesmente da habilidade dos sujeitos, mas também da dificuldade das lacunas criadas, há o problema da equivalência de notas. Por exemplo, duas proporções de acerto iguais, obtidas em dois testes de Cloze nos quais as dificuldades das lacunas sejam muito diferentes, não são diretamente comparáveis. Cohen (1975), Page (1975) e Santos (1981) encontraram diferenças de desempenho no teste de Cloze, relacionadas com o assunto abordado no texto, o que leva a crer que a familiaridade com o assunto aumenta a probabilidade de acerto das respostas dadas no Cloze, introduzindo uma variável interveniente na comparação dos desempenhos. Uma implicação radical é que a aplicação dos parâmetros de Bormuth (1968) para novos textos só pode ser feita se estes textos tiverem um nível de dificuldade semelhante ao dos textos que Bormuth utilizou em seu estudo.

Esta interdependência entre os escores e a dificuldade dos testes é uma das limitações dos modelos da psicometria clássica, na qual os escores são calculados atribuindo o mesmo peso para itens heterogêneos quanto à complexidade. Essa e outras limitações motivaram o desenvolvimento de um novo modelo chamado Teoria

de Resposta ao Item-TRI (Hambleton, Swaminatham & Rogers, 1991).

A Teoria de Resposta ao Item

A TRI propõe um modelo matemático que formaliza a relação entre os elementos essenciais da situação na qual uma pessoa responde a um problema. Nessa situação, quanto maior a habilidade da pessoa na modalidade requerida pelo problema, maior será a probabilidade que ela responda corretamente. Por outro lado, sendo a habilidade constante, quanto maior for a dificuldade do problema, menor será a probabilidade de que ela o acerte. O modelo matemático representa essa situação por meio da curva característica do item, que indica a probabilidade de acerto em função da habilidade das pessoas que o respondem e da dificuldade do problema. Dependendo do modelo, podem ser incluídas outras características do item, como a discriminação e as probabilidades de acerto ao acaso³.

Em uma situação de avaliação, aplica-se um conjunto de itens previamente calibrados, isto é, com parâmetros conhecidos, a uma pessoa cuja habilidade se pretende conhecer. Após a correção das respostas, tem-se duas informações: a probabilidade de acerto, isto é, o padrão de acertos e erros nos itens aplicados, e as dificuldades desses itens (obtidas previamente nos estudos de calibração). Uma das formas de atribuição do escore ao sujeito é feita comparando-se o perfil de acertos de uma pessoa com a dificuldade dos problemas respondidos, para atribuição de um valor numérico (θ). Este, por sua vez, reproduz o perfil de acertos da maneira mais semelhante possível. Em síntese, esse procedimento encontra o valor da habilidade mais condizente com o padrão observado de respostas, considerando, para isso, a dificuldade dos itens.

Assim, por meio da equação da curva característica do item estabelece-se uma relação que tem, de um lado, a probabilidade de acerto ($P_{ij}(\theta)$), e do outro, uma comparação entre habilidade do sujeito e dificuldade do item ($\theta - b_i$). Nesse processo de comparação, se o sujeito acertou o item é porque sua habilidade excedeu à dificuldade do item ($\theta > b_i$). Reversamente, se errou, é porque sua habilidade foi inferior à dificuldade do item ($\theta < b_i$). A estimação da habilidade é feita considerando a probabilidade, acerto ou erro, e a dificuldade do item, e encontrando o valor mais plausível da habilidade que equilibre essa relação.

Esse procedimento supera a limitação apontada da interdependência entre os escores e a dificuldade dos itens pois, diferentemente do procedimento clássico, no cálculo dos escores atribui-se pesos diferenciados dependendo da dificuldade dos itens. Devido a esse procedimento, os escores resultantes independem dos itens que compõem o teste, desde que todos os itens utilizados meçam uma mesma habilidade (sobre o postulado da unidimensionalidade, ver Hambleton & cols., 1991).

Um dos avanços mais importantes decorrentes da TRI refere-se às possibilidades de interpretação atribuídas aos escores. Como apontou Embretson (1996), quando é possível explicar por que os itens vão ficando cada vez mais difíceis, referindo-se aos processos cognitivos subjacentes requeridos por eles, pode-se estabelecer interpretações sobre o desempenho de uma pessoa, caracterizando-o diretamente com afirmações sobre o processamento cognitivo.

Esse avanço fundamenta-se em uma característica importante da TRI, que é a disposição dos valores numéricos dos escores e dos índices de dificuldade do item em uma escala comum. Disso emerge uma relação interessante: quando uma pessoa tem habilidade igual ao índice de dificuldade do item, as chances são de 50% que ela o acerte. À medida em que sua habilidade aumente em relação à dificuldade do item, suas chances de acertá-lo serão maiores do que 50%. À medida que sua habilidade seja menor do que a dificuldade do item, suas chances de acertá-lo serão menores do que 50%. Portanto, tendo-se em mãos o escore de uma pessoa, pode-se prever quais itens ela terá mais chances de acertar ou errar, informando-se o domínio que a pessoa possui do que foi avaliado.

A escala comum permite o estabelecimento de significados mais objetivos sobre quais itens o sujeito terá mais chance de acertar, isto é, aqueles com dificuldade inferior à sua habilidade. Se os diferentes níveis de dificuldade estiverem vinculados a uma definição das operações cognitivas envolvidas ou do conteúdo da tarefa, as informações obtidas na avaliação se tornam qualitativamente mais refinadas.

