

CONDIÇÕES PRIMORDIAIS E SECUNDÁRIAS PARA PRODUÇÃO DE TESTES METALINGÜÍSTICOS EM L2

CONDICIONES PRIMORDIALES Y SECUNDARIAS PARA PRODUCCIÓN DE PRUEBAS
METALINGÜÍSTICAS EN L2

ESSENTIAL AND SECONDARY CONDITIONS FOR PRODUCTION OF METALINGUISTIC
TESTS IN L2

Rosi Ana Grégis*

Universidade Feevale

RESUMO: Os testes metalingüísticos têm sido usados por muitos pesquisadores como um meio eficaz de fornecer dados sobre a natureza da linguagem. Bley-Vroman, Felix e Ioup (1988), Hawkins e Chan (1997), Johnson e Newport (1989), Schachter (1989), Schütze (1996), White (2003), entre outros, utilizaram esses testes para dar suporte a suas hipóteses sobre a aquisição e o desenvolvimento de uma segunda língua (L2). Para este artigo, foram analisadas três pesquisas, com a finalidade de criar condições relevantes para aqueles pesquisadores que necessitam construir testes metalingüísticos (TM) para validar suas investigações. Além disso, avaliamos testes utilizados na área de Aquisição de Segunda Língua (ASL), com o intuito de verificar se as condições aqui apresentadas foram atendidas. Percebemos, porém, que os testes metalingüísticos utilizados por esses autores poderiam ter sido construídos com mais cautela para que tivessem maior confiabilidade e uma melhor validação dos resultados.

PALAVRAS-CHAVE: Testes metalingüísticos. Pesquisa em segunda língua. Aceitabilidade linguística.

RESUMEN: Muchos investigadores han utilizado las pruebas metalingüísticas como un eficaz medio de proveer datos sobre la naturaleza del lenguaje. Bley-Vroman, Felix y Ioup (1988), Hawkins y Chan (1997), Johnson y Newport (1989), Schachter (1989), Schütze (1996), White (2003), entre otros, utilizaron esas pruebas para dar soporte a sus hipótesis sobre la adquisición y el desarrollo de una lengua segunda (L2). Para este artículo, fueron analizados tres trabajos a fin de crear condiciones relevantes para quienes necesitan construir pruebas metalingüísticas para validar sus investigaciones. Además, evaluamos testes utilizados en el área de Adquisición de Segunda Lengua (ASL), con la intención de verificar si fueron acogidas las condiciones aquí presentadas. Sin embargo, percibimos que las pruebas metalingüísticas utilizadas por esos autores podrían haber sido construidas con más precaución, para que tuvieran más confiabilidad y una mejor validación de los resultados.

PALABRAS CLAVE: Pruebas metalingüísticas. Investigación en segunda lengua. Aceptabilidad lingüística.

* Rosi Ana Grégis possui doutorado em Linguística Aplicada pela Pontifícia Universidade Católica do Rio Grande do Sul, realizou um período de estudos (Doutorado Sanduíche) na Universidade de Essex, Reino Unido. Atualmente, é professora e pesquisadora da Universidade Feevale, atuando na área de língua inglesa, teorias linguísticas e sociolinguísticas e aquisição de segunda língua. E-mail: rosiana@feevale.br.

ABSTRACT: Several researchers have used metalinguistic tests as an effective way to provide data on the nature of language. Bley-Vroman et al. (1988), Hawkins e Chan (1997), Johnson e Newport (1989), Schachter (1989), Schütze (1996), White (2003), among others, have used those tests to support their hypothesis on the acquisition and the development of a second language (L2). In this article, we analyzed three researches with the aim of creating relevant conditions for producing metalinguistic tests (MT) to validate inquiries. Beyond that, we evaluate some tests, used in Second Language Acquisition, aiming to find out if the conditions presented in this research were satisfied. We point out, however, the tests used by some authors could have been built more carefully, so that they would be more trustful and achieve better validation of their results.

KEYWORDS: Metalinguistic tests. Second language research. Linguistic acceptability.

1 INTRODUÇÃO

Resultados de pesquisas em segunda língua são, em sua maioria, dependentes de coletas de dados. Um dos objetivos principais dessas pesquisas é obter informações sobre o conhecimento linguístico dos aprendizes a fim de encontrar, através de suas respostas, um reflexo de sua competência linguística. Na verdade, os teóricos da ASL (Aquisição de Segunda Língua) têm o objetivo de entender os processos inseridos na competência dos aprendizes de uma L2 e não somente as características evidenciadas em suas performances. Os testes metalinguísticos (doravante, TM) podem ser confeccionados de formas variadas, mas geralmente um conjunto de sentenças é apresentado aos aprendizes para que eles façam julgamentos, afirmando se as sentenças são gramaticais, não-gramaticais, aceitáveis, não-aceitáveis, corretas, incorretas, ou outros tipos de respostas pedidas pelos pesquisadores.

À medida que estudiosos de ASL começaram a inserir os TM em suas pesquisas, iniciaram-se discussões que giravam em torno da seguinte questão: “É possível investigarmos a competência linguística diretamente?”. Apesar de haver alguns teóricos contra o uso de testes de julgamento¹ (BIRDSONG, 1989), há pesquisadores que afirmam que, embora o conhecimento linguístico não possa ser totalmente observado na performance, as intuições dos aprendizes podem ajudar nas descobertas acerca do funcionamento das interlínguas. Apesar de concordarmos que um TM mal formulado e mal aplicado possa tirar a validação de qualquer pesquisa, percebemos, através de leituras de teóricos que lidam com a questão (BIRDSONG, 1989; ELLIS; BARKHUIZEN, 2005; GASS, 1994; HAWKINS; CHAN, 1997; MACKKEY; GASS, 2005; MANDELL, 1999; SCHÜTZE, 1996; SORACE, 1996; TOWELL; HAWKINS, 1994; WHITE, 2003), que, se algumas condições forem respeitadas, essa técnica de obtenção de resultados pode ser de grande valia nas pesquisas de ASL. Nosso objetivo neste artigo é apresentar condições para que os TM aplicados em pesquisas de segunda língua sejam mais válidos e confiáveis.

2 TESTES METALINGUÍSTICOS EM L2

As razões pelas quais julgamentos linguísticos são utilizados em pesquisas dependem basicamente do que está sendo estudado pelos pesquisadores. Chaudron (1983; 2003) declara que esse tipo de metodologia surgiu com o objetivo de testar a adequação de gramáticas descritivas de uma dada língua. As habilidades sobre os TM passaram a ser objeto de estudo em pesquisas sobre o desenvolvimento de L1 e de L2, embora surgisse a preocupação de se conhecer mais a respeito de variáveis que pudessem influenciar esses julgamentos, tais como maturação cognitiva, idade, contexto linguístico e fatores sociais. A partir da década de 70, tornou-se muito comum pensar a língua não somente em termos de seu uso na comunicação diária, mas também como objeto de análise e observação em si mesma. Com isso, os TM passaram a ser um dos meios pelos quais os estudiosos obtêm respostas a fim de determinar as propriedades gramaticais da linguagem, ou seja, eles são usados para determinar quais sentenças são possíveis ou não em uma determinada língua.

