

# O que fazer com grupos de fatores não selecionados? O caso da concordância nominal no Paraná

*Odete Pereira da Silva Menon*

*Edson Domingos Fagundes*

*Loremi Loregian-Penkall*

## **Resumo**

*O objetivo deste trabalho é repensar a questão da seleção de grupos de fatores (GFs) pelo pacote Varbrul, a partir de uma divergência ocorrida quando das rodadas com duas cidades do Paraná, Irati e Pato Branco, na análise da concordância nominal. Um dos GFs estruturais, tipos de formação do plural, foi selecionado quando se rodou Irati, mas não em Pato Branco, apesar de a amostra ter sido feita com uma única metodologia, a do Projeto Varsul, e com exatamente os mesmos GFs, tanto os linguísticos, quanto os sociais. Por isso se questiona até que ponto se deve considerar, na análise, somente a seleção dos GFs pelo Varbrul, quando o objetivo de uma amostra, como é o caso, é perscrutar se e até que ponto a ocupação étnica diferenciada (sobretudo na região Sul) poderia interferir em fenômenos da língua portuguesa no Brasil.*

**Palavras-Chave:** Sociolinguística quantitativa. Seleção de grupos de fatores. Concordância nominal em Irati e Pato Branco. VARSUL-PR.

## O projeto VARSUL: diversidade étnica e sociocultural

Os dados de *concordância nominal* aqui discutidos são provenientes do Banco de Dados VARSUL, cujo objetivo geral é fornecer dados para a descrição do português falado no Sul do Brasil e está sediado em quatro Universidades, dos estados do Paraná, Santa Catarina e Rio Grande do Sul.<sup>1</sup>

O Banco de dados VARSUL é resultante da execução do projeto *Variação Linguística Urbana na Região Sul do Brasil*, cuja concepção foi idealizada em 1984, por Leda Bisol, que reuniu, em outubro daquele ano, alguns pesquisadores em Porto Alegre. O projeto proposto pela pesquisadora deveria espelhar-se no projeto pioneiro de levantamento sociolinguístico no Brasil: Projeto Censo Linguístico do Rio de Janeiro, coordenado por Anthony Julius Naro, e executado no final dos anos 70, com os primeiros resultados publicados no início dos anos 80. O Projeto Censo limitou a coleta de dados à cidade do Rio de Janeiro, realizando as entrevistas em diferentes bairros, representativos das diferentes comunidades cariocas, sobretudo do ponto de vista social.

Embora o modelo de coleta de dados do Varsul fosse o do Censo, para dar conta da diversidade étnica da região, chegou-se a um denominador comum: não bastaria pesquisar as capitais dos três estados (Curitiba, Florianópolis e Porto Alegre); seria necessário incluir algumas das etnias representativas da ocupação étnica diferenciada não só da região Sul, mas de cada estado tomado individualmente. A razão disso residia no fato de que se pretendia pesquisar se o português da região sul diferiria dos demais dialetos do PB (português do Brasil) como consequência do povoamento distinto dos três estados (essa região era praticamente despovoada no tempo em que mais entraram escravos no Brasil).

Em relação à metodologia do levantamento de dados, a equipe VARSUL seguiu a mesma diretriz laboviana do Projeto Censo, a das entrevistas ditas sociolinguísticas, com duração média de cinquenta minutos.

Para dar conta da diversidade étnica, em cada estado foram selecionados, então, três municípios representativos de grupos populacionais comprovadamente relevantes no seu processo de ocupação.<sup>2</sup> No Paraná, optou-se pelas etnias mais antigas: Irati (eslavos); Londrina (colonização mineiro-paulista); Pato Branco (colonização gaúcho-catarinense). Em Santa Catarina, foram selecionadas: Blumenau (alemães); Lages (caminho das tropas); Chapecó (italianos, provenientes do RS). No Rio Grande do Sul, uma escolha evidente: alemães (Panambi); italianos (Flores da Cunha) e região de fronteira (São Borja).