Em síntese, como o escore indica em quais itens a pessoa teve mais sucesso e simultaneamente a teoria cognitiva explica por que os itens são difíceis, pode-se generalizar as explicações cognitivas ao desempenho dos sujeitos. Tal procedimento foi denominado por Carrol (1993, 1997) de escalonamento comportamental.

A Complexidade de Avaliação dos Testes de Cloze

Para que o escalonamento comportamental seja possível é necessário conhecer o processamento cognitivo envolvido

³ Ver a respeito em: Wright e Stone (1979); Hambleton, Swaminatham e Rogers (1991); Pasquali (1996); Primi (1998a, 1998b).

nos problemas e a relação entre o processamento e a dificuldade, pois assim pode-se vincular as afirmações sobre o processamento à escala de complexidade dos itens.

Storey (1997) descreve que, nos últimos 25 anos, duas concepções têm polarizado o debate sobre quais processos cognitivos ocorrem na resolução do teste de Cloze. De um lado, figuram pesquisadores que enfatizam o processamento global contextual como elemento central na elaboração das respostas. De outro, figuram os que enfatizam o processamento local imediato ou externo ao texto (conhecimento vocabular prévio) como elementos centrais da resposta. Considera, entretanto, que tais diferenças associam-se principalmente à ênfase diferencial dos estudos ao tipo de estrutura do Cloze.

Ambas as posições podem estar corretas, pois o tipo de processamento requerido e a habilidade avaliada pelo teste de Cloze dependem basicamente do método adotado na criação das lacunas. Assim, determinado tipo de lacuna pode enfatizar o uso do contexto, comprovando a primeira posição, enquanto outro tipo de lacuna pode enfatizar a aplicação de conhecimento prévio (conhecimento lexical), favorecendo a segunda posição (Abraham & Chapelle, 1992).

Corroborando com essa visão, Bensoussan (1990, p.18) afirma que: “A leitura envolve simultaneamente o processamento top-down, pelo qual o leitor aplica o conhecimento prévio ao texto, e o processamento bottom-up pelo qual o leitor deriva significado das macro estruturas inerentes ao texto”. No entanto, é importante ressaltar que a prática isolada da habilidade de leitura tem, subjacentemente, uma visão do processamento do tipo bottom-up, muito freqüente no ensino fundamental, que determina apenas uma interação passiva do leitor com o texto, que não é a desejável para o aluno universitário.⁴

As pesquisas têm demonstrado que o processamento requerido para o tipo de Cloze usado nos estudos com eliminação aleatória de vocábulos depende principalmente da aplicação de conhecimento prévio externo ao texto (Storey, 1997).

A pesquisa de Abraham e Chapelle (1992) investigou o efeito de oito características das lacunas criadas na complexidade do preenchimento. As características foram: a) distância entre a lacuna criada e o contexto relevante ao seu preenchimento, b) número de sílabas da sentença em que a lacuna está inserida, c) freqüência que a alternativa correta ocorre em outra parte do texto, d)

tipo de palavra funcional ou semântica, e) tamanho da palavra que deve ser recuperada, f) número de respostas possíveis, g) número de formas possíveis - necessidade de inflexão morfêmica, h) tipo de resposta: múltipla alternativa - resposta construída.

Os resultados obtidos demonstraram que o efeito dependia do procedimento de construção das lacunas. Quando as lacunas são construídas pelo procedimento de razão fixa (Ex.: eliminação aleatória de todo quinto vocábulo), a complexidade está associada principalmente aos fatores já citados: freqüência em que a palavra ocorre no texto, tipo de palavra omitida e tipo de resposta necessária, escolhida entre as alternativas disponíveis ou construída. Abraham e Chapelle (1992) concluíram, a partir disto, que a habilidade avaliada por este tipo de Cloze associa-se à recuperação do vocabulário disponível na memória de longo prazo. Por outro lado, quando as lacunas são construídas por um procedimento racional (eliminação de palavras que necessariamente requerem o uso dos elementos contextuais claramente identificáveis), a complexidade associou-se principalmente ao fator a), a distância entre a lacuna e o contexto relevante ao seu preenchimento. Abraham e Chapelle (1992) consideraram que a habilidade avaliada, neste caso, é a de identificar e utilizar as informações contextuais na recuperação do vocabulário. Os autores descobriram também que os escores deste tipo de Cloze relacionavam-se ao teste GEFT (Group Embedded Figures Tests) de Witkin, Oltman, Raskin e Karp (1987), uma medida de processamento visual e inteligência fluida.

Os estudos citados têm sido realizados com textos em inglês, fazendo surgir o questionamento de sua aplicabilidade para testes de Cloze com textos em português, pois existem apenas algumas evidências de que essas descobertas também se aplicam aos textos em português. Por exemplo, em um estudo buscando verificar a eficiência da técnica de Cloze como instrumento de diagnóstico da compreensão em leitura, de prognóstico do desempenho acadêmico e de eficiência no aumento de compreensão em leitura, Santos (1991) verificou que os erros cometidos ocorriam com maior freqüência em palavras de determinadas classes gramaticais. Das categorias mais difíceis às mais fáceis, observou: adjetivos, advérbios, substantivos, verbos, pronomes, contrações, conjunções, preposições e artigos. A autora concluiu que as categorias com forte carga semântica (adjetivos, substantivos, verbos e advérbios) tenderam a se mostrar mais difíceis do que as categorias de relatores, como artigos e pronomes, por exemplo. Estes resultados foram coerentes com os encontrados por Egypto (1983).

