



A influência de fatores semânticos no uso de nomes gerais para a indeterminação do sujeito

Eduardo Tadeu Roque Amaral^{1*} e Fernanda Carla de Oliveira²

¹Faculdade de Letras, Universidade Federal de Minas Gerais, Avenida Antônio Carlos, 6627, 31270-901, Belo Horizonte, Minas Gerais, Brasil. ²Instituto Federal de Minas Gerais, Congonhas, Minas Gerais, Brasil. *Autor para correspondência. E-mail. eduamaralbh@ufmg.br

RESUMO. As formas de indeterminação do sujeito do português brasileiro vêm sendo pesquisadas por diversos autores, tais como Milanez (1982), Menon (1994), Carvalho (2010) e outros. Na maioria desses trabalhos, busca-se verificar quais fatores podem condicionar a realização de cada uma das estratégias de indeterminação. Nessa mesma linha, este artigo busca, em uma perspectiva variacionista, discutir a relevância de fatores semânticos no uso de nomes gerais para a indeterminação do sujeito. Por nomes gerais, compreendem-se certas unidades com traços semânticos mínimos, que estão entre o léxico e a gramática (Halliday & Hasan, 1976). O *corpus* da pesquisa está constituído por gravações sociolinguísticas coletadas na cidade de Bambuí-MG. A análise demonstra que, em termos de frequência, os nomes gerais identificados, ‘pessoa’, ‘pessoal’, ‘pessoas’, ‘cara’, ‘homem’, ‘povo’, ‘fulano’ e ‘ser humano’, ocupam a terceira posição no conjunto das variantes. As demais formas estudadas referem-se àquelas que já foram discutidas em trabalhos anteriores, isto é, ‘a gente’, ‘você/ocê/cê’, ‘nós/nóis’, ‘eles/ês’, entre outras. Considerando os resultados estatísticos da correlação entre as variáveis, observa-se que a indeterminação do sujeito por nomes gerais é condicionada pela ausência de anáfora e pela indeterminação total do referente ou indeterminação parcial quando este está implícito.

Palavras-chave: variação linguística; anáfora; grau de indeterminação; português brasileiro.

The influence of semantic factors in the use of general nouns for the subject indeterminacy

ABSTRACT. Subject indeterminacy in Brazilian Portuguese has been researched by several authors, such as Milanez (1982), Menon (1994), Carvalho (2010) and others. Most of the studies aim to verify which factors can condition the use of each strategy of indeterminacy. From a variationist perspective, this article aims to discuss the relevance of semantic factors in the use of general nouns to achieve the subject indeterminacy. As general nouns, we consider certain units, between the lexicon and the grammar, with minimal semantic features (Halliday & Hasan, 1976). The *corpus* of the research is composed of sociolinguistic recordings collected in Bambuí, a city of the state of Minas Gerais. The analysis shows that, in terms of frequency, the general nouns (*pessoa* ‘person’, *pessoal* ‘people’, *pessoas* ‘people’, *cara* ‘guy’, *homem* ‘man’, *povo* ‘people’, *fulano* ‘so-and-so’ and *ser humano* ‘human being’), occupy the third position in the set of the variants. The other variants refer to those that have already been discussed in previous research such as *a gente* ‘we’, *você/ocê/cê* ‘you’, *nós/nóis* ‘we’, *eles/ês* ‘they’ and others. Regarding the statistical results of the correlation between the variables, it is observed that the subject indeterminacy with general nouns is conditioned by the absence of anaphora and by the total indeterminacy of the referent or partial indeterminacy when the referent is implicit.

Keywords: linguistic variation; anaphora; degree of indeterminacy; Brazilian Portuguese.

Received on October 22, 2018.

Accepted on April 23, 2019.