Uma série de estudos já foi realizada pelos pesquisadores pertencentes ao projeto, a partir do banco de dados, seja no campo da fonética/fonologia; da morfossintaxe ou do léxico, desde 1995, quando já estavam disponibilizadas (transcritas e armazenadas

<sup>1</sup> O Projeto *Variação Linguística Urbana na Região Sul (VARSUL)*, iniciado em 1990, é composto inicialmente pelas três universidades federais dos três estados do Sul do Brasil: Universidade Federal de Santa Catarina (UFSC), Universidade Federal do Paraná (UFPR) e Universidade Federal do Rio Grande do Sul (UFRGS). Em 1993, passa a contar com a Universidade Católica do Rio Grande do Sul (PUC-RS). O Projeto VARSUL tem como horizonte armazenar amostras de realizações de fala de habitantes de áreas urbanas sócio-representativas de cada um dos três estados da região Sul do Brasil, estratificados por localidade, faixa etária, sexo e escolaridade. O Banco VARSUL vem sendo constantemente ampliado com o acréscimo de novas amostras em todas as sedes. A amostra básica, constituída de informantes sem curso superior, distribuídos por grau de escolaridade, sexo e faixa etária (acima de 25 anos), outras vêm sendo acrescentadas, contemplando novas regiões, diferentes faixas etárias, bem como níveis de escolaridade.

<sup>2</sup> Como havia a limitação de três cidades por estado, ficaram de fora algumas etnias presentes nos estados: árabes, franceses, japoneses e holandeses, por exemplo.

em microcomputadores) as entrevistas das capitais. Para detalhamento, consulte-se o site [www.cce.ufsc.br/~varsul](http://www.cce.ufsc.br/~varsul) e [www.pucrs.br/fale/pos/varsul/index.php](http://www.pucrs.br/fale/pos/varsul/index.php).

### O fenômeno em estudo: a concordância nominal no Paraná

O grupo de pesquisa do Varsul-Paraná vem se dedicando ao estudo da *concordância nominal* nas cidades do banco, a fim de verificar se a regra geral do fenômeno, classificada por Scherre & Naro (1997, p. 93) como um “caso típico de variação inerente” no PB, se aplicaria no estado, apesar de ela apresentar uma realização bastante característica, quando um dos determinantes é pronome possessivo (*o meus filho*). Para fins de comparabilidade com outras regiões do Brasil, adotou-se para os GFs linguísticos o critério de tentar utilizar aqueles já empregados em outras pesquisas do mesmo teor. A igualdade absoluta não foi possível, visto que, em alguns trabalhos, apareceram casos que não foram registrados na nossa amostra, como aquele de Dias (1993, p. 124): “Aí tinha três *filhotin*, den da barriga dela (Paul, 10, r, m)”, classificado como “redução do item lexical”. No campo das variáveis extralinguísticas, há a limitação natural do banco: sexo, idade (25-45 e mais de 50 anos), escolaridade (até cinco anos, oito e onze anos) e etnia (conforme acima).

Fizemos uma restrição na análise dos dados da alternância na concordância nominal (CN): consideramos somente a concordância **intra-sintagmática**, deixando de lado a concordância com o predicativo. Como em outros trabalhos, a variável linguística dependente foi a presença/ausência de marca morfológica de plural nos elementos do sintagma nominal (SN) e as variáveis independentes linguísticas foram: posição linear dos elementos no SN; classe gramatical das palavras; posição em relação ao núcleo do SN; tipos de formação de plural, marcas precedentes; contexto fonológico seguinte (para eliminar os casos problemáticos, como “nas porta\_? sabe?”); grau dos substantivos<sup>3</sup> e adjetivos; tonicidade das palavras; animacidade dos substantivos.

Como pretendemos fazer o levantamento das quatro cidades do Paraná, começamos por Irati, visto que, no estudo de outros fenômenos (apagamento do *se* reflexivo em Bandeira, 2007 e na alternância indicativo/subjuntivo em Fagundes, 2007), essa localidade já havia mostrado apresentar características mais conservadoras quando comparada com as outras três do Banco.