⁴ Um aprofundamento da discussão sobre os diversos modelos teóricos de compreensão em leitura pode ser obtida consultando-se Smith (1978; 1997); Anderson e Pearson (1984); Norris (1986); Kintsch (1988, 1994); Robeck e Wallace (1990).

No contexto do presente estudo, o trabalho de Abraham e Chapelle (1992) fornece informações importantes sobre os processos cognitivos envolvidos na resolução de testes de Cloze com eliminação aleatória, o procedimento empregado nesta pesquisa. Considera, ainda, que a TRI pode trazer duas vantagens para a utilização do Cloze: a primeira refere-se à possibilidade de se obter notas equiparáveis independentes da dificuldade dos textos; a segunda é a possibilidade de interpretações de desempenho referenciadas nas operações cognitivas envolvidas, isto é, referindo-se ao conteúdo da tarefa, tornando as interpretações qualitativamente mais refinadas.

Assim, esta pesquisa foi desenvolvida com o objetivo de investigar a aplicação da TRI na análise de um instrumento de Cloze, visando o refinamento técnico da avaliação da habilidade de compreensão em leitura.

Método

Participantes

Participaram desse estudo 612 estudantes universitários ingressantes que cursavam o segundo semestre do primeiro ano. Os estudantes eram provenientes de quatro cursos de dois dos três campus de uma universidade particular do interior do estado de São Paulo (Medicina, Odontologia, Administração e Psicologia), com variação de idade de 17 a 52 anos ($m=21a9m$; $dp=4a9m$). Dentre eles, 75,9% tinham de 17 a 22 anos, e a maioria, 67,8%, era do sexo feminino.

Material

Utilizou-se uma crônica de Luís Fernando Veríssimo, de 1995, chamada Desentendimento, estruturada segundo a técnica de Cloze tradicional, com omissão do quinto vocábulo, num total de 40 lacunas a serem preenchidas. A primeira oração e as orações finais do referido texto foram preservadas na íntegra, permitindo ao leitor uma visão mais completa do tema abordado (veja Anexo A).

Procedimento

O instrumento foi aplicado como parte de um projeto de avaliação das habilidades básicas de ingressantes universitários, que incluiu a aplicação de seis provas em duas sessões de avaliação, feitas coletivamente em classes com aproximadamente 80 alunos. O teste de Cloze foi aplicado em uma das sessões, juntamente com duas outras provas: de juízo moral e de conhecimentos gerais, sempre na mesma ordem de apresentação. As instruções de aplicação incluíam a informação dos objetivos do projeto e a solicitação da participação, que foi feita em caráter voluntário, com baixo índice de rejeição dos participantes (menos que 5% dos alunos de cada turma recusaram-se a participar).

Resultados

Análise geral dos itens

Na Tabela 1, apresenta-se as estatísticas resultantes das análises dos itens, ou seja, de cada lacuna criada. Na primeira coluna apresenta-se o número da lacuna. Em seguida o índice de dificuldade (ID) que constitui a proporção de acertos no item. Em seguida apresenta-se a correlação ponto biserial entre o acerto no item e o escore total no teste (r_{bi}).

Na sexta coluna, apresenta-se o índice de discriminação (D). Esse índice foi criado, em um primeiro momento, separando da amostra aproximadamente 27% dos participantes que obtiveram os maiores escores (acima de 22, $N=180$) e 27% dos que obtiveram os menores escores (abaixo de 19, $N=220$). Calculou-se, então, a proporção de acertos do primeiro (p_{at}) e do segundo grupo (p_{bt}) e, em seguida, o índice D subtraindo essas duas proporções. Ele representa, então, a diferença na proporção de acertos entre dois grupos: um que demonstrou uma habilidade acima da média em compreensão de leitura, e outro, abaixo da média. Se esse índice for próximo de zero ou negativo, ele indica que a lacuna criada não conseguiu discriminar os alunos hábeis daqueles inábeis sendo, portanto, de má qualidade.

A análise geral da Tabela 1 indicou uma peculiaridade importante: os itens tenderam a ser ou muito fáceis ou muito difíceis. A Figura 1 mostra o número de itens por níveis classificados em cinco intervalos de dificuldade. Observa-se neste gráfico duas categorias extremas predominantes: 0 a 0,20 e 0,81 a 1. Em 10 itens, ou 25% das lacunas, menos de 10% dos estudantes apresentaram o vocábulo correto.

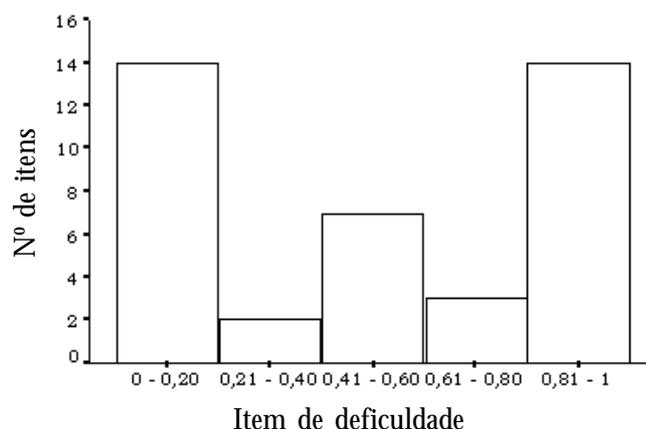


Figura 1. Distribuição dos itens entre cinco intervalos de dificuldade.