Introdução¹

A indeterminação do sujeito vem sendo estudada por muitos autores e em diversas regiões do país, tais como Milanez (1982), Menon (1994), Setti (1997), Godoy (1999), Oliveira (2006), Carvalho (2010), Souza (2014) e Barbosa (2016). São pesquisas com dados de diferentes localidades que mostram um amplo conjunto de formas que são utilizadas pelos falantes do português para indeterminar o sujeito. São

¹ Este trabalho constitui um desdobramento da dissertação intitulada *Nomes gerais e outras formas de indeterminação do sujeito no português oral de Bambuí-MG*, defendida na Faculdade de Letras da UFMG em 2018 (Oliveira, 2018).

exemplos: ‘a gente’, ‘você’ (e formas reduzidas ‘ocê’ e ‘cê’) e sintagma nominal que tem como núcleo um nome geral, conforme os exemplos (1), (2) e (3) abaixo:

(1) pessoalmente assim pro/que eu vi não... mais... acho que talvez **a gente** não perceba os milagres que aconteçam na vida da gente né (BAM 2)²

(2) assim tirano o da felicidade saúde né lógico mais... com certeza noventa por cento vai se[r] o dinheiro mesmo sempre vai se[r] o dinheiro porque com dinheiro **cê** faz tudo que **você** precisa[r] intendeu? (BAM 19)

(3) se a **pessoa** tem uma fé convicta intendeu? se a **pessoa** tem uma fé convicta o milagre ele acontece... (BAM 12)

Nas pesquisas sobre o tema, têm-se discutido as diferenças existentes entre indeterminação, indefinição e impessoalização. Milanez (1982), ao tentar diferenciar tais termos, demonstra que o conceito de indeterminação foi confundido com o de impessoalização durante muito tempo, e que só a partir de 1957, quando o Ministério da Educação e Cultura sugeriu a Nova Nomenclatura Gramatical Brasileira, é que os autores parecem ter passado a fazer a distinção entre indeterminação e impessoalização. Assim, o que se passou a ver nas gramáticas foram explicações como as de Cunha e Cintra (2008):

Não deve ser confundido o ‘sujeito indeterminado’, que existe, mas não se pode ou não se deseja identificar, com a inexistência de sujeito. Em orações como as seguintes ‘Chove, anoitece, faz frio’, interessa-nos o processo verbal em si, pois não o atribuímos a nenhum ser. Dize-se então que o verbo é ‘impessoal’; e o sujeito, ‘inexistente’ (Cunha & Cintra, 2008, p. 143, grifos do autor).

Nota-se, portanto, que Cunha e Cintra (2008) tentam fazer uma distinção entre indeterminado e impessoal, mostrando que, no primeiro caso, o sujeito existe, mas não pode ou deseja ser identificado e, no segundo, que ele de fato não existe. Milanez (1982) acrescenta que o sujeito indeterminado existe, ou seja, refere-se a um elemento semântico, contudo este elemento não é expresso sintaticamente nas duas ocorrências de indeterminação que a gramática prevê: verbo na terceira pessoa do plural sem sujeito explícito e verbo na terceira pessoa do singular mais o pronome ‘se’.

Outra questão observada pela pesquisadora é que a indeterminação é também, muitas vezes, confundida com a indefinição, ainda que sejam consideradas distintas pela análise sintática. Milanez (1982) mostra que a diferença está no fato de que quando dizemos ‘Alguém bateu à porta’, o sujeito é determinado, ainda que indefinido. A indefinição pressupõe um conjunto fechado de elementos e remete à noção de quantificação. Dessa forma, Milanez (1982, p. 39) aponta que, em (4), tem-se um exemplo de indeterminação e em (4a) de indefinição, tal como ilustrado abaixo:

(4) ‘Lá no Norte come-se fruta no café da manhã’.

(4a) ‘Lá no Norte, **todos** comem frutas no café da manhã’.

Considerando o exposto pela autora, assim como por Menon (1994) e outros autores, o que consideramos neste trabalho como indeterminado é o referente extralinguístico, ainda que a terminologia utilizada seja a da gramática tradicional, ou seja, indeterminação do sujeito. Desse modo, nos exemplos (1), (2) e (3), tem-se um referente extralinguístico indeterminado. Todos os casos podem ser parafrazeados por ‘se’: ‘talvez não **se** percebam os milagres...’; ‘com o dinheiro faz-**se** tudo que se precisa; se **se** tem uma fé convicta, o milagre acontece’.

