

A VARIAÇÃO DIATÓPICA NO USO DO PRETÉRITO PERFEITO SIMPLES E COMPOSTO NA LÍNGUA ESPANHOLA: ANÁLISE DE CORPUS¹

Leandra Cristina de Oliveira² - UFSC

RESUMO: Esta pesquisa analisa o uso do pretérito perfeito simples e do pretérito perfeito composto (PS e PC, respectivamente) na língua espanhola. O *corpus* elaborado apresenta a frequência do PS e do PC nos sete países selecionados: Argentina, Bolívia, Chile, Cuba, Espanha, México e Peru. A análise quantitativa de uma amostra de 1678 casos conta com o suporte tecnológico da lingüística computacional e com pacotes estatísticos, os quais favorecem um exame mais preciso. Objetivamos apresentar numericamente a variação geográfica no uso dessas formas verbais a qual não deve ser ignorada no ensino de línguas.

ABSTRACT: *This research analyzes the usage of simple and composed forms of the past tense (PS and PC, respectively) in the Spanish language. The corpus that has been elaborated presents the frequency of the PS and of the PC in the seven selected countries: Argentina, Bolivia, Chile, Cuba, Mexico and Peru. The quantitative analysis of 1678 samples has the technological support of Computational Linguistics and statistical packages, which converges to a more precise analysis. The purpose is then to present numerically the geographic variation of the usage of such verbal forms, which can not be ignored in language teaching.*

1 – Introdução

A elaboração de um *corpus* lingüístico do castelhano surgiu com a necessidade de analisar dados recentes da língua escrita, especialmente no que diz respeito ao uso das formas verbais simples e composta do pretérito perfeito. Essa preocupação se dá por alguns motivos: 1) as diferentes manifestações dessas formas em espaços geográficos distintos, Espanha e América Latina, por exemplo; 2) o contraste entre as normas gramaticais estabelecidas e o uso da língua; 3) a dificuldade de um aprendiz do espanhol como língua estrangeira em identificar as tênues diferenças entre o pretérito perfeito simples e pretérito perfeito composto (doravante, PS e PC, respectivamente).

Considerando a contribuição que a estatística traz às ciências humanas e sociais, tomamos este método a fim de discutir o primeiro item acima. Portanto, esta pesquisa está ancorada na análise estatística dos dados presentes no *Corpus de notícias mundiais no panorama hispânico*. O objetivo é verificar a variação diatópica entre o PS e o PC, ou seja, observar a frequência em que ocorre uma forma ou outra em contextos geográficos distintos.

A “população acessível” analisada nesta abordagem é composta por dados da língua escrita, isto é, aquela que, segundo as teorias gramaticais, deve estar mais próxima da norma padrão. A decisão por adotar o termo “população acessível” é uma forma de evidenciar que neste estudo há um pequeno conjunto de elementos passíveis de serem observados, isto é, não há pretensão de generalizar as características da língua escrita com esta simples análise.

“Quando houver diferença razoável entre a população alvo e a população acessível, pode haver grande viés ao generalizar os resultados da análise para toda a população alvo. Nestes casos, é recomendável citar (...) que seus resultados valem especificamente para a população definida como acessível, evitando, assim, que seus resultados sejam usados de maneira inadequada.” (BARBETTA, 1994: 19)

Esta população é representada pelo *corpus* supracitado o qual apresenta notícias de jornais virtuais publicadas na Espanha e em seis países da América Latina (Argentina, Bolívia, Chile, Cuba, México e Peru) entre os meses de fevereiro de 2005 a janeiro de 2006. O *corpus* é constituído por 42.276 palavras e 1.678 dados para análise, neste caso, as formas simples e composta do pretérito perfeito.

Vale destacar o motivo da escolha pelas notícias de jornais virtuais. Primeiramente, porque a rede mundial é um veículo de comunicação bastante acessível, fornecendo de maneira rápida e econômica a publicação das notícias mais recentes. Em segundo lugar, porque ao escolher as notícias, especialmente

¹ Agradeço aos professores Dra. Luizete Guimarães Barros e Dr. Marco Antônio Esteves da Rocha pela leitura atenciosa e pelas sugestões.

² leandraletras@hotmail.com

aquelas publicadas por jornais conceituados dentro e fora do país de origem, espera-se que haja monitoramento na escrita, ou seja, que certas normas gramaticais sejam cuidadosamente respeitadas.