Birdsong (1989) e Selinker (1972) chamaram atenção para o fato de que falantes de uma L2 não possuem a gramática final dessa língua como na sua L1 e por isso não têm a mesma capacidade de fazer julgamentos como os falantes nativos. Conforme Gass (1994), no caso dos falantes de L2, pesquisadores fazem inferências não somente a respeito do assunto específico da pesquisa, mas também

¹ Neste artigo, utilizaremos os termos “testes metalinguísticos”, “testes de julgamento” e “testes de aceitabilidade” intercambiavelmente.

sobre partes do sistema que os aprendizes têm internalizado. Em outras palavras, a autora defende que pode haver um desencontro entre os dois sistemas em questão. É por esse motivo que Chaudron (2003) lembra que os TM fazem parte de uma atividade bastante complexa e que devemos ter conhecimento de suas possíveis limitações. Gass (1994) afirma haver inúmeras interpretações equivocadas sobre o que os testes metalinguísticos representam: eles não têm como ser uma representação direta do que se considera competência linguística, mesmo porque competência é um construto teórico. Contudo, Gass é categórica ao dizer que não há dúvida de que os TM oferecem informações sobre a performance e que, através dessas informações, é possível obtermos *insights* sobre a competência linguística dos informantes. Assim como Gass (1994), White (2000) concorda que a performance pode nos auxiliar muito sobre o que caracteriza a competência linguística. Gass é bastante positiva quanto ao uso dos TM, declarando que, embora eles não nos levem a um acesso direto à competência do falante, eles fornecem informações sobre o que é possível e impossível na linguagem do aprendiz.

2.1 OBSERVAÇÕES TEÓRICAS SOBRE OS TESTES METALINGUÍSTICOS

Conforme Chaudron (1983; 2003), as questões teóricas que envolvem os TM são bastante complexas. Corder (1971) e Selinker (1972), que iniciaram os debates sobre essa questão na ASL, não concordavam quanto ao seu uso. Selinker sugeriu que os pesquisadores fossem cautelosos no uso dos TM; por outro lado, Corder considerava os aprendizes de L2 bons informantes sobre suas interlínguas e agradava-lhe o fato desses falantes terem condições de traduzir suas línguas maternas e de muitos deles estarem familiarizados com a terminologia metalinguística.

De acordo com Schachter e Yip (1990), esses pesquisadores partem do pressuposto de que os TM são um meio adequado para auxiliar o entendimento sobre a competência linguística. Assim, é possível argumentar o seguinte: “If L2 learners are able to judge correctly ungrammatical sentences ruled out by universal constraints just on the basis of positive evidence from the target language, then they have access to UG” (SCHACHTER; YIP, 1990, p.381)².

Assim como Chaudron (1983, 2003), as autoras Schachter e Yip (1990) concordam que, embora seja comum a afirmação de que os TM sejam um meio de acessar a competência, na verdade eles refletem um sistema extremamente complexo, no qual é muito difícil separar as variáveis relacionadas à performance, presentes nas atividades metalinguísticas. Segundo as autoras, testes metalinguísticos são o produto final de fatores gramaticais e extra-gramaticais. Por esse motivo, todos que trabalham com esse tipo de testagem devem estar cientes de que muitas interpretações equivocadas podem ocorrer. Por esse motivo, nosso objetivo é apresentar condições relevantes para os pesquisadores da área de ASL que desejam utilizar os TM em suas pesquisas.

Entendemos que a divisão dessas condições em primordiais e secundárias pode ajudar os pesquisadores a decidirem quais delas devem ser seguidas na formulação dos seus testes. Cada pesquisa, com suas particularidades, exigirá que o TM siga certas condições. Nossa intenção também é a de ajudar aqueles pesquisadores que estão em dúvida sobre qual tipo de teste pode ser aplicado, pois muitas vezes testes de julgamento ou aceitabilidade não serão os mais adequados, dependendo do tipo de estrutura estudada, dos objetivos do autor, da faixa etária dos sujeitos, entre outros fatores.

Iniciaremos mostrando as condições que são indispensáveis em pesquisas de L2 e, a seguir, apresentaremos as condições que consideramos recomendáveis ou secundárias. Esclarecemos que, embora as condições estejam numeradas por uma questão prática, isso não quer dizer que a primeira seja mais importante que a segunda, a segunda mais importante que a terceira e assim por diante. Muitas condições são bastante próximas e estão intrinsecamente ligadas, podendo haver características de algumas que se mesclam com as de outras.

²Se os aprendizes de L2 são aptos a julgar corretamente sentenças agramaticais, excluídas pelas restrições universais, somente com base em evidência positiva da língua alvo, então esses aprendizes têm acesso à UG (Todas as traduções neste artigo são de nossa autoria).

3 CONDIÇÕES PARA PRODUÇÃO DE TESTES METALINGÜÍSTICOS

3.1 CONDIÇÕES PRIMORDIAIS

Condição 1

- ✓ O teste deve ser totalmente claro e de fácil entendimento para os sujeitos. É bastante provável que eles não aceitem as sentenças se não compreenderem seu sentido.

Apesar de poder parecer uma condição bastante óbvia, não são todos os pesquisadores que prestam atenção em algo tão evidente. É muito comum observarmos que alguns TM possuem léxicos e estruturas gramaticais que não condizem com o nível de aprendizagem dos sujeitos. Isso pode confundir os pesquisadores no momento da avaliação das respostas, porque não há como saber se os sujeitos não responderam a questão corretamente porque não entenderam o sentido das frases (vocabulário difícil, uso de termos técnicos, uso de gírias, expressões idiomáticas etc.) ou se não entenderam as estruturas gramaticais por essas serem utilizadas principalmente em etapas posteriores de sua aprendizagem — **sujeitos são considerados de nível básico e as estruturas são apresentadas em níveis avançados, por exemplo.**

Assim como Mackey e Gass (2005), sugerimos que as instruções dos TM sejam traduzidas para a língua materna dos sujeitos (oralmente por quem está aplicando o teste ou escrita abaixo da instrução em língua estrangeira), principalmente se os informantes tiverem conhecimento básico ou intermediário da língua alvo. O melhor é que se adapte o teste ao nível de conhecimento dos sujeitos. Deve-se observar cautelosamente o uso de certas preposições, verbos, conjunções, advérbios, certas regências verbais e nominais que podem alterar facilmente alguns resultados.

Condição 2

- ✓ É preciso que seja definido claramente, nas instruções do teste, o que os pesquisadores entendem por gramatical, não gramatical, aceitável e não aceitável.

Esta condição está diretamente relacionada com as instruções dadas aos sujeitos. Deve haver um treinamento antes do início do teste propriamente dito e, neste momento, quem está aplicando o teste deve dar exemplos aos sujeitos, mostrando o que os pesquisadores esperam de suas respostas. Segundo Schütze (1996), muitas vezes os sujeitos não sabem exatamente o que fazer e como responder o teste. Para que essa condição seja seguida de modo adequado, alguns pesquisadores devem se apropriar melhor dos termos “gramaticalidade” e “aceitabilidade”. Não podemos esperar que, somente pelo fato de essas palavras estarem escritas em um teste, estaremos sendo claros quando os informantes são *naive*.³ Aliás, Cowart (1994) observa que até mesmo alguns linguistas desconhecem a diferença entre aceitabilidade e gramaticalidade. Chomsky (1965) parece concordar com esse autor ao afirmar que se os pesquisadores não forem suficientemente diretos nas instruções dos TM, os sujeitos responderão o que quiserem e, dessa forma, seus resultados não podem ser considerados satisfatórios. Bley-Vroman, Felix e Ioup (1988) perceberam que muitas respostas inconsistentes em seu teste piloto foram ocasionadas porque as instruções não estavam claras nem objetivas. Quando esses autores formularam instruções mais detalhadas, acompanhadas de exemplos, seus resultados foram muito mais satisfatórios.