Ao se afirmar que o referente extralinguístico é o elemento indeterminado, as formas de indeterminação analisadas neste trabalho alinham-se com as construções chamadas *R-impersonals* ou *human impersonal pronouns* por estudos mais recentes (Siewierska & Papastathi, 2011; Gast & van der Auwera, 2013). Entretanto, algumas características do português brasileiro, como língua parcialmente *pro-drop* e alta frequência de nomes gerais em formas de indeterminação (Amaral & Ramos, 2014), conduzem a uma descrição semântica específica, tal como será indicado na análise.

Neste artigo, discute-se a relevância de dois fatores semânticos no fenômeno da indeterminação do sujeito: presença ou ausência de anáfora e grau de indeterminação. Para isso, consideram-se tanto as variantes mencionadas, ‘a gente’, ‘você/ocê/cê’, nomes gerais (‘cara’, ‘fulano’, ‘homem’, ‘pessoa’, ‘pessoal’, ‘pessoas’, ‘povo’, ‘ser humano’), quanto outras já identificadas em vários trabalhos sobre o tema: ‘nós/nóis’, ‘eles/ês’, Ø+V3PS (verbo na terceira pessoa do singular sem sujeito explícito), Ø+V3PP (verbo na terceira

² Código de identificação do arquivo de áudio. Maiores informações se encontram na seção ‘Procedimentos metodológicos’.

pessoa do plural sem sujeito explícito), Ø+VINF (verbo no infinitivo sem sujeito explícito) e 'se'. Os dados analisados provêm de transcrições de entrevistas sociolinguísticas realizadas na cidade de Bambuí-MG, município do centro-oeste mineiro, localizado a 270 km de Belo Horizonte.

No que diz respeito à estrutura deste trabalho, inicialmente, é apresentado o referencial teórico, em que se discutem os conceitos mais relevantes para esta pesquisa. Tem-se, assim, um breve histórico da sociolinguística variacionista, seguido de discussões sobre a indeterminação do sujeito e dos nomes gerais.

A segunda parte trata da metodologia, isto é, uma descrição dos procedimentos realizados para que os dados fossem obtidos e, posteriormente, analisados. Nela está contida a descrição da cidade, o período em que ocorreram as gravações, a apresentação dos fatores analisados e a forma como as análises quantitativas foram realizadas. Após a finalização da metodologia, são apresentados e analisados os resultados encontrados, tanto quantitativa quanto qualitativamente. A seção contém análises de dados a partir de resultados obtidos com o uso do programa *GoldVarb X* (Taglimonte, 2017).

Dessa forma, o nosso objetivo principal é analisar a influência de dois fatores semânticos nas variantes que indeterminam sujeito, discutindo-se ainda a relevância que os nomes gerais têm no conjunto das variantes.

Sociolinguística variacionista

A Sociolinguística Variacionista constitui uma área do conhecimento que visa ao estudo da relação entre língua e sociedade (Labov, 2008). Dois conceitos importantes para essa área de estudo são os de 'variável' e de 'variante'. Variável é o lugar no sistema linguístico que permite identificar uma variação e variantes são as formas individuais que se alternam na expressão de uma variável (Coelho, Görski, Nunes de Souza, & May, 2015).

Neste artigo, a variável em análise é a indeterminação do sujeito e as variantes são as formas distintas em que essa indeterminação se apresenta. Nos estudos clássicos de linguística variacionista, considera-se que um fenômeno variável ocorre quando duas ou mais formas podem ocorrer no mesmo contexto com o mesmo significado ou com o mesmo valor de verdade (Labov, 2008). Entretanto, neste estudo, o valor é funcional, assim como proposto por Lavandera (2014). Essa autora mostra que a noção de variante foi criada, inicialmente, com base em estudos fonológicos e, devido a isso, a definição, muitas vezes, não consegue abarcar pesquisas voltadas para variantes sintáticas e morfológicas. Desse modo, segundo a autora, para se definir variantes que não sejam fonológicas, "[...] a proposta é enfraquecer a condição de que o significado deve ser o mesmo para todas as formas alternantes e substituí-las por uma condição de comparabilidade funcional"³ (Lavandera, 2014 p. 47, tradução nossa). Além disso, conforme aponta Silva (2004) ao tratar do mesmo tema, os campos de estudo da sociolinguística têm se aberto cada vez mais e, em trabalhos que contemplam as várias formas de se indeterminar o sujeito, por exemplo, "[...] deve-se estabelecer não apenas uma escolha entre duas alternativas, mas uma escala de possibilidades para aquilo que poderíamos chamar de um domínio funcional, e aí estudar a variação" (Silva, 2004, p. 71).