2 – Metodologia

Consciente da necessidade da organização criteriosa e objetiva no desenvolvimento desta pesquisa torna-se fundamental o cumprimento do planejamento que visa, primeiramente, a elaboração de um *corpus* que possa representar a população.

2.1 – Dos critérios para a elaboração do *corpus*

Em estudos de *corpora*, é comum levantar a questão sobre a representatividade do *corpus*, normalmente relacionada à sua extensão (número de palavras). Para Berber Sardinha (2004: 22), quanto maior o *corpus*, maior a sua representatividade. No entanto, ele destaca “o *corpus* é uma amostra de uma população cuja dimensão não se conhece (a linguagem como um todo). Desse modo, não se pode estabelecer o tamanho ideal da amostra” (BERBER SARDINHA, 2004: 22). Segundo o autor, sendo a língua um fenômeno em constante modificação, torna-se impossível analisá-la em sua totalidade.

Destarte, acreditamos que um *corpus* cuja população é a linguagem, não pode ser considerado representativo em termos absolutos. Não temos, portanto, a pretensão de afirmar que nossa amostra representa como a língua espanhola é praticada nos contextos analisados. Elaboramos um *corpus* e pretendemos, a partir dele, investigar a manifestação do PS e do PC em dados reais sem apontarmos generalizações.

Assim, tendo definida a população, foi preciso pensar em alguns critérios que permitissem a elaboração de um *corpus* representativo, porém de curta extensão respeitando o tempo reduzido para esta pesquisa e a complexidade da identificação de cada verbo para o processamento do programa SPSS 10.0 FOR WINDOWS®.

Para representar, parcialmente, a diversidade geográfica da língua espanhola, foi selecionado, a princípio, um jornal em cada um dos sete países supracitados: www.clarin.com (Argentina); www.elmundo.com.bo (Bolívia); www.lun.com.cl (Chile); www.granma.cu (Cuba); www.el-mundo.es (Espanha); www.cronica.com.mx (México); www.larepublica.com.pe (Peru). Considerando a dificuldade em encontrar um evento noticiado nestes contextos, na mesma data, surgiu a necessidade de pesquisar outros jornais. Acrescentamos, portanto, os periódicos www.elpais.es (Espanha); www.diario.com.mx e www.lapinion.com.mx (México); www.lajornada.com.pe (Peru).

Foram coletadas aproximadamente quinze notícias de cada jornal, priorizando as de repercussão mundial, ou melhor, buscou-se respeitar o critério de escolher o mesmo fato apresentado em espaços distintos. O número restrito de notícias em um espaço de tempo, consideravelmente, longo (de fevereiro de 2005 a janeiro de 2006) ocorreu devido à dificuldade em encontrar simultaneidade temporal na apresentação dos acontecimentos. Determinados eventos eram apresentados quase em tempo real³ por alguns periódicos e somente no dia seguinte, por outros. Deste modo, tivemos uma amostra um pouco mais limitada do que esperávamos. Entretanto, acreditamos que o cumprimento dos requisitos supracitados, previamente estabelecidos, era primordialmente importante.

Mesmo reconhecendo a importância da inclusão de diferentes tipos de textos no *corpus*, conforme recomenda Rocha (2000), esta pesquisa apresenta uma outra dimensão, ou seja, analisar textos equivalentes em dois aspectos: o assunto tratado e o tempo em que eles ocorrem.

2.2 – Da escolha das variáveis

Um modelo pode ter variáveis dependentes, que podem ser influenciadas ou previstas por outra independente, e variáveis independentes, que podem influenciar ou prever os valores da dependente.

O modelo elaborado para esta investigação apresenta duas variáveis: (V1) – variável dependente que representa as formas verbais o PS e o PC; (V2) – variável independente representando os países selecionados. Assim, partindo de uma análise quantitativa, é possível perceber a relação entre as variáveis: *Forma Verbal* (V1) e *Contexto Geográfico* (V2).

³ “Tempo real” neste contexto, significa que o evento era noticiado no mesmo dia do seu acontecimento.

2.3 – Da elaboração do modelo

Com o uso do programa computacional SPSS 10.0 FOR WINDOWS®, foi preciso elaborar um modelo para que o *corpus* fosse processado pelo programa. Atendendo a esta exigência, cada forma verbal do pretérito perfeito (somente) foi codificada com suas respectivas variáveis entre colchetes angulares imediatamente após o verbo, sem espaço.