Os resultados de Irati foram apresentados no Gel de 2008 e os de Pato Branco no CELLIP de 2009. Em Irati, foram nove os GFs selecionados: relação com o núcleo; posição linear no SN; escolaridade; grau do substantivo; tipos de formação de plural; sexo do informante; traço semântico; idade e classe gramatical. Quando rodamos Pato Branco, apareceu um problema: o GF *tipos de formação de plural*, que havia sido selecionado em quinto lugar

<sup>3</sup> É interessante observar que, no caso dos aumentativos que passaram a constituir entrada lexical independente, como *giletão*, *machão*, *bailão*, *bodegão*, a não-concordância foi categórica.

em Irati, não foi selecionado porque apresentou nocaute no item VEZ<sup>4</sup>. Os resultados de Irati podem ser conferidos na Tabela 1 (*Input* .76).

TABELA 1 – Tipos de formação de plural – Irati

Fatores	Apl./Total	%	PRel.
Item lexical VEZ	24/28	86	.86
Itens terminados em -ão/-ões	59/72	82	.81
Itens terminados em -r	84/106	79	.75
Plural duplo	15/20	75	.73
Itens terminados em -l	33/43	77	.65
Itens terminados em -s/-z	15/22	68	.59
Itens com plural regular	3.316/4.358	76	.48
Itens terminados em -ao/-s	19/42	45	.33
Total	3.565/4.691	76	

Para Pato Branco, como não tinha sido selecionado o GF, foi feita uma rodada só com a variável dependente (com ou sem concordância) e o GF tipos de formação de plural, para se obter pesos relativos, a fim de comparar com os resultados de Irati (ainda que esses tivessem sido objeto de iterações com outros GFs). O resultado da rodada está na Tabela 2, com *Input* de .81 (onde se incluiu o nocaute com o item VEZ só para constar o número de ocorrências visto que, evidentemente, foi retirado do GF para se poder rodar o Varb2000):

TABELA 2 – Tipos de formação de plural – Pato Branco

Fatores	Apl./Total	%	P.Rel.
Item lexical VEZ	02	100	
Itens com plural regular	2.187/2.695	81	.51
Plural duplo	11/14	79	.47
Itens terminados em -ão/-s	24/31	77	.45
Itens terminados em -ão/-ões	32/43	74	.41
Itens terminados em -r	52/75	69	.35
Itens terminados em -l	19/28	68	.34
Itens terminados em -s/-z	2/3	67	.32
Total	2.329/2.891	81	

<sup>4</sup> Pela alta incidência da locução nominal *às vezes*, ela não foi considerada, como já aconteceu em outros trabalhos. O item lexical VEZ foi analisado em separado, conforme critério adotado por Dias (1993).

Nessa rodada, foram selecionados os GFs: posição linear no SN; relação com o núcleo; traço semântico; grau do substantivo; sexo do informante; idade e escolaridade.

As diferenças (e/ou aparentes incongruências) que se pode apontar entre os resultados contidos nas Tabelas 1 e 2 compreendem: para os *itens terminados em -ão/-ões* temos grande favorecimento da aplicação da regra de concordância em Irati (.81) e desfavorecimento da concordância para Pato Branco (.41). O mesmo ocorre em relação ao *plural duplo*, com Irati favorecendo a aplicação da regra com .73 e Pato Branco desfavorecendo com .47. Seguem na mesma direção os *itens terminados em -r* (Irati com .75 e Pato Branco com .35); os *itens terminados em -l* (Irati com .65 e Pato Branco com .34) e os *itens terminados em -s/-z* (Irati com .59 e Pato Branco com .32).

Como ambas as cidades tiveram idêntico tratamento, tanto na gravação das entrevistas como no tratamento dos dados quantificados, não havia explicação lógica para tal situação. Mais: não seria possível a comparação linear dos resultados, pois havia as divergências acima.