De um modo geral, os índices de discriminação foram adequados, ou seja, foram valores maiores do que zero. Isso indicou que, comparando os estudantes que acertaram um determinado item com os que erraram, encontrou-se, entre os que acertaram, uma tendência ao

acerto nos outros itens e a atingir níveis mais altos de compreensão.

Em alguns casos, os índices foram próximos de zero, nomeadamente: 7, 12, 18, 23, 31, 32, 35 e 36, e por serem especificamente identificados como palavras de conteúdo,

Tabela 1
Propriedades Psicométricas dos Itens (lacunas) do Teste de Cloze

	Lacuna	ID	r_{bi}	p_{bx}	p_{at}	D	a	b
Adjetivos	4	0,08	0,24	0,01	0,22	0,20	0,76	2,33
	18	0,01	0,02	0,01	0,01	0,00	0,92	3,00
	32	0,01	0,06	0,00	0,02	0,02	0,79	3,00
	36	0,01	-0,02	0,01	0,00	-0,01	0,83	3,00
	40	0,46	0,34	0,28	0,70	0,42	0,48	0,12
Verbos	14	0,33	0,28	0,20	0,51	0,31	0,48	1,07
	16	0,11	0,26	0,03	0,27	0,24	0,74	2,08
	23	0,02	0,06	0,01	0,03	0,02	0,74	3,00
	27	0,08	0,25	0,00	0,21	0,21	0,81	2,14
	31	0,01	0,06	0,00	0,02	0,02	0,88	3,00
	38	0,75	0,37	0,61	0,86	0,25	0,50	-1,66
Advérbios	9	0,12	0,24	0,04	0,23	0,19	0,64	2,22
	30	0,61	0,48	0,31	0,85	0,55	0,71	-0,51
Pronomes	13	0,50	0,34	0,34	0,75	0,41	0,51	0,09
	19	0,57	0,33	0,39	0,74	0,35	0,46	-0,33
Substantivos	1	0,86	0,44	0,71	0,94	0,23	0,69	-1,96
	3	0,53	0,39	0,30	0,81	0,51	0,58	-0,05
	5	0,59	0,37	0,38	0,79	0,42	0,53	-0,36
	6	0,26	0,41	0,03	0,59	0,57	0,84	1,04
	7	0,00	0,04	0,00	0,01	0,01	0,99	3,00
	11	0,81	0,55	0,56	0,98	0,42	0,95	-1,26
	12	0,01	0,02	0,01	0,01	0,00	0,86	3,00
	24	0,83	0,54	0,60	0,99	0,39	0,89	-1,41
	28	0,93	0,66	0,81	1,00	0,19	1,07	-2,46
	35	0,02	0,06	0,01	0,02	0,01	0,79	3,00
Preposições	37	0,89	0,57	0,74	0,99	0,25	0,92	-2,13
	2	0,82	0,50	0,61	0,96	0,35	0,80	-1,48
	8	0,92	0,59	0,81	0,98	0,16	0,91	-2,28
	17	0,86	0,54	0,67	0,98	0,31	0,81	-1,71
	20	0,91	0,61	0,77	1,00	0,23	0,96	-2,07
	21	0,50	0,37	0,27	0,71	0,44	0,53	0,09
	22	0,91	0,56	0,79	0,99	0,20	0,87	-2,12
	25	0,77	0,40	0,62	0,89	0,27	0,57	-1,50
	26	0,87	0,50	0,74	0,97	0,23	0,75	-1,98
	29	0,86	0,57	0,66	0,99	0,33	0,90	-1,82
Artigos	34	0,17	0,19	0,09	0,23	0,14	0,49	2,22
	39	0,20	0,12	0,16	0,24	0,08	0,40	1,21
	10	0,85	0,52	0,70	0,97	0,27	0,79	-1,74
	15	0,55	0,44	0,26	0,81	0,55	0,64	-0,10
	33	0,89	0,56	0,75	0,99	0,24	0,88	-2,43

isto é, característicos da função semântica do discurso, esses itens foram também muito difíceis o que provavelmente contribuiu para a redução dos índices de discriminação. Este achado indica que uma possível medida objetivando a otimização das propriedades psicométricas do teste de Cloze seria a eliminação racional de palavras, evitando eliminar aquelas cujo conteúdo fosse muito específico, já que possuem uma proporção de acerto muito baixa e não contribuem para a diferenciação das habilidades dos sujeitos. Uma outra solução seria a análise destas omissões, para as quais seria realizada uma correção por sinônimos. Isto aumentaria a proporção de acertos, uma vez que possibilitaria mais de uma resposta, o que também poderia contribuir para o aumento do índice de discriminação. No entanto, vale ressaltar que tal procedimento seria muito mais trabalhoso do que considerar correta apenas a palavra originalmente utilizada pelo autor do texto.

Nas últimas duas colunas apresentam-se os parâmetros de discriminação (a) e dificuldade (b). Esses parâmetros foram calculados utilizando o modelo logístico de dois parâmetros com o auxílio do software XCALIBRE (Assessment Systems Corporation, 1996). Estes parâmetros são os elementos que caracterizam cada item e são variáveis da equação da Curva Característica do Item (CCI) que descreve a relação entre a habilidade e a probabilidade de acerto (exemplos podem ser vistos na Figura 2). O índice de dificuldade habitualmente situa-se entre -2 a $+2$ e indica o valor de habilidade correspondente a probabilidade de acerto igual a 0,50. Este índice também é chamado parâmetro de localização já que, dependendo de seu valor, a curva se posicionará mais para a esquerda (itens fáceis, $b < 0$) ou mais para a direita (itens difíceis, $b > 0$). O índice de discriminação geralmente varia de 0 a 2,0 e está associado à inclinação da CCI. Nos itens cujas curvas são mais inclinadas, o aumento da habilidade de um ponto q_1 para outro ponto q_2 corresponderá a uma mudança mais acentuada na probabilidade de acerto do que ocorreria se a inclinação fosse menos acentuada. Quanto maior a inclinação da curva, maior o valor do índice de discriminação. Nos itens com discriminação elevada, variações de probabilidade de acerto estão mais fortemente associadas à habilidade.