Assim, a (V1) foi codificada com *s* ou *c* – formas verbais simples e composta do pretérito perfeito, respectivamente; a (V2) com *ar, bo, cl, cu, es, mx, pe* – contexto geográfico. Por exemplo:

- (1) “Wilma **castigó**⁴<**s;ar**> a México y avanza hacia Florida.” (www.elclarin.com – Argentina, sábado, 22/10/05)
- (2) “El ojo del huracán 'Wilma' **tocó**<**s;bo**> tierra firme en México en la costa noreste de la península de Yucatán tras pasar por encima de la isla de Cozumel.” (www.elmundo.com.bo – Bolívia, sábado, 22/10/05)
- (3) “Wilma **ha azotado**<**c;cl**> con especial fuerza al idílico balneario de Cancún.” (www.lun.com.cl – Chile, sábado, 22/10/2005)
- (4) “... en el balneario mexicano de Cancún, una estación meteorológica automática **marcó**<**s;cu**> vientos de hasta 215 kilómetros por hora...” (www.granma.cu – Cuba, sábado, 22/10/05)
- (5) “El poderoso huracán 'Wilma' golpea con fuerza desde el viernes las paradisíacas playas del Caribe mexicano, donde **ha provocado**<**c;es**> importantes inundaciones y graves daños.” (www.el-mundo.es – Espanha, sábado, 22/10/05)
- (6) “Seis municipios de Yucatán **fueron**<**s;mx**> severamente afectados por Wilma.” (www.cronica.com.mx – México, sábado, 22/10/05)
- (7) “Potente huracán **tocó**<**s;pe**> tierra firme en la costa noreste de la península de Yucatán, en México.” (www.larepublica.com.pe – Peru, sábado, 22/10/05)

A interpretação da codificação entre os colchetes angulares após os verbos destacados acima deve ser:

- i) À esquerda: variável dependente (V1), codificada por *s* ou *c* identificando as respectivas formas verbais simples e composta do pretérito perfeito.
- ii) À direita: variável independente (V2), codificada por *ar, bo, cl, cu, es, mx, pe*, representando os países: Argentina, Bolívia, Chile, Cuba, Espanha, México e Peru, respectivamente.

Para testar a significância entre duas variáveis, optou-se por utilizar um dos testes mais utilizados em pesquisa social: o teste qui-quadrado (χ^2).

3 – Frequência do PS e do PC nos diferentes contextos geográficos

De acordo com diversas teorias sobre os dois pretéritos, há diferentes possibilidades de uso dessas formas verbais. Muitos fatores contribuem para essa diversidade, mas apenas um deles é de interesse nesta pesquisa: a variação geográfica.

Considerando que a língua espanhola falada nos seis países analisados, entre outros, é subordinada a uma única norma padrão, estabelecida pela RAE (Real Academia Espanhola), partimos de duas hipóteses⁵:

⁴ Grifos nossos.

⁵ Conforme Barbeta (1994), a hipótese nula (H_0) deve ser, basicamente, uma negação daquilo que o pesquisador quer provar. Quando os dados mostram que essa hipótese é falsa, aceita-se a alternativa (H_1), que geralmente reflete o que se quer provar.

- Hipótese nula (H_0): o uso do PC e do PS não apresenta variação diatópica, ou seja, a diferença geográfica não influencia na escolha dessas formas verbais.
- Hipótese alternativa (H_1): o uso do PC e do PS apresenta variação diatópica, ou seja, a diferença geográfica influencia na escolha dessas formas verbais.

A tabela abaixo representa, parcialmente, o processamento dos dados feita pelo programa SPSS.

Tabela 1 – Frequência e percentual do PC e do PS nos países selecionados

V1	V2							Total
	ar	bo	cl	cu	es	mx	pe	
PC	11 4,7%	13 6,6%	14 7,7%	16 8,0%	80 25,9%	27 9,4%	34 12,6%	195 11,6%
PS	224 95,3%	184 93,4%	168 92,3%	183 92,0%	229 74,1%	259 90,6%	236 87,4%	1483 88,4%
Total	235	197	182	199	309	286	270	1678

Fonte: *Corpus* de notícias mundiais no panorama hispânico

Mesmo antes da aplicação do teste de significância é possível observar que existe diferença relevante no que diz respeito à frequência das formas simples e composta nos países analisados. Vemos o predomínio da forma simples em todos os contextos. No entanto, a diferença entre a frequência do PS e do PC não é tão díspar na Espanha quanto nos demais países. Queremos dizer com isso que o índice de 80 aparições para o perfeito composto na Espanha é número bastante maior que 11 casos, da Argentina, ou 13 e 14, de Bolívia e Chile, respectivamente. Contrapor duzentos (média relativa de aparições do perfeito simples) a quinze ou vinte (média numérica do PS nos países americanos) faz o índice da Espanha (oitenta) parecer significativo.