Schütze é bastante categórico quanto a essa questão: “In fact, if we were to ignore all studies in which we believe the instructions to subjects were inadequate to convey the subtlety of a linguistic definition, the remaining studies could likely be counted on one hand.” (SCHÜTZE, 1996, p.133)⁴.

³ Um sujeito *naive* (*naive subject*) é aquele que não tem conhecimentos de linguística, ou seja, não é um linguista ou um estudante de linguística.

⁴Na verdade, se ignorássemos todos os estudos nos quais acreditamos que as instruções aos sujeitos estejam inadequadas ao atender as sutilezas de uma definição linguística, os estudos que restariam caberiam em uma mão.”

Conforme esse autor, muitos testes demonstram terem sido bastante ineficazes porque suas instruções não foram suficientemente claras. Infelizmente, não há como dizermos aos pesquisadores que temos um método ou um modelo de instruções que funcionarão para qualquer teste porque isso é impossível. Sugerimos que as instruções dos testes não sejam muito breves ou vagas e que sempre haja, junto com elas, exemplos de como vai ser o funcionamento do teste e de como as questões podem ser respondidas. As instruções devem ser específicas e, de preferência, devem esclarecer por que motivo uma sentença é considerada ruim, não aceitável ou não gramatical. Para um melhor entendimento por parte dos sujeitos, pode-se ler ou escrever em um quadro, ou ainda ter frases, no próprio teste, que sejam exemplos claros do que é uma frase gramatical e do que é uma frase não gramatical, por exemplo. Conforme Bley-Vroman, Felix e Ioup (1988), o ideal é sempre, antes do início da aplicação do teste, dar uma folha com instruções aos sujeitos, na qual contenha explicações detalhadas sobre o que significa dizer que uma sentença é possível. Nessa explicação, deve estar claro que a noção de “possível” está mais ligada a um sentimento do que a uma questão de ter conhecimento de regras explícitas⁵. Birdsong (1989) lembra que as noções de gramaticalidade e aceitabilidade podem ser usadas como sinônimos para muitos linguistas, embora para alguns o mais adequado seja usarmos algo como: “Esta sentença parece correta ou não para você?” Mesmo assim, é evidente que esse tipo de estratégia pode não funcionar: é comum acontecer de os sujeitos rejeitarem sentenças em que apareçam palavras que eles não compreendam. Birdsong (1989, p. 114-115) oferece alguns exemplos do que ele considera instruções mal formuladas:

- a) As sentenças seguintes lhe parecem corretas?
- b) Diga-me se para você isso é uma sentença.⁶

Birdsong destaca que, na primeira, os sujeitos têm problemas para diferenciar as sentenças que apresentam problemas gramaticais das sentenças que possam ser semanticamente estranhas. O problema da segunda instrução está no fato de que orações como “Quando tua tia vai chegar?” podem ser rejeitadas pelos sujeitos simplesmente por ser uma interrogação e não uma afirmação.

Mackey e Gass (2005) sugerem que os pesquisadores também devem estar atentos ao tamanho das sentenças nos TM: para esses autores, elas não podem ser muito longas para que não excedam a memória de trabalho (*working memory* ou *short term memory*). Eles sugerem orações que não tenham mais de nove ou dez palavras.

Condição 3

- ✓ Um teste de julgamento não deve ser muito longo. Além disso, deve haver número equilibrado de sentenças “aceitáveis” e “não aceitáveis”, e também um número suficiente de *fillers*.

Uma das preocupações frequentes na preparação de um TM é quanto ao seu número de sentenças. É comum vermos testes que variam entre dez e mais de duzentas orações. Percebemos que o ideal seria que os testes tivessem no mínimo trinta sentenças — com léxico bem diversificado — **em** máximo em torno de sessenta, pois há um risco muito grande de os sujeitos ficarem cansados e, por isso, responderem o teste sem prestar muita atenção ao que estão lendo.

Mackey e Gass (2005) são bastante incisivos quanto a isso: para eles, o número de questões não pode ser muito elevado e deve haver um balanço entre sentenças gramaticais e não gramaticais, pois é evidente que os sujeitos ficam cansados após algum tempo. Para que os resultados sejam válidos, esses autores sugerem que os testes tenham no máximo cinquenta questões, afirmando que “Although studies have been carried out with as many as 101 (HAWKINS; CHAN, 1997) e 282 (JOHNSON; NEWPORT, 1989), we recommended no more than approximately 50 sentences.” (MACKEY; GASS, 2005, p. 51)⁷.

⁵ O que queremos dizer é que os sujeitos devem entender que eles não precisam saber nomear as partes da frase, ou seja, eles não precisam saber, por exemplo, se uma palavra está funcionando como adjunto ou complemento.

⁶ “Do the following sentences sound right?” e “Tell me if for you this makes a sentence”.

⁷ Embora alguns estudos tenham sido conduzidos com 101 (HAWKINS; CHAN, 1997) e 282 (JOHNSON; NEWPORT, 1989) sentenças, nós não recomendamos mais de 50 sentenças aproximadamente.”

Schütze (1996) é bastante categórico quanto ao número de sentenças gramaticais e aceitáveis ou não gramaticais e não aceitáveis que devem ser colocadas em um TM; o autor advoga que deve haver um número bem equilibrado entre os dois tipos de sentenças. Ele também propõe que “[...] the lexical content of the sentences should be varied to guard against the influence of imagery, and any other potential biases of lexical items, such as word length, frequency, and semantic peculiarities”⁸. Além disso, é pertinente que coloquemos nessa condição a questão dos *fillers*. Conforme autores como Birdsong (1989) e Schütze (1996), o ideal é que todo teste contenha uma porcentagem entre 33% e 50% de *fillers* para que, dessa forma, os respondentes não tenham dicas – pelo menos aparentes – do que está sendo trabalhado pelos pesquisadores.

Condição 4

- ✓ O tempo para que os sujeitos respondam o teste deve ser controlado pelos pesquisadores.

Muitos autores (BIALYSTOK, 1979; BIRDSONG, 1989; ELLIS, 1991; COOK, 1993; SCHÜTZE, 1996; MACKEY; GASS, 2005), que trabalham com a questão do uso dos TM, mencionam a importância da questão do tempo total para os sujeitos responderem o teste e, preferencialmente, que se controle o tempo para cada questão respondida. Logicamente, é muito mais fácil controlarmos esse fator quando o teste é oral ou realizado no computador. Quando o teste é escrito, isso fica um pouco mais difícil, embora não seja impossível. O ideal é definirmos o tempo que os sujeitos têm para realizar o teste.

Greenbaum (1977) tem posicionamento interessante sobre o tempo para realização do teste: para o autor, os sujeitos devem ter tempo suficiente para entender e julgar as sentenças, mas não podem ter tempo em excesso. Segundo ele, quanto mais tempo é disponível para os sujeitos, mais erros eles encontram nas sentenças. O autor argumenta que: “The more they look, the more they find wrong things on them”⁹. (GREENBAUM, 1977, p. 205).

Embora essa afirmação seja bastante interessante, o autor não explica claramente o porquê de pensar dessa forma. Entendemos que a quantidade de tempo para a realização do TM deve ser decidida pelo pesquisador com base no seu bom senso e na sua experiência. Se os sujeitos forem iniciantes na aprendizagem da L2 em questão, eles precisarão de mais tempo para a leitura, compreensão e julgamento das questões. Com certeza, a realização do teste piloto pode ajudar nessa decisão.