Tentamos, então, reparar a situação, fazendo outra rodada para Irati, sem o item lexical VEZ (que ali contava com 28 ocorrências, das quais 24 com concordância). A nova rodada consistiu em aplicar a regra também rodando somente a variável dependente (com ou sem concordância) com o GF *tipos de formação de plural*. Os resultados são os que constam nas duas tabelas abaixo (*Input .76*):

TABELA 3 – Tipos de formação de plural – Irati (sem vez)

Fatores	Apl./Total	%	PRel.
Itens terminados em -r	84/106	79	.55
Itens terminados em -ão/-ões	59/75	79	.54
Itens terminados em -l	33/43	77	.51
Itens com plural regular	3.320/4.365	76	.50
Plural duplo	15/20	75	.49
Itens terminados em -s/-z	15/22	68	.40
Itens terminados em -ão/-s	19/42	45	.21
Total	3.547/4.676	76	

Como ainda se apresentassem divergências em algumas variantes – o que, em princípio, poderia ser atribuído à questão *etnia* ou povoamento diferenciado –, resolvemos fazer uma série de testes, com ambas as cidades em separado, rodando a variável dependente com o GF *tipos de formação de plural*, eliminando ora umas, ora outras variantes do grupo, ou amalgamando variantes com resultados semelhantes.

Iniciamos com a rodada sem o item *vez* e sem itens terminados em *-r*, cujos resultados podem ser comparados nas tabelas 4 e 5 (*input* .76 e .81, respectivamente):

TABELA 4 – Tipos de formação de plural: Irati (sem *vez* e *-r*)

Fatores	Apl./Total	%	PRel.
Itens terminados em <i>-ão/-ões</i>	59/75	79	.54
Itens terminados em <i>-l</i>	33/43	77	.51
Itens com plural regular	3.320/4.365	76	<b>.50</b>
Plural duplo	15/20	75	.49
Itens terminados em <i>-s/-z</i>	15/22	68	.41
Itens terminados em <i>-ão/-s</i>	19/42	45	.21
Total	3.463/4.570	76	

Como se pode observar nessas tabelas, há resultados semelhantes em alguns itens testados mas, em relação aos itens terminados em *-ão/-ões* (Irati .54; Pato Branco .41), aos itens terminados em *-l* (Irati .51; Pato Branco .33) e aos itens terminados em *-ão/-ãos* (Irati .21; Pato Branco .45) continua havendo certas discrepâncias entre os pesos atribuídos às duas cidades.

TABELA 5 – Tipos de formação de plural: Pato Branco (sem *vez* e *-r*)

Fatores	Apl./Total	%	PRel.
Itens com plural regular	2.187/2.695	81	<b>.50</b>
Plural duplo	11/14	79	.46
Itens terminados em <i>-ão/-s</i>	24/31	77	.45
Itens terminados em <i>-ão/-ões</i>	32/43	74	.41
Itens terminados em <i>-l</i>	19/28	68	.33
Itens terminados em <i>-s/-z</i>	2/3	67	.32
Total	2.329/2.891	81	

Face a esses números, e porque o plural regular do ditongo nasal *-ão* (em *-ãos*) apresentou divergência de comportamento, resolvemos rodar juntas as duas cidades (*Input* .78), em relação ao GF *tipos de formação de plural*, sem o item *vez* e sem os itens terminados em *-ão/-s*:

TABELA 6 – Tipos de formação de plural: IRT/PBR  
(sem *vez* e -ão/-s)

Fatores	Apl./Total	%	PRel.
Itens com plural regular	5.510/7.063	78	.50
Itens terminados em -ão/-ões	91/118	77	.49
Plural duplo	26/34	76	.48
Itens terminados em -r	136/181	75	.46
Itens terminados em -l	52/71	72	.44
Itens terminados em -s/-z	17/25	68	.38
Total	5.832/7.492	78	

Observa-se que o item *tipo de formação de plural* não foi selecionado e houve favorecimento da marca de concordância para a cidade de Pato Branco (.54) em relação a Irati (.48).