Conforme a tabela 1, na Argentina há uma disparidade entre o uso das duas formas verbais. Neste contexto, o PC corresponde a 4,7% das 235 ocorrências do pretérito perfeito e o PS corresponde a 95,3%.

Bolívia e Peru, mesmo representando a mesma área dialetológica – Região Andina – apresentam diferença curiosa quanto ao uso dos dois pretéritos: dos 197 dados da Bolívia, o PC aparece num percentual de 6,6% e o PS de 93,4%; das 270 manifestações do pretérito perfeito no Peru, 12,6% representa a forma composta e 87,4%, a forma simples.

No Chile, o uso das duas formas verbais não é muito diferente dos demais países que analisamos até aqui. Neste contexto, o perfeito composto corresponde a 7,7% e o perfeito simples, a 92,3% - considerando os 182 dados.

Os resultados obtidos na amostra de Cuba também confirmam o predomínio da forma simples: 8% contra 92% da forma composta, de um número total de 199.

Seguindo a análise do contexto latino-americano, observamos que no México, o PC equivale a uma parcela de 9,4% e o PS a 90,6% das 286 manifestações do pretérito perfeito. Verificamos, portanto, que na América Latina há “*una clara preferència por el pretérito indefinido frente al perfecto compuesto*” (ANDIÓN HERRERO, 2004: 27).

Após a análise da amostra dos seis países latino-americanos, partimos para os dados da Espanha a fim de observar a possível variação diatópica no uso dos dois pretéritos na conjugação castelhana.

Vemos por meio da tabela 1 que, comparado aos países americanos, há uma preferência relativamente menor pelo uso do perfeito simples na Espanha. Embora ainda haja uma predominância do PS (74,1%), a forma composta é bem recorrente neste último contexto, marcando um percentual de 25,9% de um total de 309 dados.

Considerando o fato de que apenas a Espanha apresentou diferença significativa em relação aos demais países, torna-se relevante agrupar os resultados obtidos nestes últimos a fim de confrontar apenas dois contextos geográficos: América e Espanha.⁶

⁶ Não pretendemos com isso, afirmar que há homogeneidade absoluta no espanhol americano.

Tabela 2 – Freqüência e percentual do PC e do PS na América e na Espanha

V1	AMÉRICA	ESPAÑA	TOTAL
PC	115	80	195
	8,4%	25,9%	11,6%
PS	1254	229	1483
	91,6%	74,1%	88,4%
TOTAL	1369	309	1678

Fonte: Tabela 1

Com estes resultados, faz-se necessário verificar a validade das hipóteses lançadas no início deste subtítulo, ou seja, verificar se há influência de uma variável sobre a outra. Para tanto, aplica-se o teste de significância escolhido para esta investigação: teste do qui-quadrado – χ^2 , “método que permite testar a significância da associação entre duas variáveis qualitativas.” (BARBETTA, 1994; p. 222)

Tabela 3 – Contribuições do χ^2 – Análise da V1 e V2

	valor	grau de liberdade (df)	significância (p)
qui-quadrado χ^2	76,05	1	,000

Fonte: Tabela 2

Vemos na tabela 3 o resultado do teste χ^2 pelo qual se obtém a probabilidade de significância (p). Vale lembrar que a significância (p) mede a probabilidade de o efeito ser devido ao acaso. Observamos na tabela acima que a chance de este resultado ser aleatório é menor que uma em mil. Um outro fator a ser destacado é que “quando os dados observados derivam um χ^2 grande (e, em consequência, um p pequeno), o teste rejeita H_0 , em favor da H_1 ...” (BARBETTA, 1994: 228). Nota-se na tabela 3, obtida a partir da tabela 2 (Freqüência e percentual do PC e do PS na América e na Espanha), que o valor do χ^2 é muito alto, pois numa tabela 2x2, cujo grau de liberdade (df) é igual a 1, o resultado deve ser de no mínimo 3.84 para a rejeição da hipótese nula. O valor obtido nesta análise foi de 76,05. Logo, com este valor relativamente alto, rejeitamos a H_0 e aceitamos H_1 – “o uso do PC e do PS apresenta variação diatópica, ou seja, a diferença geográfica influencia na escolha destas formas verbais”.⁷

A citação a seguir evidencia que a variação diatópica no uso dos dois pretéritos tem sido interesse de muitos investigadores da conjugação castelhana.