Ellis e Barkhuizen (2005) afirmam que, a partir do exame do fator tempo nos julgamentos de TM, é possível especificar até que ponto certas estruturas serão julgadas gramaticais ou não gramaticais pelos sujeitos. Os autores também sugerem que os sujeitos, tanto falantes de língua materna quanto de uma segunda língua, levam mais tempo para julgar as sentenças não gramaticais do que as gramaticais, mas infelizmente não há evidências suficientes sobre a legitimidade desses argumentos¹⁰.

Condição 5

- ✓ Para que o teste tenha maior confiabilidade, sempre deve haver um teste piloto com falantes de língua materna e também com aprendizes de L2.

Embora esta condição possa parecer bastante óbvia, é comum observarmos a não menção dessa etapa em muitos artigos, teses e dissertações. É possível que pesquisadores não explicitem como foi feito seu teste piloto ou porque pensam que ele não tem importância suficiente para ser mencionado no trabalho ou porque realmente não o realizaram. Sugerimos que haja, em qualquer pesquisa, algumas informações com dados sobre quantos sujeitos responderam o piloto. Contudo, acreditamos que para ter maior legitimidade, toda a pesquisa em L2 que faz uso de TM deveria ser precedida de um teste com um grupo de controle composto por falantes de língua materna (entre dez e quinze sujeitos). Embora concordemos que isso seria um fator bastante útil para percebermos

⁸O conteúdo lexical das sentenças deve ser variado para que haja precaução contra a influência de imagens (ideias) e quaisquer outros pré-conceitos potenciais de itens lexicais, tais como tamanho das palavras, frequência, e peculiaridades semânticas.”

⁹“Quanto mais eles olham as questões, mais erros encontram.”

¹⁰ Para maiores detalhes acerca desse assunto, consultar Birdsong (1989) e também Mackey e Gass (2005).

se há sentenças menos naturais para falantes de L1, essa condição só é fácil de ser executada em países onde há número suficiente de estudantes estrangeiros ou imigrantes falantes da língua alvo.

Condição 6

- ✓ O TM terá problema de legitimidade se não apresentar, no mínimo, 60 informantes.

Esta questão é uma das mais controversas entre os autores que utilizam testes gramaticais escritos e também orais. Flanigan (1995) utilizou 23 sujeitos, falantes de inglês como L2; Myles, Mitchell e Hooper (1999), dezesseis sujeitos falantes de francês como L2; Mackey (1999), 34 falantes de inglês como L2; Juffs (1996), 105 falantes de inglês como L2, 22 falantes de chinês como língua materna e 19 falantes de inglês como L1; Inagaki (2001) obteve ajuda de 22 falantes de inglês como L1, 42 falantes de japonês como L1 e 63 falantes de inglês ou japonês como L2. Ainda há autores como Bley-Vroman e Yoshinaga (1992) que utilizaram praticamente o mesmo número de informantes falantes de L1 e falantes de L2 — **84 falantes de japonês como L2 e 85 falantes de inglês como língua materna**. Nossa lista poderia ser imensa e a quantidade de sujeitos extremamente variável. Percebemos, porém, que o que mais importa é o número mínimo de sujeitos que irão responder o teste. O número máximo, na realidade, parece não ser um problema para os resultados da pesquisa, se os números finais forem adquiridos a partir de cálculos e percentagens exatas. Assim, concordamos com Birdsong (1989) e Schütze (1996): quando possível, deve haver um número mínimo de oitenta sujeitos e, no caso de pesquisas em L2, esse número deve ser dividido em diferentes níveis de aprendizagem. Nossa condição 6 exige em torno de sessenta sujeitos, porque julgamos que um número menor do que esse não é satisfatório e, por conseguinte, não traz credibilidade à pesquisa realizada.

3.2 CONDIÇÕES SECUNDÁRIAS

Condição 7

- ✓ Nem sempre é aconselhável pedir aos informantes que expliquem suas respostas.

Bialystok (1997), Birdsong (1989), Schütze (1996) e Ellis e Barkhuizen (2005) são alguns dos autores que discorrem sobre a possibilidade de perguntarmos para os sujeitos das pesquisas que utilizam TM o porquê de eles terem respondido as questões de tal forma. Como outros fatores que envolvem o tema, aqui também existem opiniões controversas.

Segundo Bialystok (1997) e Ellis e Barkhuizen (2005), os aprendizes de uma L2 possuem conhecimento implícito sobre muitas estruturas e nem sempre sabem explicar o porquê de terem falado, escolhido ou escrito alguma coisa. Ellis e Barkhuizen (2005) argumentam que, na maior parte das vezes, os sujeitos dirão que estavam cansados, entediados ou simplesmente que não sabem explicar por que escolheram tal resposta. Esses autores lembram que devemos estar atentos para a distinção que há entre processar o conhecimento linguístico e conseguir verbalizá-lo. E isso, de certa forma, não está vinculado somente ao conhecimento de uma L2: possivelmente um falante de sua língua materna também não saberia explicar algumas de suas escolhas em um TM sobre sua própria L1.

Bialystok (1979) discute a importante questão sobre mudança no uso do conhecimento implícito por aprendizes de uma L2, em suas interlínguas: na medida em que eles vão ficando mais fluentes, passam a utilizar menos conhecimento explícito e são mais rápidos para responderem as tarefas. Bialystok (1979) reconhece haver três grupos distintos que utilizam seu conhecimento implícito com mais eficácia: crianças, aprendizes avançados e aprendizes que falam outras línguas estrangeiras¹¹.

Mesmo entendendo que as diferenças entre conhecimento explícito e implícito são discutidas de diferentes formas, por vários autores (BIALYSTOK, 1997; COOK, 1993; ELLIS, 1994; TOWELL; HAWKINS, 1994) e que suas linhas divisórias não são suficientemente claras, percebemos que a decisão do pesquisador em perguntar sobre as respostas dadas pelos sujeitos deve ser pensada com muita cautela.

¹¹ Para maiores detalhes sobre julgamentos implícitos em L2 e também sobre diferenças entre aprendizes, ver Bialystok (1979).

Condição 8

- ✓ É preferível que os sujeitos não sejam linguistas ou estudantes de linguística.

Essa é outra questão citada na maior parte dos trabalhos que discutem os TM. Embora autores como Hawkins (2001) admitam que, tanto nos seus testes piloto como nos seus TM válidos, utilizam respostas dadas por outros linguistas ou por estudantes da área, autores como Birdsong (1989) e Ellis (1991; 1994) condenam veementemente os pesquisadores que não utilizam somente sujeitos *naive* em seus experimentos. O motivo de suas ressalvas é compreensível: por terem uma visão da língua como ciência, os linguistas podem observar detalhes, nos TM, que não são observados por outros sujeitos e, assim, suas respostas podem influenciar de alguma maneira os resultados da pesquisa. Além disso, é provável que muitos linguistas — principalmente quando já conhecem o pesquisador e já sabem com qual assunto este trabalha — percebam exatamente qual é o foco de estudo da pesquisa, mesmo com a adição de *fillers* no teste. Conforme Schütze (1996), é provável que linguistas sejam inconscientemente preconceituosos em relação a certas estruturas devido a suas posições teóricas, podendo fazer julgamentos de acordo com suas versões do que é ou não gramatical¹².

Contudo, na opinião de Schütze (1996), não podemos ser categóricos quanto a essa questão, assegurando que os julgamentos de linguistas serão sistematicamente diferentes de julgamentos de outros sujeitos. O autor salienta que pesquisas como as de Spencer (1973), Snow e Meijer (1977), e Greenbaum (1977) mostram que há muita controvérsia quanto a isso. Porém, parece-nos claro que esses autores percebem que não deveríamos confiar nas respostas de um grupo de sujeitos composto somente por linguistas.