TABELA 7 – distribuição por cidade (sem *vez* e -ão/-s)

Fatores	Apl./Total	%	PRel.
Pato Branco	2.303/2.858	81	.54
Irati	3.531/4.637	76	.48
Total	5.834/7.495	78	

Na nova rodada foram consideradas as duas cidades em relação ao GF tipos de formação de plural (*Input* .78), sem o item *vez* e amalgamando os itens terminados em -s/-z (*mês/meses*) e -r (*plural regular: cama/camas*); *ão/-ões* e -ão/-s (*pão/pães* e *mão/mãos*) e o plural das palavras terminadas em -r e -l (*colher/colheres* e *animal/animais*).

TABELA 8 – Tipos de formação de plural com dados amalgamados e sem *vez*

Fatores	Apl./Total	%	P.Rel.
<i>plural regular</i> e terminados em -s/-z	5.527/7.088	78	.50
Itens terminados em -ão/-ões e -ão/-s	134/191	70	.49
<i>Plural duplo</i>	26/34	76	.48
Itens terminados em -r e -l	188/252	75	.46
Total	5.875/7.4	78	

Observamos que, em todos os casos, a CN fica no limite do ponto neutro, com nítida tendência à queda. Essa talvez tenha sido a razão de o GF não se ter mostrado estatisticamente relevante. Assim, somente *cidade* é que foi selecionado, com mais possibilidade de aplicação da regra no PB.

TABELA 9 – distribuição por cidade dados amalgamados e sem *vez*

Fatores	Apl./Total	%	P.Rel.
Pato Branco	2.327/2.889	81	.54
Irati	3.550/4.679	76	.47
Total	5.877/7.568	78	

A seguir, outra rodada considerando as duas cidades em relação ao GF *tipos de formação de plural*, sem o item *vez* e sem *plural regular* (Input .73).

TABELA 10 – Tipos de formação de plural (sem *vez* e *plural regular*)

Fatores	Apl./Total	%	PRel.
Itens terminados em -ão/-ões	91/118	77	.55
Plural duplo	26/34	76	.55
Itens terminados em -r	136/181	75	.53
Itens terminados em -l	52/71	73	.50
Itens terminados em -s/-z	17/25	68	.44
Itens terminados em -ão/-s	43/73	59	.35
Total	365/502	73	

Nessa rodada, embora haja leve favorecimento para algumas variantes, a significância foi de (.848) e nenhum dos fatores foi selecionado (o *stepdown* confirmou a eliminação).

TABELA 11 – distribuição por cidade (sem *vez* e *plural regular*)

Fatores	Apl./Total	%	PRel.
Irati	227/311	73	.50
Pato Branco	140/194	72	.49
Total	367/505	73	

Estaríamos, então, em face de um problema de falta de ortogonalidade? Ao se retirar *plural regular*, sobraram poucos dados. No entanto, é quando se retira *plural regular* que a expectativa de que Irati favoreceria a CN se concretiza, mesmo se o GF não se tenha mostrado relevante do ponto de vista estatístico.

Para refinar a análise foram feitas duas outras rodadas, via *Crosstab* ou tabulação cruzada, considerando todos os GFs e selecionando alguns para testar o cruzamento. Um dos resultados que fica evidenciado é que **quanto mais regular o item, maior o percentual de concordância**. No cruzamento efetuado entre *escolaridade* e *processos morfológicos de formação do plural*, por

exemplo, nos itens terminado em *-ão/-ões* temos para falantes do primário 62%; para os do ginásio 74% e para os do colegial 90% de concordância (confirmando os *pesos relativos* das rodadas que consideraram a variável escolaridade).

Já para os itens terminados em *-ão/-ões*, não se verifica tal acréscimo de percentual de concordância, pois temos o primário com 50%; o ginásio com 78% e o colegial com 54% de frequência da concordância.

Nesse tipo de análise podem surgir algumas questões que se originam principalmente da “necessidade analítica de comparação controlada” dos dados (cf. Guy e Zilles, 2007, p. 51), sendo as principais: (i) a ortogonalidade dos grupos de fatores e (ii) a assimetria da distribuição dos dados.