A indeterminação do sujeito

O fenômeno da indeterminação do sujeito tem sido tratado por vários trabalhos em diversas regiões do país, e, a partir dessas pesquisas, é possível observar que existem muito mais formas de se indeterminar o sujeito do que as gramáticas normativas prescrevem. Isso é o que também observa Neves (2000), ao tratar das formas que possibilitam o que a autora chama de 'referência genérica'. A partir de dados da língua escrita, Neves (2000) cita o pronome 'eles', verbos na terceira do plural, verbos na terceira pessoa do singular, verbos na terceira pessoa do singular mais o pronome 'se', o pronome 'você', o pronome 'eu', a primeira pessoa do plural, ou seja, 'nós' e, ainda, a forma 'a gente' e sintagmas nominais como 'a pessoa', 'o cara' e 'o pessoal', estes dois últimos, segundo a autora, utilizados mais comumente na linguagem popular.

Neves (2000) destaca que a forma 'eles' possui uma indeterminação parcial por se referir ao universo, somente, das terceiras pessoas e, ainda, que a forma 'nós' também não indetermina totalmente porque sempre haveria, pelo menos, uma referência determinada, o 'eu'. No caso de 'a gente', Neves (2000, p. 470) acrescenta que esta forma "[...] sempre deixa indicado o envolvimento da primeira pessoa no conjunto".

³ "Mi propuesta es debilitar la condición de que el significado deba ser el mismo para todas las formas alternantes, y reemplazarla por una condición de comparabilidad funcional".

As diferenças do grau de indeterminação do sujeito comentadas são registradas na análise variável dos dados que será apresentada mais adiante. Antes, porém, serão descritos os resultados de um conjunto de pesquisas que vêm, ao longo das últimas décadas, dedicando-se ao fenômeno em questão. Milanez (1982), Menon (1994), Setti (1997), Godoy (1999), Oliveira (2006), Carvalho (2010), Souza (2014) e Barbosa (2016) desenvolveram pesquisas com dados de diferentes localidades e mostram um amplo conjunto de formas que são utilizadas pelos falantes do português para indeterminar o sujeito.

Milanez (1982) pesquisa a fala culta paulistana a partir de dados do projeto NURC. Os itens analisados pela autora foram as expressões denominadas de itens lexicais que normalmente ocupam posição de sujeito ('a gente', 'você', 'eles', 'eu', 'nós', 'o/um indivíduo', 'o/um sujeito', 'a/uma pessoa', 'o/um cara' e 'a/uma pessoa') e as constituídas por formas verbais despronominalizadas ((\emptyset + 3ª p. sing.), (\emptyset + 3ª p. pl.), a construção com 'se', ou seja, (\emptyset + VERBO + se) e (\emptyset + infinitivo)). Os resultados apontam que, das 1058 ocorrências de expressões indeterminadoras, 38% ocorreram com formas despronominalizadas e 62% com sujeito lexical. Além disso, a grande maioria ocorreu com pronomes pessoais ('a gente', 'você', 'eles', 'eu'), em detrimento de recursos como 'o indivíduo', 'o sujeito', 'o cara' etc. O pronome 'a gente' foi o mais recorrente de todas as formas de indeterminação encontradas.