Embora muitos desses pesquisadores tenham realizado estudos em Aquisição de L1, as mesmas preocupações devem ser consideradas em estudos de L2 e, em nossa opinião, talvez o mais indicado seria que os sujeitos com conhecimentos teóricos sobre a linguagem só respondessem o teste piloto. Assim, podemos evitar algumas dúvidas quanto aos resultados, como também críticas quanto à metodologia e à validação dos TM.

Condição 9

- ✓ Ao utilizar diferentes tipos de testes, a comparação dos resultados deve ser feita com extrema cautela.

Chaudron (1983; 2003) afirma que uma das melhores maneiras de tornarmos uma pesquisa válida é usarmos diferentes tipos de instrumentos. Apesar de confiar na legitimidade e na importância dos testes de julgamento, em uma pesquisa com o uso de um TM, é interessante que os pesquisadores também realizem outro teste com exercícios de tradução, ou de compreensão, de imitação, ou ainda de identificação de figuras, para poder comparar os resultados obtidos. Dificilmente é possível obtermos conclusões confiáveis através de somente um tipo de teste, sugere o autor. Chaudron (2003) assegura que esse é um dos principais motivos pelos quais certos pesquisadores, tais como Doughty e Long (2000), estejam usando mais de um tipo de tarefa: “This is why many researchers today employ multiple measures, in order to ‘triangulate’ their findings, and to differentiate the possible effects of the method employed from the stable or developing traits of the learners’ underlying capacity.” (CHAUDRON, 2003, p. 766)¹³.

Por outro lado, Birdsong (1989) não concorda com a ideia de “quanto mais evidências, melhor”, pois é necessário muito cuidado quando os pesquisadores têm a intenção de comparar resultados provenientes de diferentes fontes. O autor enfatiza que cada método possui características que não podem ser simplesmente transferidas para outros métodos ou teorias. Bialystok e Ryan (1985) concordam com Birdsong (1989), pois declaram que as análises de diferentes conhecimentos linguísticos variam de tarefa para tarefa, tornando a comparação entre resultados bastante difícil e discutível. Sendo assim, se o pesquisador realmente desejar cruzar os resultados de um TM com outro tipo de teste, isso deve ser feito com muito rigor. Sugerimos também que o pesquisador pilote

¹² Obviamente, não podemos deixar de considerar que qualquer sujeito terá seus preconceitos e suas próprias versões do que é ou não gramatical. Esse é um dos motivos pelos quais se deve ter um número mínimo de sujeitos respondendo o teste.

¹³ Este é o motivo pelo qual muitos pesquisadores, hoje, aplicam múltiplas medidas, com o objetivo de “triangularizar” seus resultados, bem como para diferenciar os possíveis efeitos do método aplicado, a partir das características estáveis ou em andamento da capacidade de linguagem subjacente dos aprendizes.

os dois testes e compare seus resultados, buscando perceber até que ponto isso acrescenta dados relevantes nos resultados encontrados.

Condição 10

✓ O pesquisador deve ter cautela ao optar por julgamentos binários ou julgamentos escalares em um Teste de Julgamento Gramatical

Decidir que tipos de respostas serão oferecidas aos sujeitos não é tarefa das mais fáceis. Segundo alguns autores (BIRDSONG, 1989; SCHÜTZE, 1996; SORACE, 1996; COWART, 1997; MACKEY; GASS, 2005), devemos considerar vários fatores antes de optarmos por julgamentos absolutos ou binários (aqueles que possuem somente duas opções de respostas, tais como “gramatical x não gramatical”, “aceitável x não aceitável”, “bom x ruim” etc.) ou por julgamentos escalares (aqueles que possuem graus de gramaticalidade, tais como “não sei”, “não tenho certeza”, “provavelmente correto”, “provavelmente incorreto” etc.). Na verdade, é surpreendente o quanto as opiniões desses autores diferem entre si. Schütze (1996), por exemplo, mesmo não sendo contra o uso dos julgamentos escalares, aconselha que, se o pesquisador quiser adotá-los, ele deve ser bastante cuidadoso, pois um número grande de opções pode confundir os sujeitos e, além disso, poderá ser mais difícil fazer os cálculos estatísticos dos resultados.

Schachter e Yip (1990), por outro lado, oferecem quatro opções de respostas em sua pesquisa — claramente gramatical, provavelmente gramatical, provavelmente não gramatical e claramente não gramatical — e concluem que, em seu estudo, a escala de quatro opções ajudou muito seus sujeitos a mostrar seu grau de incerteza quando deparam com questões mais complexas. Chaudron (1983) e Sorace (1996) também usaram julgamentos escalares e concluíram que as opções de respostas ajudaram seus TM a se tornarem mais confiáveis e válidos. Embora Sorace (1996) tenha destacado os aspectos positivos do uso dos julgamentos escalares, ela observa que os sujeitos optam demasiadamente pelas respostas consideradas neutras, ou seja, diante da primeira dúvida, escolhem respostas como “não sei” e “provavelmente”.

De acordo com Birdsong (1989), tanto os julgamentos binários quanto os escalares possuem vantagens e desvantagens, mas, em sua opinião, os testes que apresentam mais opções de respostas refletirão mais detalhes e sutilezas sobre a estrutura estudada. Contudo, o autor não é favorável a um número de opções maior do que cinco. Na verdade, percebemos que essa questão ainda está longe de ser resolvida: a discrepância que há entre as opiniões da maioria desses autores é digna de atenção, pois nos mostra como essa questão ainda carece de uniformidade.

Autores como Juffs (1996) e Hawkins e Chan (1997) não são a favor do uso de respostas como “não sei” ou “não tenho certeza”, mesmo quando a opção é por escalas de somente três ou quatro respostas. Verificamos que a maior parte dos autores que escrevem sobre os testes metalingüísticos (BIRDSONG, 1989; GASS, 1994, 2003; SCHÜTZE, 1996; HAWKINS; CHAN, 1997; MACKEY; GASS, 2005) concordam que os julgamentos ocorrem em um contínuo, não estaticamente. Chomsky (1965, p. 10) observou que “Obviously, acceptability will be a matter of degree, along various dimensions.”¹⁴. Talvez esse seja um dos motivos pelos quais esses autores sugerem que haja mais de duas opções de respostas nos TM. Contudo, não encontramos homogeneidade quando esses autores sentenciam o número de opções e o que deve ser escrito nas alternativas — gramatical, não gramatical, correto, provavelmente correto, aceitável, não-aceitável, não sei etc.

Nossa sugestão, então, é que os pesquisadores procurem usar o bom senso, observando o que é mais apropriado, especificamente para o seu TM e para a estrutura linguística que estejam trabalhando, porém não ultrapassando o número de cinco opções de respostas.

¹⁴ “Obviamente, aceitabilidade será um caso de graduação, através de várias dimensões.”

Condição 11

- ✓ Se possível e conveniente, o teste pode ser aplicado duas vezes aos mesmos sujeitos.

Schütze (1996) afirma que, para os TM terem maior utilidade em pesquisas de ASL, os pesquisadores podem usar diversos resultados experimentais e tentar alocá-los em um modelo coerente. De acordo com o autor, é preciso muito esforço para que tenhamos resultados exatos e consistentes através desses testes. Uma das maneiras de se conseguir esses resultados pode ser através da aplicação do teste mais de uma vez. Com o objetivo de encontrar resultados mais rigorosos, Chaudron (1983) recomenda que os mesmos sujeitos respondam a dois instrumentos, com diferença de pouco tempo entre um e outro, por volta de uma ou duas semanas. No caso das pesquisas em L2, temos um agravante: se no período entre um teste e outro os sujeitos realizam exercícios abordando direta ou indiretamente a estrutura pesquisada, eles podem mudar de opinião quanto às questões respondidas em um momento anterior. Por outro lado, os aprendizes podem ter deixado de utilizar certas estruturas e, dessa forma, terem mais dúvidas quanto a suas respostas na segunda vez que respondem o exercício.