Embora não totalmente comparáveis, o significado geral destes dois parâmetros assemelha-se muito ao índice de dificuldade e à correlação item-total da psicometria clássica. Portanto, as constatações com relação a estes índices são as mesmas que foram discutidas a partir dos índices clássicos.

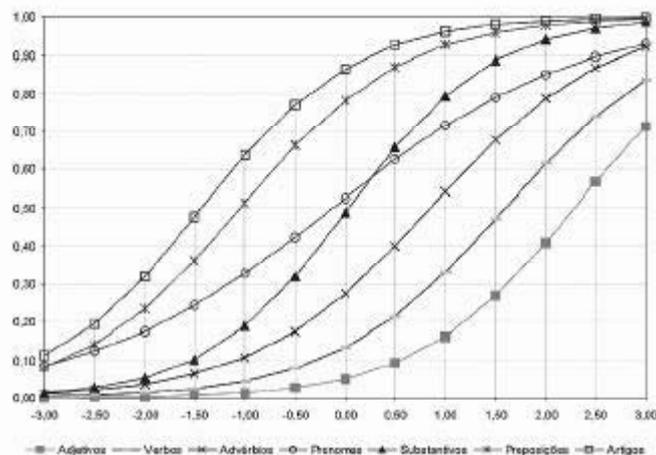


Figura 2. Curvas características dos subgrupos de itens com classes gramaticais semelhantes.

Uma questão específica sobre a aplicação da TRI, que necessitou ser examinada, referiu-se a eficácia do modelo em representar satisfatoriamente os padrões de resposta observados. Para verificar a adequação do uso do modelo de dois parâmetros analisou-se os resíduos padronizados. Esses resíduos foram obtidos: a) criando-se subgrupos com escores semelhantes; b) calculando-se a probabilidade esperada, segundo o modelo, para cada subgrupo, através da função da curva característica do item; c) calculando-se a probabilidade real observada, e d) descobrindo a diferença entre o valor teórico e o observado e expressando-o através dos resíduos padronizados.

Seguindo as orientações do manual do programa empregado, que indica resíduos maiores do que 2 como indicadores de desajuste, observou-se que somente o item 39 satisfaz esse critério. Portanto, concluiu-se que o modelo de dois parâmetros foi adequado para a representação dos resultados.

Na Tabela 2 apresentam-se as estatísticas da escala final composta pela soma do número de vocábulos corretos calculados a partir do resultado dos 612 participantes. Em função das características dos itens, o instrumento atingiu um bom nível de consistência interna, ou seja, os itens avaliam, homoganeamente, em maior ou menor grau, uma mesma capacidade cognitiva.

O escore total, para o qual foram calculadas as estatísticas apresentadas na Tabela 2, inclui todos os itens independentemente de sua natureza. Obviamente, alguns itens muito difíceis e pouco discriminativos poderiam ter sido eliminados, sem que isso influenciasse negativamente as propriedades psicométricas obtidas. No entanto, decidimos manter todos os itens para que fosse possível

Tabela 2
Estatísticas Finais da Escala Formada pela Soma dos Vocábulos Corretos

Número de itens	40
Média	19,49
Variância	26,11
Desvio-padrão	5,11
Assimetria	-0,83
Curtose	2,58
Mínimo	0
Máximo	31
Mediana	20
Alfa	0,829
Média ID	0,487
Média r_{bi}	0,345

Tabela 3

Médias e Desvios-Padrão (dp) dos Índices de Dificuldade dos Itens Agrupados em Função da Classe Gramatical

Categorias gramaticais	Média dos ID's	dp	Média dos b's	dp	Número de itens
Adjetivos	0,114	0,197	2,289	1,249	5
Verbos	0,217	0,286	1,604	1,750	6
Advérbios	0,364	0,341	0,853	1,931	2
Substantivos	0,521	0,382	0,038	2,149	11
Pronomes	0,536	0,049	-0,118	0,301	2
Preposições	0,709	0,285	-1,039	1,518	11
Artigos	0,764	0,188	-1,423	1,193	3
Média global	0,487	0,357	0,181	2,008	40

investigar quais características psicométricas um teste de Cloze construído pelo procedimento tradicional apresenta; isto é, atender nosso objetivo, o de explorar as características do instrumento tal como ele fora construído, deixando para futuros estudos a aplicação de procedimentos de otimização das propriedades psicométricas no teste de Cloze.

As Classes Gramaticais das Palavras e sua Dificuldade

Nesta parte do trabalho procurou-se investigar se a categoria gramatical das palavras que deveriam ser inseridas exerceria algum efeito na complexidade do processo de descoberta da palavra.

Inicialmente, classificou-se cada palavra omitida definindo sete subgrupos reunindo lacunas de classes semelhantes. Na Tabela 3 apresentam-se as médias dos índices de dificuldade, ID e b, e o número de itens que compunham cada categoria gramatical. O teste não paramétrico de Kruskal-Wallis indicou que as diferenças entre os subgrupos foram significativas (Qui-Quadrado = 14,84, gl=6, p= 0,008, significância exata estimada pelo método de Monte Carlo).