Baseando-se em Milanez (1982), Carvalho (2010) aponta alguns aspectos que servem para diferenciar a indeterminação da indefinição. A indeterminação restringe-se a seres humanos; refere-se a qualquer pessoa do discurso, ainda que sua marcação seja a uma das três pessoas do discurso; possui um caráter generalizador como uma de suas condições primárias; considera a quantificação como irrelevante; depende do contexto para ser interpretada e não identifica de modo exato as pessoas do discurso. Contrariamente à indeterminação, a indefinição refere-se a seres humanos e não humanos; não remete a outras pessoas do discurso, apenas à 3ª pessoa; toma a generalização como uma possibilidade; considera a quantificação como relevante; pode ser recuperada no nível frasal e é determinada, visto que uma das pessoas do discurso pode ser recuperada. No Quadro 1, apresenta-se um quadro com um resumo das propriedades da indeterminação e da indefinição:

Quadro 1. Propriedades da indeterminação e da indefinição.

Indeterminação	Indefinição
Restringe-se a seres humanos;	Refere-se a seres humanos e não humanos;
refere-se a qualquer pessoa do discurso;	remete apenas à 3ª pessoa do discurso;
possui caráter generalizador;	toma a generalização como uma possibilidade;
é irrelevante a quantificação;	é relevante a quantificação;
depende do contexto para ser identificada;	pode ser recuperada no nível frasal;
não identifica de modo exato as pessoas do discurso.	é determinada.

Fonte: Adaptado de Milanez (1982) e de Carvalho (2010).

Menon (1994), a partir de dados do NURC/SP, realiza um estudo que serviu de referência para muitos trabalhos que vieram posteriormente. Com base na sociolinguística variacionista, selecionou 12 variantes para serem estudadas, sendo elas: 'a gente', 'eles', 'eu', 'formas nominais', 'nós', 'se', 'você', 'vocês', 'voz passiva sem agente', 'voz passiva sintética', 'verbo na terceira pessoa do singular' e 'verbo na terceira pessoa do plural'. Vale ressaltar que Menon (1994) entende as formas nominais como aquelas constituídas de um artigo definido mais um substantivo e que servem como formas de generalização. Os exemplos encontrados pela autora foram: 'o camarada', 'o homem', 'o indivíduo', 'a pessoa', 'o sujeito', 'a turma', 'o cara', 'o cidadão' e 'o público'. A maior parte dessas formas têm sido incluídas como membros do conjunto dos nomes gerais do português, conforme será descrito adiante.

Menon (1994), assim como Milanez (1982), também discutiu em sua tese a questão da indeterminação, indefinição e impessoalização. A autora argumenta que a indeterminação se dá quando não se quer ou não se pode nomear o agente, ou seja, o referente extralinguístico. A indefinição, semelhante à proposta de Milanez (1982), também remete à quantificação, pois a autora explica que um sujeito indefinido diz respeito

a uma classe de indivíduos em que todos possuem características semelhantes. A impessoalização, ainda que seja um processo tal qual a indeterminação, ou seja, aquele em que não se quer nomear o referente, abarca alguns tipos específicos de verbos, semelhante ao ilustrado por Cunha e Cintra (2008) acima.

Setti (1997) faz uma análise sobre as formas de indeterminação do sujeito em uma abordagem variacionista, a partir de dados do VARSUL referentes às capitais do Sul do Brasil: Curitiba, Florianópolis e Porto Alegre. As variantes estudadas como indeterminadoras do sujeito são: 'a gente', 'eles', 'eu', 'formas nominais', 'nós', 'se', 'tu', 'você(s)', 'voz passiva sem agente', 'voz passiva sintética', 'verbo na 3ª pessoa do singular' e 'verbo na 3ª pessoa do plural'. Constatou-se que os falantes de Curitiba indeterminam muito mais o sujeito e que nesse local não existe ocorrência do pronome 'tu'. Em Porto Alegre e Florianópolis, existe uma variação no uso de 'tu' e 'você', entretanto, em Porto Alegre o 'tu' aumenta, e em Florianópolis esse aumento ocorre com 'você'.