Gass (1994) reaplicou um TM e concluiu que a maior parte das informações dadas na primeira aplicação do teste continuou a mesma na segunda e, por esse motivo, a autora destaca a confiabilidade inerente a um teste bem formulado. Assim, para ela, quando usados de modo apropriado, os TM são uma ferramenta válida e confiável para os estudos da ASL. Ellis (1991) é totalmente contrário a essa ideia por não ter encontrado resultados semelhantes aos de Gass em sua pesquisa. Os sujeitos que responderam o instrumento de Ellis, em duas oportunidades, mostraram muito mais inconsistência do que os sujeitos da pesquisa de Gass. Porém, nesse caso, é muito complicado fazermos uma comparação direta entre os estudos desses dois autores, pois eles utilizaram estruturas gramaticais e sujeitos diferentes. Dessa forma, talvez seja melhor não adotarmos a reaplicação do instrumento enquanto não haja mais estudos que revelem se isso pode ou não afetar a validação e a confiabilidade dos TM. Se o pesquisador quiser realmente reaplicar seu instrumento, sugerimos que isso seja feito, no máximo, após duas semanas.

Condição 12

- ✓ Pesquisadores devem ser muito prudentes ao escolher as sentenças do teste, pois muitas vezes confundem sentenças não-gramaticais com sentenças nas quais os sujeitos têm dificuldade de processamento.

Conforme Schachter e Yip (1990), os julgamentos gramaticais são resultado tanto de fatores processuais quanto gramaticais. Como não há como separar esses agentes, pois eles estão o tempo todo interagindo e, muitas vezes, pode haver confusão entre sentenças agramaticais e sentenças com problemas de processamento (*parsing*). As autoras asseguram que os sujeitos podem fazer julgamentos equivocados devido a muitas razões: a sentença pode ser semanticamente ou pragmaticamente estranha ou ainda ter alguma expressão não aceita nos dialetos de prestígio, por exemplo. Para elas, quando os pesquisadores escolhem seus exemplos sem prestarem atenção em possíveis problemas de processamento que possam dificultar a realização do teste, há possibilidade de haver equívocos na avaliação dos resultados. Esse é o caso das *garden-path sentences*, que são sentenças sobre as quais há, comumente, dificuldade de se fazer julgamentos gramaticais, mesmo quando os sujeitos estão respondendo a um instrumento na sua L1. Vejamos alguns exemplos:

a) The guy who they don't know whether he wants to come or not.

(O homem que eles não sabem se quer ou não quer vir.)

b) Which book did you say John believes offended many people?

(Qual livro que você disse que John acha que ofendeu muitas pessoas?)¹⁵

Os exemplos acima podem ser avaliados, por um leitor desatento, como sendo agramaticais, pois há certa dificuldade em processá-los, principalmente para aprendizes de inglês como L2. Se esses sujeitos tiverem níveis de conhecimento linguístico básico ou

¹⁵ Esses exemplos foram extraídos de Schütze (1996, p. 31).

intermediário, é muito provável que considerem essas sentenças estranhas e, conseqüentemente, não as aceitem como gramaticais ou corretas.

Conforme Schütze (1996), dependendo do estudo em questão, os pesquisadores devem evitar colocar sentenças tipo *garden-path sentences* em seus instrumentos, pois é muito provável que os sujeitos as interpretem como sentenças “ruins”, não-aceitáveis ou agramaticais. É claro, continua o autor, que isso não se aplica quando o estudo tem interesse em averiguar como os sujeitos processam sentenças em sua L1 ou L2. Contudo, quando o propósito do estudo é saber se algum tipo de estrutura é ou não aceito pela gramática de uma língua, não há razão alguma para confundir os sujeitos. Schütze também declara que sentenças com problema de *parsing* são, sem dúvida, casos em que há maior evidência de que o tempo seja fator determinante nos julgamentos. Além disso, o autor observa que a questão do processamento está diretamente relacionada com as correções feitas pelos sujeitos às sentenças e faz a seguinte afirmação: “the closer a bad sentence is to satisfying all the parser’s constraints, the easier it will generally be to correct.” (SCHÜTZE, 1996, p. 164)¹⁶.

Por esses motivos, nossa sugestão aqui é bastante categórica: a não ser que o estudo seja sobre processamento, deve-se evitar quaisquer sentenças do tipo *garden-path sentences*.

Ainda que tenhamos apresentado doze condições, divididas entre primordiais e secundárias, concordamos que pode haver outras condições importantes para esse fim que não foram abordadas nesta pesquisa. Mesmo assim, acreditamos que, se os pesquisadores levarem em conta as condições por nós apresentadas, suas pesquisas terão menos problema com as questões de validação e de confiabilidade, aspectos largamente discutidos na área da Aquisição da Linguagem.

A seguir, faremos uma breve análise de alguns TM utilizados por alguns dos autores da área de ASL.

4 OBSERVAÇÕES SOBRE TESTES METALINGÜÍSTICOS ELABORADOS PARA PESQUISAS EM L2

Uma das dificuldades para escrever esta seção foi encontrar testes metalingüísticos – no corpo do texto ou em anexo –, em artigos e capítulos de livros sobre aquisição de L2. Na maior parte desses trabalhos, encontramos somente pequenos extratos dos TM. Analisaremos então, principalmente pelo espaço ser restrito, três estudos de autores com vasta experiência em pesquisas em L2¹⁷.

Teste 1 – Ellis (1991)

Neste artigo, o autor tem o objetivo de analisar criticamente o uso dos testes de julgamento em ASL. Além de examinar teorias que dão suporte à construção desses testes, o autor apresenta sua pesquisa realizada com 21 aprendizes avançados de inglês, falantes de chinês como primeira língua. Seu intuito é verificar se os sujeitos sabem utilizar *dative alternation*¹⁸. Dois testes foram aplicados: o primeiro escrito, com todos os sujeitos; o segundo, oral, com apenas oito. Surpreendentemente, esse foi um dos únicos trabalhos em que encontramos, nos anexos, os dois testes na íntegra. Ellis é um autor que, além de utilizar os TM em suas pesquisas, também tem trabalhos que abordam a questão da utilidade e da validação desses testes. Por isso, não nos causa surpresa que os testes que ele apresenta são meticulosamente elaborados, com instruções e sentenças objetivas e de fácil entendimento. O número de questões a serem julgadas totalizam quarenta e há somente três opções de respostas: gramaticalmente incorreta (x), gramaticalmente correta (c) ou não tenho certeza (?). Além disso, é solicitado que os sujeitos reescrevam as sentenças que julgaram incorretas.

¹⁶ Quanto mais próxima uma sentença ruim estiver de satisfazer todas as restrições de processamento, mais fácil será de corrigi-la.

¹⁷ Para melhor entendimento desta seção, sugerimos a leitura dos artigos analisados na íntegra.

¹⁸ Verbos dativos, em língua inglesa, como give (dar), sell (vender) e buy (comprar), aceitam duas formas. Uma das formas é preposicionada (ex.: Mark sent a book to John – Mark enviou um livro a John) e a outra, sem preposição (ex.: Mark sent John a book – Mark enviou a John um livro).