### Ortogonalidade de grupos de fatores e assimetria na distribuição dos dados

Para que se possa efetuar análises e comparações precisas e confiáveis de dados, faz-se necessário que os grupos de fatores sejam ortogonais, ou quase ortogonais. De acordo com Guy e Zilles (2007, p. 52), para que tenhamos grupos de fatores ortogonais é preciso que eles coocorram livremente e que não sejam sub- ou supercategoriais uns dos outros. Em outras palavras, quando não temos dados bem distribuídos por grupos de fatores temos a ausência de ortogonalidade: todas as ocorrências de “y” são também as ocorrências de “x”, e todas as ocorrências de “x” são também as ocorrências de “y”.

Guy e Zilles ressaltam que para o programa VARBRUL não importa se as não-ortogonalidades ocorrem **por acaso** ou **por impossibilidade estrutural**. Isto porque “quaisquer que sejam as origens do problema, nenhum procedimento analítico pode parcializar efeitos separados para dois grupos de fatores que apresentem esse tipo de distribuição dos dados” (GUY; ZILLES, 2007, p.53).

Nesse estudo, há dois casos de figura:

- a) aparentemente, o *plural regular* (muito mais numeroso e, portanto, mais suscetível de apresentar ortogonalidade) está puxando para baixo a concordância em Irati, quando, pelos dados passados no pente fino, os falantes apresentam maior frequência de concordância (inclusive no caso do plural em *-ões*);
- b) há restrições de ordem estrutural, uma vez que a formação de plural irregular é numericamente inferior à do plural regular (veja-se o número de ocorrências de cada um).

Essas divergências entre as duas cidades parecem apontar para algum fator de ordem estrutural, que seria de aplicação dife-

renciada da regra, por conta da ocupação diferenciada das duas cidades. No entanto, não foi possível localizar exatamente onde residiria essa diferença pois, além de tudo, no caso dos plurais irregulares, caso específico de *-ãos*, o número de itens lexicais é muito pequeno nas duas cidades (19 ocorrências – 9 de *irmão/-s* – em Irati e 24 em Pato Branco – 17 de *irmão/-s*). Isso nos leva a um questionamento: será que o número de palavras com plural em *-ãos* é menor do que o das palavras cujo plural é *-ões*? Não se dispõe – que seja do nosso conhecimento – de levantamento estatístico dessa frequência que pudesse nos auxiliar na análise. Além disso, vamos ver que, nas duas últimas grandes rodadas, tanto o plural regular quanto o em *-ões* têm comportamento similar (desfavorecendo levemente a concordância, com .49 e .43, respectivamente). É no plural *-ãos* que se localiza a divergência, com favorecimento da CN, em .76, talvez resultado daquele pequeno número de itens e, portanto, de não-ortogonalidade dos dados. E aí é que reside o impasse: Irati apresentaria maior concordância nesse plural irregular (talvez marca de posição/avaliação social?) e menor nos plurais regulares?

Guy e Zilles (2007) afirmam que nem sempre a ortogonalidade se dá em termos categóricos, isto porque ela também pode ocorrer como uma propriedade escalar ou relativa. Nesse sentido, a **assimetria** tem relação direta com a questão de como os dados se distribuem entre as categorias nos diferentes grupos de fatores. “O caso analítico ideal é o de uma distribuição equilibrada com todas as células contendo números iguais (...). Com uma distribuição equilibrada, obtemos resultados analíticos robustos e significativos (...), mas quando começamos a ter distribuições altamente sobrepostas, todos esses resultados começam a cair” (GUY; ZILLES, 2007, p. 58-59).

Passamos, então, a considerar como caso de sobreposição (acavalamento) de grupos de fatores dois dos GFs também utilizados em outros trabalhos sobre CN: *ordem linear dos elementos dentro do sintagma nominal* e *posição em relação ao núcleo*: isso estaria representando uma espécie de burla para o programa, pois ambas as variáveis estariam testando coisas semelhantes, contrariando o princípio da independência dos fatores, proposta pelo pacote Varbrul.