“... esta distribución entre una forma u otra vacila fuertemente en Hispanoamérica y que en general se emplea mucho más corrientemente el simple popular que el pretérito perfecto compuesto (con menos frecuencia en Perú y Bolivia, y otras áreas limitadas, así como en San Luis, Argentina.” (KANY, 1976: 200)

Considerando a ausência de uma delimitação mais precisa nos países envolvidos nesta análise, fica como sugestão uma pesquisa mais aprofundada de áreas específicas inseridas nos países apresentados no corpus.

⁷ Sendo uma pesquisa fundamentalmente lingüística, pretende-se aplicar o teste χ^2 como recurso para testar a significância entre as variáveis. Não é de interesse de esta investigação detalhar sobre as fórmulas e números utilizados os quais são facilmente encontrados em bibliografias sobre estatística, como Barbetta (1994) e Walsh (1990)

Conclusão

Analisar o uso da forma simples e composta do pretérito perfeito na conjugação castelhana por meio da lingüística de *corpus* torna esta investigação mais abrangente e confiável em termos numéricos.

Nos diversos estudos que atentam para este tema é comum a menção sobre as variações geográficas no uso do PS e do PC sem a apresentação de uma análise quantitativa de dados observados. Por este motivo, surgiu a necessidade da elaboração de um *corpus* que representasse sincronicamente dados recentes da língua escrita, mais especificamente da língua usada nas notícias de jornais virtuais.

Com este recurso, pôde-se verificar que o uso dessas formas verbais se difere nos sete países analisados. O país de menor índice da forma composta é a Argentina, com um percentual de 4,7% contra 95,3% da forma simples. Depois vêm Bolívia e Chile, com um percentual semelhante: 6,6% e 7,7% do PC, e 93,4% e 92,3% do PS, respectivamente. Em Cuba, este resultado não se difere significativamente: 8% do PC e 92% do PS. Nos dois últimos países do contexto latino-americano – México e Peru – os resultados se aproximam: 9,4% e 12,6% da forma composta, e 90,6% e 87,4% da simples, respectivamente. Indubitavelmente, a Espanha apresenta uma grande diferença no uso dessas formas verbais, isto é, o PC é bem mais recorrente neste contexto, 25,9% contra 74,1% do PS.

Considerando a disparidade dos números obtidos na análise dos dados da Espanha, optou-se por agrupar todos os países americanos a fim de confrontar o espanhol peninsular e o americano. A tabela 2 evidencia a variação diatópica no uso das formas do pretérito perfeito em espanhol. Por fim, o teste de significância (χ^2) comprova que a variável *Contexto Geográfico* (V2) tem influência sobre a variável *Forma Verbal* (V1), e que os resultados obtidos não são devido ao acaso.

Acreditamos que a análise da variação diatópica da língua espanhola representa uma ferramenta para professores e alunos verificarem o registro numérico dos diferentes usos da forma simples e composta do pretérito perfeito, pois sendo o espanhol uma língua falada por quase quinhentos milhões de pessoas, distribuídas em vinte e dois países, é inevitável que as normas estabelecidas não sejam igualmente seguidas nos diferentes contextos dos países hispânicos. Não queremos afirmar com isso que os padrões devam ser ignorados. Contudo, queremos tornar visível, a todos aqueles interessados na língua castelhana, a variação no uso das duas formas verbais.

Referências

Livros

ANDIÓN HERRERO, Maria A. *Variiedades del español de América: una lengua y diecinueve países*. Madrid: Espasa, 2004.

BARBETTA, P. A. *Estatística aplicada às ciências sociais*. Florianópolis: Editora da UFSC, 1994.

BERBER SARDINHA, T. *Lingüística de corpus*. Barueri: Manole, 2004.

KANY, Charles E. *Sintaxis Hispanoamericana*. Madrid: Editorial Gredos, 1976.

Artigos

ROCHA, Marco. *Métodos estatísticos em lingüística de corpus*. UFSC – Não publicado. Florianópolis, 2000.

WALSH, A. *Statistics for the social sciences*. New York: Harper and Row, 1990.

Endereços eletrônicos

www.clarin.com

www.cronica.com.mx

www.diario.com.mx

www.elmundo.com.bo

www.el-mundo.es

www.elpais.es

www.granma.cu

www.lajornada.com.pe

www.laopinion.com.mx

www.larepublica.com.pe

www.lun.com.cl

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.