Exemplos de sentenças extraídas da Parte 1 (Ellis, 1991)

- My wife sent me a letter. ()
 (Minha mulher me mandou uma carta.)
- The president offered him a job. ()
 (O presidente ofereceu um emprego a ele.)
- The policeman reported me the accident. ()
 (O policial me relatou o acidente.)
- The teacher explained me the answer. ()
 (O professor me explicou a resposta.)
- My mother bought me a shirt. ()
 (Minha mãe me comprou uma saia.)
- The manager reserved me a seat. ()
 (The manager me reservou uma mesa.)
- Mr. Smith opened the door to me. ()
 (O Sr. Smith abriu a porta para mim.)
- Mrs. Jones designed her a dress. ()
 (A sra. Jones criou uma saia pra ela.)

Observando as orações do primeiro teste, percebemos que o autor não incluiu *fillers* em seu instrumento. Por esse motivo, é muito provável que os sujeitos tenham percebido que o objetivo do pesquisador era investigar o conhecimento sobre *dative alternation*. Nesse caso, parte da condição 3, considerada primordial, não foi atendida. Ademais, não é possível saber, pelas instruções, se o tempo de execução do teste foi limitado (condição 4).

Observações sobre a Parte 2 (Ellis, 1991)

Uma semana após a realização da Parte 1, oito aprendizes receberam a Parte 2, com um pequeno agradecimento por terem completado o teste anterior e com novas instruções para a segunda parte. As instruções são muito semelhantes nos dois testes, embora neste só haja dez questões que se dividem entre uma atividade de audição e outra para a qual também foi solicitado que os sujeitos gravassem suas respostas em um aparelho.

O autor, ao final do artigo, ratifica que é necessário muito cuidado ao usar testes metalinguísticos. Não há dúvidas que seu teste foi elaborado com prudência. Não compreendemos, porém, o motivo de o autor ter incluído somente 21 sujeitos na pesquisa, o que fere a condição 6, que dita que o ideal é que haja o mínimo de sessenta respondentes. Mesmo assim, é um artigo muito importante de ser consultado por quem deseja aplicar um TM, tanto em L1 como em L2.

Teste 2 – Juffs (1996)

Nesta pesquisa, Juffs sugere que línguas podem ter parâmetros lexicais, que o sentido de certos verbos pode incluir causa e mudança de estado. Em inglês, causa e estado podem estar em um mesmo verbo, porém essa sobreposição de sentido é impossível em chinês. Exemplos em inglês podem ser verificados em verbos que apresentam estado psicológico, como *anger* (odiar), *disappoint* (desapontar, frustrar) e *frighten* (amedrontar). O verbo desapontar, por exemplo, pode significar que algo fez com que alguém ficasse frustrado ou desapontado, ou seja, há causa e estado em um mesmo verbo. O objetivo do autor foi avaliar como os aprendizes chineses aceitavam orações com esses verbos em inglês, já que em sua língua materna não há esse parâmetro.

No apêndice do artigo, o teste é apresentado por inteiro, com as instruções dadas aos sujeitos, explicando que falantes de uma L2 parecem desenvolver um *feel* (palavra usada e grafada pelo autor) para saber o que é possível em uma sentença. Há quatro exemplos para que os sujeitos compreendam a ideia do autor.

Teste 3 – Tremblay (2005)

A pesquisa de Tremblay tem o objetivo de criar um teste de julgamento que siga as orientações de Schütze (1996) e não tivesse os problemas apontados por autores como Sorace (1996) e Chaudron (2003). A autora administrou um TM com oitenta questões a vinte sujeitos, falantes de inglês americano. Os resultados mostraram que os participantes julgaram como gramaticais a maior parte das orações, mesmo quando eram agramaticais. Como a pesquisa fazia parte de outra de maior dimensão sobre conhecimento gramatical e restrições nas interlínguas, a autora avaliou todas as questões elaboradas para o teste que foi respondido *online*, com tempo de no máximo vinte segundos para cada resposta. Após esse tempo, a tela passava automaticamente para outra questão e os sujeitos não podiam retornar para a tela anterior (condição 4).

Nas instruções do teste, há explicações detalhadas sobre o que a pesquisadora considera intuição linguística:

In this task, you will be asked to judge whether individual sentences sound good or not TO YOU. A sentence sounds good if you think you would or could say it under appropriate circumstances. By contrast, a sentence sounds bad if you think you would never say it under any circumstances. Here we are interested in your linguistic intuitions only, not in the rules of “proper English” that you might have been taught in school. (TREMBLAY, 2005, p. 166)¹⁹

Após, a autora explica os procedimentos do teste, no qual o sujeito terá que escolher uma das quatro opções abaixo:

Perfect	Okay	Akward	Horrible
4	3	2	1
(perfeito)	(ok)	(estranha)	(horrível)

Se os sujeitos não souberem ou não tiverem nenhuma intuição, a autora pede que não tentem acertar e que pressionem a tecla X, que na pesquisa quer dizer “No intuition” (sem intuição). Observamos que Tremblay seguiu nossa condição 10, pois cinco opções de respostas, bem diferentes umas das outras, auxiliam os sujeitos a optar por uma única resposta, embora talvez ela pudesse utilizado somente três: *Okay*, *Akward*, e *Horrible*.

Apesar de as instruções terem sido bem elaboradas e com descrições detalhadas, a autora peca por repetir informações e grifar várias vezes expressões como “DO NOT GUESS”, “IT IS CRUCIAL”, “TO YOU”, “ESSENTIAL”, entre outras. São duas páginas inteiras de instruções e exemplos que poderiam ter sido resumidos, talvez sem a necessidade de usar caixa alta e de repetir palavras e instruções.

Além disso, um teste de oitenta sentenças é considerado longo, segundo nossa condição 3 e, conforme a própria autora, o número de informantes (somente vinte) foi o maior problema de sua pesquisa (condição 6). Ademais, em todo o artigo, não é dito se houve ou não um teste piloto, que é uma das sugestões feitas por Schütze (1996). O mais interessante acerca do trabalho de Tremblay é ela ter declarado que seu objetivo principal era produzir um TM que não tivesse os problemas encontrados em artigos da área de aquisição da linguagem. Porém, mesmo assim, ela não conseguiu seguir todas as recomendações de Schütze, o que nos mostra quanto é difícil e complexo produzir um bom teste metalinguístico.

¹⁹ Neste teste, será solicitado que você julgue se sentenças parecem boas ou não PRA VOCÊ (grifo da autora). Uma sentença é boa se você acha que poderia usá-la em certas circunstâncias. Ao contrário, a sentença é ruim se você nunca a usaria sob nenhuma circunstância. Aqui, nós estamos interessados somente nas suas intuições linguísticas, não nas regras do “inglês padrão” que você deve ter aprendido na escola.

5 CONSIDERAÇÕES FINAIS

Apesar de ainda haver dúvidas quanto à eficiência dos TM, nosso propósito neste artigo foi o de mostrar evidências de que esses testes podem ser bons instrumentos na obtenção de resultados em pesquisas de ASL. Os TM carecem de elaboração cuidadosa, e, para que isso seja possível, propusemos uma série de condições fundamentadas em estudos de vários autores, tanto por aqueles que pesquisaram especificamente sobre a utilização dos TM (CHAUDRON, 1983; BIRDSONG, 1989; ELLIS, 1991; GASS, 1994; SCHÜTZE, 1996; SORACE, 1996; TREMBLAY, 2005), quanto por outros que os usaram em seus estudos (JUFFS, 1996; WHITE; GENESEE, 1996; HAWKINS; CHAN, 1997; INAGAKI, 2001).