Além disso, parece que a ordem linear não testa a regra, visto que o núcleo pode estar em todas as posições consideradas no GF e, assim, não se pode localizar o que está à esquerda do núcleo. Se esse GF é selecionado em algumas rodadas, é porque ele tem ortogonalidade (tem uma boa distribuição do número de ocorrências em cada variante) e, portanto, é **lido** pelo programa como estatisticamente relevante. Qualitativamente, ele não acrescenta nada à análise.

Realizou-se, então, uma dobradinha de rodadas (*Input .78* em ambas): em cada uma delas eliminou-se um dos GFs acima

mencionados, para se verificar se, quando um não está presente, o outro tem comportamento diferenciado. Dito e feito: quando se alterna o padrão dos GFs, alguns dos outros GFs passam a ter comportamento aleatório: é o caso de *classes de palavras*, que é selecionado em primeiro lugar quando se mantém o GF *ordem linear* dos constituintes do SN. Quando se exclui *ordem linear* e se inclui *posição em relação ao núcleo*, *classes de palavras* cai para a penúltima posição, antes de animacidade:

Tabela 12 – Rodadas com *ordem linear* (sem posição) e sem *ordem linear* (com posição) – GF *classes de palavras*

Com ordem linear – 1º. selecionada				Com posição em relação ao núcleo 7º			
Fatores	Apl./Total	%	P.R.	Fatores	Apl./Total	%	P.R.
Indefinidos	518/523	99%	.83	Indefinidos	518/523	99%	.87
Determinantes	2707/2753	98%	.73	Determinantes	2707/2753	98%	.79
Possessivos	234/246	95%	.73	Possessivos	234/246	95%	.57
Adjetivos	306/422	73%	.35	Adjetivos	306/422	73%	.35
Quantificadores	101/111	91%	.34	Quantificadores	101/111	91%	.44
Substantivos	2003/3497	57%	.27	Substantivos	2003/3497	57%	.22
Total	5869/7552	78%		Total	5869/7552	78%	

Isso demonstra que também *classes de palavras* está interagindo com *ordem linear*, uma vez que *determinantes*, *possessivos* e *indefinidos* **sempre** ocorrem nas primeiras posições – favorecedoras da regra; *adjetivos*, *quantificadores* e *substantivos* podem aparecer nas últimas posições, o que desfavorece a CN.

No entanto, os outros GFs (*escolaridade*; *grau dos substantivos* e *adjetivos*; *idade*; *cidade*; *morfologia* e *animacidade dos substantivos*) permanecem nas mesmas posições, o que garante a confiabilidade dos resultados e a certeza de que a sua distribuição e seleção não ficam alteradas por conta do tropeço do acasalamento daqueles dois GFs. Esse equívoco se reflete somente na seleção dos dois GFs, evidenciando que *ordem linear* não é relevante para a concordância nominal, uma vez que, quando está presente nas variáveis, só é selecionado em terceiro lugar. Quando *posição em relação ao núcleo* está entre as variáveis, é selecionado em primeiríssimo lugar, demonstrando, como em outras pesquisas sobre o assunto, ser este o mais proeminente no grupo de fatores estruturais, na variação da concordância nominal do PB.

Quanto aos GFs sociais, vemos que *idade* é selecionado depois de *escolaridade*, apresentando resultado semelhante ao de outras regiões, ou seja, os mais jovens estão realizando menos concordância (.43) que os mais velhos (.55), podendo se caracterizar o fenômeno como de mudança em curso.

*Escolaridade* é selecionado em segundo lugar, nos dois casos. Podemos inferir, então, que a passagem pela escola, no tocante à concordância nominal, surte algum efeito. Não que acreditemos que a escola seja altamente competente: trata-se da ação, mais intensificada pela escola, sobre o efeito social de que se reveste o uso da não-concordância. De fato, diferentemente de outros fenômenos de variação do PB (como o de *nós/a gente*, p. ex.), há um **estigma** linguístico-social gritante que marca negativamente as pessoas que não produzem concordância nominal (e também a verbal, cf. Naro & Scherre, 1999) no Brasil. Esta sempre é atribuída a pessoas de baixa renda e de pouca ou nenhuma escolaridade. Nos nossos dados, os informantes de colegial não vão além de 64 na aplicação da regra.