Observa-se que os adjetivos, verbos e advérbios foram as palavras mais difíceis de serem descobertas, os substantivos e pronomes foram de dificuldade média e as preposições e artigos as mais fáceis. Obviamente, a dificuldade das palavras dentro de cada classe variou consideravelmente como indicam os desvios-padrão, fato que pode ser observado mais detalhadamente na Tabela 1. Entretanto, parte significativa da variação da dificuldade esteve associada às classes gramaticais.

Ordenando-se as classes gramaticais da mais difícil a mais fácil, e atribuindo-se a ordem de 1 a 7 para cada classe, criou-se uma nova variável correspondente à dificuldade da classe. Essas ordens foram atribuídas a cada palavra membro de uma determinada classe. Verificou-se, então, a correlação dessa nova variável com a dificuldade dos itens. Essa correlação foi de 0,62 ($p < 0,001$), indicando que 38,4 % da variância da dificuldade foi prevista em função do conhecimento da classe gramatical de cada palavra.

A partir dessa constatação calculou-se uma curva característica para cada classe gramatical. Essas curvas foram calculadas: a) agrupando-se os itens de classes

gramaticais comuns, b) obtendo-se as médias dos índices de discriminação e dos índices de dificuldade para itens de classe gramatical comum e c) calculando-se as curvas características de cada classe, considerando as estimativas médias de dificuldade e discriminação.

Na Figura 2 apresentam-se essas curvas. Esse gráfico permite distinguir, a cada nível de compreensão, qual a probabilidade média de acertar uma palavra de uma determinada classe gramatical. Como os participantes tiveram suas notas estimadas nessa escala, é possível entender os resultados com referência a essas classes.

As curvas apresentadas não devem ser consideradas como curvas diretamente aplicáveis a qualquer palavra de uma determinada classe. Isso porque, como se observou anteriormente, ainda existe uma variância considerável entre as palavras de uma determinada classe (ver Tabela 1). Estas curvas apresentam uma probabilidade de acerto médio para uma palavra de determinada classe em razão da habilidade, o que seria a melhor suposição possível a partir da informação da categoria gramatical da palavra.

Discussão

Dada a importância da compreensão em leitura entre estudantes universitários, este trabalho procurou compreender mais detalhadamente as propriedades psicométricas da técnica de Cloze para a avaliação dessa habilidade. Especificamente, procurou-se refinar as interpretações atribuídas às notas obtidas pelo teste.

Verificou-se que os adjetivos, verbos e advérbios foram as lacunas mais difíceis de serem completadas. Esse resultado foi em parte coerente com os dados obtidos na análise de erros feita por Santos (1991), que encontrou como mais difíceis, as lacunas referentes a adjetivos, advérbios, substantivos e verbos. Abraham e Chapelle (1992) classificam esses tipos de palavras como palavras de conteúdo, relativas à função semântica do discurso, em oposição às palavras funcionais, ligadas à função coesiva do discurso. Com relação ao processamento cognitivo, as palavras de conteúdo requerem, com maior ênfase, o processamento semântico (recuperação das palavras da memória de longo prazo), em oposição ao processamento sintático (recuperação das palavras a partir da estrutura sintática das frases, nas quais a palavra omitida se insere).

A compreensão em leitura, tal como exigida no teste de Cloze, depende, entre outras variáveis, da habilidade do leitor em estabelecer relações entre os elementos do texto e, também, de sua capacidade em desenvolver associações apropriadas entre o conhecimento anteriormente adquirido e a informação expressa (Riley, 1986). Considerando-se essas duas vertentes à luz dos resultados que apontam

relação entre a dificuldade no preenchimento das lacunas com a complexidade do conteúdo semântico das palavras, pode-se inferir que, no presente instrumento, observa-se a predominância do segundo elemento, isto é, a extensão e profundidade do conhecimento semântico lexical dos alunos.

As preposições e os artigos foram as lacunas mais fáceis de serem preenchidas. Essas palavras formam unidades com as palavras imediatamente adjacentes e, portanto, na presença dessas últimas palavras, elas são recuperadas mais facilmente, o que confirma uma maior facilidade de processamento do tipo bottom-up. Um segundo aspecto associado à facilidade é o número de opções de resposta (Abraham & Chapelle, 1992): o número de artigos e preposições dos quais o aluno tem que escolher para preencher a lacuna é muito menor do que o número de adjetivos e verbos. Em termos de processamento cognitivo pode-se inferir que o preenchimento dessas lacunas requer o uso do contexto imediato, isto é, o processamento sintático simples, uma vez que os elementos sintáticos que devem ser processados estão muito próximos à lacuna a ser preenchida. Portanto, a descoberta dessas palavras não depende de uma compreensão mais global do texto, o que deve ser avaliado à luz das teorias sobre compreensão de leitura propostas por Smith (1978, 1997), Norris, 1986, e Kintsch (1988, 1994).