Esclarecemos, mais uma vez, que estamos cientes de que os GJTs não são um método eficiente para todos os tipos de pesquisas na área de aquisição de uma segunda língua, embora concordemos com Gass (1994) e com Sorace (1996) que eles podem ser, com certeza, válidos e eficientes, ajudando muito a entendermos melhor como ocorre a aquisição de uma segunda língua. Não há como negar que os resultados encontrados nessas pesquisas dependem largamente da performance escrita ou oral dos aprendizes e de que o tipo de teste gramatical que iremos aplicar estará diretamente ligado à pergunta inicial e também à linha teórica seguida pelo estudo em questão. Conforme Mackey e Gass (2005), não importa qual seja o principal objetivo ou hipótese da pesquisa, os TM nunca serão fáceis de serem empregados e interpretados e, por esse motivo, precisam ser confeccionados com cautela, devendo o pesquisador estar consciente de todas as suas vantagens e desvantagens. As controvérsias que permeiam o uso dos TM ainda é alvo de discussões e, por isso, quanto mais autores se interessarem por esclarecer essas questões, maiores serão as chances de sabermos como e quando utilizá-los.

Esperamos que outros pesquisadores reconheçam a necessidade de haver mais estudos nessa área e almejamos termos contribuído, de alguma forma, para o entendimento de uma questão tão controversa em pesquisas linguísticas.

REFERÊNCIAS

- BIALYSTOK, E. Explicit and implicit judgments of L2 grammaticality. *Language Learning*, v.29, n.1, p. 81-103, 1979.
- BIALYSTOK, E.; RYAN, E. A metacognitive framework for the development of first and second language skills. In: FORREST-PRESLEY, D. et al. (Ed.). *Metacognition, cognition, and human performance*. Vol. 1: Theoretical Perspectives. Orlando: Academic Press, 1985. p. 207-52.
- BIRDSONG, D. *Metalinguistic performance and interlinguistic competence*. New York: Springer-Verlag, 1989.
- BLEY-VROMAN, R.; YOSHINAGA, N. Broad and narrow constraints on the English dative alternation: some fundamental differences between native speakers and foreign language learners. *University of Hawaii working papers in ESL*, n. 11, p. 157-99, 1992.
- CHAUDRON, C. Research on metalinguistic judgments: a review of theory, methods and results. *Language Learning*, v.33, n.3, p. 343-77, 1983.
- _____. Data collection in SLA research. In: DOUGHTY, C.; LONG, M. *The handbook of second language acquisition*. Oxford: Blackwell, 2003. p. 762-828.
- CHOMSKY, N. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press, 1965.
- COOK, V. *Linguistics and second language acquisition*. New York: Palgrave, 1993.
- CORDER, S. P. Idiosyncratic dialects and error analysis. *International Review of Applied Linguistics*, n.9, p. 147-159, 1979.

- COWART, W. Anchoring and introspective judgments of sentence acceptability. *Perceptual and Motor Skills*, n. 79, p. 1171-82, 1994.
- DOUGHTY, C.; LONG, M. Eliciting second language speech data. In: MENN, L. et al. *Methods for studying language production*. Mahwah, NJ: Erlbaum Associates, 2000. p.149-77.
- ELLIS, R. Grammaticality judgments and second language acquisition. *Studies in second language acquisition*, v.13, n.2, p. 161-86, 1991.
- _____. *The study of second language acquisition*. Oxford: Oxford University Press, 1994.
- ELLIS, R.; BARKHUIZEN, G. *Analysing learner language*. Oxford: Oxford University Press, 2005.
- _____. *Experimental syntax: applying objective methods to sentence judgments*. London: Sage Publications, 1997.
- FLANIGAM, B. Anaphora and relativization in child second language acquisition. *Studies in second language acquisition*, v.17, n.3, p. 331-51, 1995.
- GASS, S. The reliability of second-language grammaticality judgments. In: TARONE, E.; GASS, S.; COHEN, M. (Ed.). *Research methodology in second language acquisition*. Hillsdale, New Jersey: Lawrence Erlbaum, 1994. p. 303-322.
- GREENBAUM, S. (Ed.) *Acceptability in language*. The Hague: Mouton, 1977.
- HAWKINS, R. The theoretical significance of Universal Grammar in second language acquisition. *Second Language Research*, v.17, n.4, p. 345-367, 2001.
- HAWKINS, R.; CHAN, Y. The partial availability of Universal Grammar in second language acquisition: the 'failed functional features hypothesis'. *Second Language Research*, n.13, p. 187-226, 1997.
- INAGAKI, S. Motion verbs with goal PPs in the L2 acquisition of English and Japanese. *Studies in second language acquisition*, n. 23, p. 153-70, 2001.
- JOHNSON, J. NEWPORT, E. Critical period effects in second language learning: the influence of maturational state of acquisition of ESL. *Cognitive Psychology*, n. 21, p. 60-99, 1989.
- JUFFS, A. Semantics-syntax correspondences in second language acquisition. *Second Language Research*, n.12, 177-221, 1996.
- MACKEY, A. Input, interaction and second language development: an empirical study of question formation in ESL. *Studies in second language acquisition*, n.21, p. 557-88, 1999.
- MACKEY, A.; GASS, S. *Second language research: methodology and design*. London: Lawrence Erlbaum, 2005.
- MANDELL, P. On the reliability of grammaticality judgment tests in second language acquisition research. *Second Language Research*, v. 15, n.1, p. 73-99, 1999.
- MYLES, F; MITCHELL R.; HOOPER, J. Interrogative chunks in French L2: a basis for creative construction? *Studies in Second Language Acquisition*, n.21, p. 49-80, 1999.

SCHACHTER, J. Testing a proposed universal. In: GASS, S.; SCHACHTER, J. (Ed.). *Linguistic perspectives on second language acquisition*. Cambridge, UK: Cambridge University Press, 1989. p. 73-88.

SCHACHTER, J. YIP, V. Grammaticality judgments: why does anyone object to subject extraction? *Studies in Second Language Acquisition*, n.12, p. 379-92, 1990.

SCHÜTZE, C. *The empirical base of linguistics: grammaticality judgments and linguistic methodology*. Chicago: The University of Chicago Press, 1996.

SELINKER, L. Interlanguage. *International Review of Applied Linguistics*, n.10, p. 209-231, 1972.

SNOW, C.; MEIJER, G. On the secondary nature of syntactic intuitions. In: GREENBAUM, S. (Ed.). *Acceptability in language*. The Hague: Mouton, 1977. p.163-177.

SORACE, A. The use of acceptability judgments in second language acquisition research. In: BHATIA, T; RITCHIE, W (Ed.). *Handbook of language acquisition*. New York: Academic Press, 1996. p. 375-409.

SPENCER, N. Differences between linguists and nonlinguists in intuitions of grammaticality-acceptability. *Journal of Psycholinguist Research*, v.2, n.2, p. 83-98, 1973.

TOWELL, R.; HAWKINS, R. *Approaches to second language acquisition*. Bristol, UK: Multilingual Matters, 1994.

TREMBLAY, A. Theoretical and methodological perspectives on the use of grammaticality judgment tasks in linguistic theory. *Second languages studies*, v. 24, n.1, p.129-167, 2005.

WHITE, L. Second language acquisition: from initial to final state. In: ARCHIBALD, J. *Second language acquisition and linguistic theory*. Massachusetts: Blackwell, 2000. p. 130-155.

_____. *Second language acquisition and universal grammar*. Cambridge, UK: Cambridge, 2003.

Recebido em 20/06/2016. Aceito em 13/01/2017.