Godoy (1999) investiga o tema 'indeterminação do sujeito' em cidades do interior do Paraná, sendo elas: Irati, Londrina e Pato Branco, a partir de dados do VARSUL. As variantes são: 'a gente', 'eles', 'eu', 'formas nominais', 'nós', 'se', 'tu', 'você (s)', 'VPSA (voz passiva sem agente)', 'VPASSINT (voz passiva sintética)', '0V3PS (verbo na 3ª pessoa do singular)' e '0V3PP (verbo na 3ª pessoa do plural)'. A autora pôde observar que a variante 'você' é a mais empregada em índices percentuais. Acrescenta-se, ainda, que as formas nominais mostram um equilíbrio nas faixas etárias, bem como 0V3PP, conferindo-lhes característica de variáveis estáveis. Além disso, verifica-se que existe certa tendência de o falante preencher a posição de sujeito e, como já citado em outras pesquisas, que o PB tem perdido sua característica de sujeito nulo.

Carvalho (2010), ao pesquisar em Salvador as formas de indeterminação do sujeito no português oral, teve como base metodológica a 'teoria variacionista'. São analisados 44 inquiridos, sendo 32 do PEPP – Programa de Estudo sobre o Português Popular falado em Salvador e 12 do NURC/SSA – Projeto Norma Urbana Culta de Salvador. As formas selecionadas no *corpus* são as previstas pela gramática, tais como: 'verbo na terceira pessoa do plural (\emptyset + V3PP)', 'verbo na terceira pessoa do singular mais o pronome 'se' (\emptyset + V3PS + SE)', 'verbo no infinitivo pessoal (\emptyset + VINF)' e outras estratégias como 'você', 'a gente', 'nós', 'eles', 'eu', 'voz passiva sem agente (VPSA)', 'voz passiva sintética (VPASSINT)', e as 'formas nominais (o cara, o indivíduo, o sujeito, nego, o pessoal, as pessoas, o povo, o público, o homem e gente)'. Constatou-se que em Salvador as formas 'você', 'a gente' e as 'formas nominais' foram as mais utilizadas pelos falantes que compõem os *corpora*.

Por último, Barbosa (2016) encontrou no falar ludovicense dez formas que indeterminam o sujeito. São elas: 'eu', 'tu', 'você', 'a gente', 'nós', 'eles', 'formas nominais', 'vocês', 'se' e 'verbo na terceira pessoa do plural'. Assim como na maioria dos trabalhos realizados sobre o tema, as formas nominais ocuparam a terceira posição em número de ocorrências, reforçando, mais uma vez, a tese de que essas formas ocupam uma posição estável na descrição do fenômeno de indeterminação.

Todos os trabalhos citados corroboram o fato de que existem muito mais formas de se indeterminar o sujeito do que as propostas pela gramática tradicional. Além disso, todos comprovam também que as formas nominais têm tido papel representativo nessas pesquisas. Entretanto, nenhum deles dá atenção especial ao peso que fatores semânticos exercem nas formas nominais, aqui vinculadas ao conjunto dos nomes gerais.

Os nomes gerais

O estudo dos nomes gerais, para esta pesquisa, tem como base os trabalhos de Halliday e Hasan (1976), Amaral e Ramos (2014), Amaral (2017) e Mihatsch (2017). Apesar de existirem algumas pesquisas que lidam com elementos de traço [-humano], tais como Fronek (1982), Mihatsch (2002), Mihatsch (2006), Ramos (2013) etc, neste trabalho nos concentramos nos itens de traço [+humano].

Em geral, Halliday e Hasan (1976) são considerados os precursores sobre os estudos dos nomes gerais, ou, como denominado por eles, *general nouns*⁴. Para os autores, esses elementos constituem um caso limítrofe entre um item lexical e um item gramatical e possuem traços mínimos de significado. No conjunto dos elementos de traço [+humano], os autores incluem: *people, person, man, woman, child, boy, girl*. Além disso, do ponto de vista semântico, os nomes gerais seriam membros superordenados de grandes conjuntos lexicais e, de um ponto de vista gramatical, a combinação desses itens com um determinante resultaria em algo semelhante a um item referencial⁵.

⁴ Autores mais antigos, como Havers (1931), já chamavam a atenção para elementos desse tipo na linguagem coloquial. Para uma abordagem de caráter contrastivo, veja-se também Koch e Oesterreicher (2007).