Os substantivos apresentaram dificuldade média; entretanto, entre eles, houve uma grande dispersão dos índices de dificuldade (-2,46 a 3,0). Santos (1991) observou que a ordem de dificuldade de palavras de diferentes classes gramaticais mudava, dependendo do instrumento utilizado. As ordens se alteravam principalmente entre substantivos, verbos e pronomes. Esse resultado é coerente com o encontrado neste estudo e poderia ser explicado pela grande variabilidade na dificuldade dos substantivos. Esse fato chama atenção para uma limitação do presente estudo: grande parcela da variância (61,6%) da dificuldade permaneceu inexplicada, o que indica que o simples conhecimento da classe gramatical não garante a compreensão de todas as fontes de dificuldade das lacunas. É necessária a sofisticação do sistema de classificação para se atingir um nível mais alto de previsibilidade da dificuldade. Provavelmente, a classificação sintática das palavras pode constituir um aprimoramento importante para este fim.

As análises psicométricas iniciais indicaram que os itens se agruparam coerentemente, formando uma escala precisa. Embora a escala tenha atingido um nível alto de consistência interna, observou-se que várias lacunas não preenchem os critérios psicométricos aceitáveis, isto é, apresentam dificuldade muito elevada e baixa discriminação. Este fato sugere que o procedimento de

eliminação racional possa ser mais adequado que o procedimento de eliminação aleatória, se o objetivo for o de criar instrumentos de dificuldade mediana com propriedades psicométricas otimizadas.

Ainda que parcialmente, este estudo trouxe novas informações sobre a viabilidade de utilização da técnica de Cloze, reafirmando sua boa qualidade como instrumento para a avaliação de compreensão em leitura. Sugere que haja um incremento em pesquisas que investiguem, não apenas o seu potencial de diagnóstico, como também, outra dimensão promissora, como técnica de intervenção psicopedagógica. Tal dimensão, referida como alternativa viável de desenvolvimento da compreensão em leitura por Condemarim e Milicic (1988) e Santos (1997), implica na utilização de formas alternativas do procedimento de Cloze, que introduzem as dificuldades dentro de um padrão hierarquizado, permitindo aos sujeitos ganhos de domínio gradual de compreensão e a conseqüente superação das dificuldades de leitura. Dessa forma, professores universitários das mais diferentes disciplinas poderiam incorporar às suas práticas educativas, a realização de atividades de leitura, valendo-se da técnica de Cloze, o que propiciaria aos alunos a utilização de estratégias metacognitivas de leitura, envolvendo o monitoramento da sua aprendizagem.

Referências

- Abraham, R. G. & Chappelle, C. A. (1992). The meaning of Cloze test scores: An item difficulty perspective. *The Modern Language Journal*, 76(4), 468-479.
- Anderson, R. C. & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. Em P. D. Pearson, R. Barr, M. L. Kamil & P. Mosenthal (Orgs.), *Handbook of reading research* (Vol. 1, pp. 251-291). New York: Longman.
- Assessment Systems Corporation (1996). *User's manual for the MicroCat Testing System*. ASC: St. Paul.
- Bensoussan, M. (1990). Redundancy and the cohesion Cloze. *Journal of Research in Reading*, 13(1), 18-37.
- Bormuth, R. J. (1968). Cloze test readability: Criterion reference scores. *Journal of Educational Measurement*, 5, 189-196.
- Carrol, J. B. (1993). Test theory and the behavioral scaling of test performance. Em N. Frederiksen, R. J. Mislevy & I. I. Bejar (Orgs.), *Test theory for a new generation of tests* (pp. 297-322). Hillsdale, NJ: Lawrence Erlbaum.
- Carrol, J. B. (1997). The three-stratum theory of cognitive abilities. Em D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Orgs.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 122-130). New York: The Guilford Press.
- Cohen, J. H. (1975). The effect of content area material on Cloze test performance. *Journal of Reading*, 19(3), 247-250.
- Condemarim, M. & Milicic, N. (1988). *Test de Cloze: Procedimiento para el desarrollo y la evaluación de la comprensión lectora*. Santiago, Chile: Editorial Andres Bello.
- Cronbach, J. L. (1996). *Fundamentos da testagem psicológica*. Porto Alegre: Artes Médicas.
- Egypto, M. S. R. (1983). *Leitura em estudantes universitários: Estudo contrastivo de caracterização de nível de desempenho*. Dissertação de Mestrado não-publicada, Curso de Pós-Graduação em Psicologia, Universidade Federal da Paraíba. João Pessoa, PB.
- Embretson, S. (1996). The new rules of measurement. *Psychological Assessment*, 8(4), 341-349.
- Hambleton, H. K., Swaminatham, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension. A construction-integration model. *Psychological Review*, 95(2), 163-182.
- Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, 49(4), 294-303.
- Marini, A. (1986). *Compreensão da leitura no ensino superior: Teste de um programa para treino de habilidades*. Tese de Doutorado não-publicada, Curso de Pós-graduação em Instituto de Psicologia, Universidade de São Paulo. São Paulo, SP.
- Norris, D. (1986). Word recognition: Context effects without priming. *Cognition*, 22, 93-136.
- Pasquali, L. (1996). Teoria da resposta ao item - IRT: Uma introdução. Em L. Pasquali (Org.), *Teoria e métodos de medida em ciências do comportamento* (pp. 173-195). Brasília, INEP.
- Page, W. D. (1975). The post oral reading Cloze test: New link between oral reading and comprehension. *Journal of Reading Behavior*, 7(4), 383-389.
- Primi, R. (1998a). *Desenvolvimento de um instrumento informatizado para avaliação do raciocínio analítico*. Tese de Doutorado não-publicada, Curso de Pós-graduação em Instituto de Psicologia, Universidade de São Paulo, São Paulo-SP.
- Primi, R. (1998b). Avanços tecnológicos na avaliação da inteligência. Em F. C. Capovilla, M. J. Gonçalves & E. C. Macedo (Orgs.), *Tecnologia em (re)habilitação cognitiva* (pp. 159-166). São Paulo: EDUNISC.
- Riley, J. D. (1986). Progressive Cloze as a remedial technique. *The Reading Teacher*, 39(6), 576-581.
- Robeck, M. C. & Wallace, R. R. (1990). *The psychology of reading: An interdisciplinary approach*. New Jersey: Erlbaum.
- Santos, A. A. A. (1981). *Desenvolvimento do hábito de leitura e compreensão de textos através da aplicação de fichas: Um estudo com adolescentes carentes*. Dissertação de Mestrado não-publicada, Curso de Pós-graduação em Psicologia, Pontifícia Universidade Católica de Campinas. Campinas, SP.
- Santos, A. A. A. (1990). *Compreensão em leitura na universidade: Um estudo comparativo entre dois procedimentos de treino*. *Estudos de Psicologia*, 7(2), 39-53.
- Santos, A. A. A. (1991). *Desempenho em leitura: Um estudo diagnóstico da compreensão e hábitos de leitura em universitários*. *Estudos de Psicologia*, 8(1), 6-19.
- Santos, A. A. A. (1997). *Psicopedagogia no 3º grau: Avaliação de um programa de remediação em leitura e estudo*. *Pro-posições*, 1(22), 27-37.
- Smith, F. (1978). *Reading*. Cambridge: Cambridge University Press.
- Smith, F. (1997). *Reading without nonsense*. Toronto: Teachers College Press.
- Storey, P. (1997). Examining the test-taking process: A cognitive perspective on the discourse Cloze test. *Language Testing*, 14(2), 214-231.
- Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 30, 415-433.
- Witkin, H. A., Oltman, P. K., Raskin, E. & Karp, S. A. (1987). *Tests de figuras enmascaradas*. Madrid: TEA Ediciones S.A.
- Witter, G. P. (1990). *Pesquisa documental, pesquisa bibliográfica e busca de informação*. *Estudos de Psicologia*, 7(1), 5-30.
- Wright, B. D. & Stone, M. H. (1979). *Best Teste Design*. Chicago: MESA.