### À guisa de conclusão...

Dessa forma, se não se conseguiu localizar exatamente o problema da inversão de resultados da CN entre Irati e Pato Branco (apesar de vislumbrarmos, pelas múltiplas rodadas do Varb2000 e dos *cross-tabs* realizadas, algum resultado estranho em Irati – mais conservador?) para caracterizá-lo como resultante de ocupação étnica diferenciada, pudemos transitar pelos efeitos causados por alguns GFs, quando retirados das rodadas gerais. Esses efeitos seriam o resultado de uma sobreposição (involuntária?) de GFs e que, até agora, não havia sido questionada em outros trabalhos sobre a CN no Brasil.

### Abstract

*This paper aimed at thinking over the issue of selecting groups of factors (GFs) for the Varbrul package after a discrepancy occurred during the rounds with two cities of the state of Paraná – Irati and Pato Branco – regarding nominal agreement analysis. One of the structural GFs – types of plural formation – was selected during the Irati round but not in Pato Branco, although the sample was prepared according to a single methodology (that of the Varsul Project) and with exactly the same GFs, both linguistic and social. That is why the question was raised on whether the analysis should consider exclusively the GF selection by Varbrul when the sample objective – as in this case – was to investigate if and to what extent the distinct ethnic occupation (especially in southern Brazil) could interfere with phenomena of the Portuguese language in Brazil.*

**Keywords:** *Quantitative sociolinguistics, factors group selection, nominal agreement in Irati and Pato Branco, VARSUL-PR.*

## REFERÊNCIAS

- BANDEIRA, G. A. F. *O apagamento de SE nas funções sujeito e objeto: um estudo variacionista com dados do VARSUL do Paraná*. Tese (Doutorado em Letras – Estudos Linguísticos). Curitiba: UFPR. 2007. 262 p.
- DIAS, M. C. A. C. *A variação na concordância nominal: um contraste entre o urbano e o rural na fala brasiliense*. Dissertação (Mestrado). Brasília: UnB. 1993. 177 p.
- FAGUNDES, E. D. *As ocorrências do modo subjuntivo nas entrevistas do VARSUL no estado do Paraná e as possibilidades de variação com o modo indicativo*. Tese (Doutorado em Letras-Estudos Linguísticos). Curitiba: UFPR. 2007. 220 p.
- FAGUNDES, E.D.; LOREGIAN-PENKAL, L.; MENON, O.P.S. Concordância nominal em Pato Branco. Comunicação apresentada no XIX Seminário do CELLIP. Cascavel, Unioeste, 2009.
- GUY, G.; ZILLES, A. *Sociolinguística quantitativa*. São Paulo: Parábola, 2007.
- MENON, O. P. S.; FAGUNDES, E.D.; LOREGIAN-PENKAL, L. Concordância nominal em Irati-PR. Comunicação apresentada no 56º Seminário do GEL. São José do Rio Preto, SP, 2008.
- NARO, A.J.; SCHERRE, M.M.P. A influência de variáveis escalares na concordância verbal. *In: A Cor das Letras*. Revista do Departamento de Letras e Artes da Universidade Estadual de Feira de Santana. Nº 3, dezembro. Bahia, Feira de Santana, 1999. p.17-34.
- SCHERRE, M. M. P.; NARO, A. J. A concordância de número no português do Brasil: um caso típico de variação inerente. *In: HORA, Dermeval da. (org.). Diversidade linguística no Brasil*. João Pessoa: Ideia. 1997. p. 93-114.
- WEINREICH, U. W.; LABOV, W.; HERZOG, M. Empirical foundation for a theory of language change. *In: LEHMANN, W.; MALKIEL, Y. (eds.). Directions for historical linguistics*. Austin: University of Texas Press, 1968. p. 97-195.