⁵ Para uma discussão a respeito da posição dos nomes gerais em uma hierarquia lexical, veja-se Mihatsch (2006).

Francis (2003), ao tratar, ainda, dessa possibilidade de referência, explicita que o grupo nominal em que se encontram os nomes gerais funciona “[...] como uma proforma ou um item referencial” (Francis, 2003, p. 196). Nesse sentido, um item desse tipo “[...] é sempre apresentado como uma informação ‘dada’ em sua oração, em relação à qual a nova mensagem – o foco informacional – é formulada” (Francis, 2003, p. 196-197, grifo do autor).

Amaral e Ramos (2014) analisam o comportamento morfossintático, semântico-textual e sociolinguístico de quatro nomes gerais: ‘coisa’, ‘negócio’, ‘trem’, ‘pessoa’. Em muitos casos, são elementos dêiticos ou anafóricos, e, portanto, podem ser parafraseados por pronomes. Os autores destacam ainda que esses itens possuem outras funções e, ao citar Mihatsch (2006), elucidam que um nome geral é utilizado quando não há uma designação, naquele momento, acessível ao falante. Outra função é a de que eles podem substituir um referente que seja incômodo para aquele que fala ou ainda para fazer uma referência genérica.

Amaral (2017), em um estudo contrastivo com dados do português e do espanhol, demonstra que existe certa diferença de aceitabilidade de alguns nomes gerais nessas línguas. Além disso, observa como o item ‘pessoa’ possui uma situação diferente dos demais nomes explorados, e como seu uso como forma de indeterminação do sujeito é recorrente. Ainda segundo o autor, nomes gerais como ‘ser humano’ e ‘indivíduo’ são mais comuns em textos acadêmicos, e ‘sujeito’ e ‘pessoa’ são mais frequentes em registros ficcionais e orais. Tendo isso vista, essa constatação pode ser corroborada por este estudo, visto que, conforme ficará evidente na análise, não foi encontrada, em nosso *corpus*, a forma ‘indivíduo’, e, ainda, ‘ser humano’ ocorreu raríssimas vezes.

Mihatsch (2017) pesquisa a gramaticalização dos nomes gerais, isto é, nomes que apresentam uma função tipicamente gramatical. A pesquisadora apresenta um estudo de itens com traços [+ humano] e salienta que existe um grupo que contém nomes bastante eruditos e outro que contempla itens mais usuais da língua, como o caso de ‘pessoa’. A autora conclui que ‘pessoa’, por exemplo, já mostra indícios de pronominalização.

Dessa forma, todos os trabalhos citados anteriormente fazem este tipo de explicitação quando se trata de sujeito indeterminado. Nomes gerais, assim como as formas de indeterminação do sujeito, são utilizadas pelos falantes quando não se quer, não se pode ou não se sabe nomear o referente. Assim, uma forte relação se dá entre os nomes gerais e as formas de indeterminação e, por isso, faz-se necessário um estudo aprofundado do tema. Além disso, como há outras formas de indeterminação do sujeito, é possível supor que fatores linguísticos, especialmente de caráter semântico, possam influenciar na escolha das variantes.

Procedimentos metodológicos

Os dados analisados neste trabalho foram coletados em Bambuí, cidade localizada no centro-oeste mineiro que, de acordo com o censo de 2016, realizado pelo Instituto Brasileiro de Geografia e Estatística [IBGE] (2017), possui 22.734 habitantes. É uma cidade interiorana, localizada a 270 km da capital de Minas Gerais, Belo Horizonte.

Tendo sido fundada em 1886, não se tem notícia de estudos anteriores sobre a fala de Bambuí. Por ser uma cidade interiorana, observa-se que a localidade sofre menos influências linguísticas que os grandes centros urbanos, tal como a localidade do interior da Bahia, em que Ponte (2008) investigou o mesmo fenômeno.