Recebido: 03/12/2001

Revisado: 25/02/2002

Aceite Final: 14/05/2002

Sobre os autores

Acácia A. Angeli dos Santos é Psicóloga Escolar e do Desenvolvimento Humano pela Universidade de São Paulo; é Docente do Programa de Estudos Pós-graduados em Psicologia da Universidade de São Francisco.

Ricardo Primi é Psicólogo, Doutor em Psicologia Escolar e do Desenvolvimento Humano pela Universidade de São Paulo, com parte dos estudos desenvolvidos na Yale University (EUA). Coordenador do Laboratório de Avaliação Psicológica e Educacional (LabAPE) que recebe financiamento da FAPESP, é coordenador do Programa de Estudos pós-graduados em Psicologia da Universidade São Francisco em Avaliação Psicológica. É pesquisador CNPq.

Fernanda de O. S. Taxa é Pedagoga, Doutora em Educação pela Universidade Estadual de Campinas.

Claudette M. M. Vendramini é Estatística, Doutora em Educação pela Universidade Estadual de Campinas, é Professora da Graduação e do Programa de Estudos Pós-graduados em Psicologia da Universidade São Francisco.

Anexo A

Apresenta-se abaixo o texto utilizado neste estudo, de Luís Fernando Veríssimo (1995) chamado Desentendimento. As palavras grifadas correspondem às lacunas (itens) formadas no original:

“A comunicação com garçons pode ser uma provação em qualquer parte do mundo. Costumo tomar leite frio, de manhã. Em Tóquio, na primeira vez em que nos aventuramos a tomar café fora do hotel, fui munido da palavra japonesa para leite. *Miruku*, ou coisa parecida. Experimentei-a no garçom, que me devolveu um olhar cheio de perplexidade. Tentei de novo, em várias flexões. Finalmente acertei: era preciso dizer a palavra rapidamente. Veio o leite. Quente. Não sei que mímica usei para dizer que o queria frio - duvido que tenha me abraçado e simulado uma tremedeira, o que só faria o garçom sair atrás de alguma corrente de ar perdida dentro do restaurante -, mas ele entendeu, levou o copo e o trouxe de volta. Com pedras de gelo dentro do leite. Um japonês que observava nosso desentendimento de uma mesa ao lado se ofereceu para ajudar. Traduziu meu pedido para o garçom. Tudo esclarecido. Agradei a intermediação enquanto o garçom voltava para a cozinha com o copo e pouco depois o trazia de volta. Afastada a ameaça à paz mundial e à amizade entre os povos, levantei o copo em direção do tradutor num brinde agradecido. O leite estava morno.

Em Budapeste entramos num restaurante em que o menu era em húngaro e alemão, mas o húngaro é tão difícil que entendemos em alemão. Não tive dificuldade em me fazer entender pelo garçom. Apontei com o dedo para a única palavra em alemão que qualquer pessoa precisa saber, além de bite e danke: forelle. Quer dizer truta. Eu poderia andar pela Alemanha comendo trutas em todas as refeições - salvo no café da manhã, claro. São as melhores do mundo, e não havia razão para as trutas húngaras não serem parecidas. Não fiquei sabendo. Não havia forelle em qualquer língua no restaurante. Acabei apontando com o dedo para a palavra com menos consoantes que vi, e que acabou sendo o nome de uma almôndega do tamanho aproximado de uma granada. Sobrevivi.”