Os dados analisados neste trabalho foram coletados durante os meses de janeiro e fevereiro de 2017. Foram realizadas gravações de entrevistas semiespontâneas, com informantes distribuídos nas faixas etárias 1, a qual contempla informantes de 18 a 25 anos, faixa etária 2, que corresponde a informantes de 30 a 45 anos, e faixa etária 3, para a qual selecionaram-se informantes acima de 50 anos.

No conjunto das variáveis internas, foram incluídas: a) tempo verbal; b) modo; c) presença ou ausência de anáfora; d) grau de indeterminação; e) tipo de oração; f) presença ou ausência de locativo; g) mais ou menos coletivo. Neste artigo, discutem-se a presença ou ausência de anáfora e o grau de indeterminação, variáveis que foram consideradas estatisticamente relevantes.

No que concerne aos critérios adotados para a delimitação da variável presença ou ausência de anáfora, consideraram-se aqueles adotados por Santana (2006) e Carvalho (2010), sendo eles:

a) seleção de apenas um turno de fala, isto é, não deveria haver nenhuma interferência do entrevistador;

- b) limite de dez linhas para cada turno de fala;
- c) ausência de mudança do tema tratado – caso houvesse, a ocorrência seria considerada a primeira de uma nova série.

Apesar de ser uma análise semelhante à dos autores citados, optou-se por considerar apenas três formas para classificar os dados: a) retomada anafórica com o mesmo núcleo, ou seja, no caso das variantes que possuem pronomes, por exemplo, só foi considerada como retomada quando o informante repetiu uma forma idêntica à citada anteriormente; b) ausência de anáfora, isto é, quando o informante não retoma nada ou muda a forma de retomada e, por fim; c) retomada anafórica com outro elemento, a qual ocorreu com aquelas formas que não possuem sujeito explícito, pois, muitas vezes, o informante utiliza a variante ‘a gente’, por exemplo, e retoma, apenas, com um verbo no plural.

No caso da variável grau de indeterminação, partiu-se do trabalho de Siewierska e Papastathi (2011). Os autores, em uma pesquisa sobre a terceira pessoa do plural em inglês, apresentam cinco maneiras de se indeterminar a terceira pessoa do plural, sendo elas: universal, corporativa, vaga, inferencial e existencial específica. Tomando tal classificação e o que foi apresentado por Santana (2006) e Carvalho (2010), estabeleceram-se três classificações para este trabalho, sendo: a) indeterminação parcial com referente explícito; b) indeterminação parcial com referente implícito; c) indeterminação total.

A indeterminação parcial com referente explícito contempla aqueles casos em que é possível inferir de alguma forma o referente e que, no texto, ocorre alguma marca explícita. Esses casos assemelham-se ao que Siewierska e Papastathi (2011) classificam como inferencial. A segunda, no entanto, diz respeito àqueles casos em que é possível inferir de quem se trata pelo contexto, mas não existe nenhuma marca que seja possível identificar, tal como o que Siewierska e Papastathi (2011) nomeiam de existencial específico. Já a indeterminação total diz respeito àqueles casos, tal como nomeado por Siewierska e Papastathi (2011) de universal, em que não é possível nenhum tipo de inferência e, ainda, servem para classificar qualquer pessoa em qualquer lugar do mundo.

Os dados coletados foram codificados e submetidos ao programa de análise de dados variáveis *Goldvarb X* (Taglimonte, 2017). Neste artigo, apresentam-se os resultados estatisticamente significativos após uma rodada que contemplava, como variantes, os nomes gerais e as demais formas de indeterminação⁶. A seguir, são discutidos os fatores considerados relevantes pela análise estatística: presença ou ausência de anáfora e grau de indeterminação.

Resultados gerais

Na análise geral do *corpus* de dados de Bambuí, foram identificadas 1357 ocorrências distribuídas entre as formas de indeterminação. A variante ‘a gente’ foi a forma mais frequente, seguida de ‘você/ocê/cê’ e, em terceiro lugar, os nomes gerais, tal como mostra a Tabela 1.

Tabela 1. Número de ocorrências e porcentagem de cada uma das variantes.

Item	N	%
a gente	376	27,7
você/ocê/cê	325	23
nomes gerais	191	14,1
nós/nóis/1ª pessoa do plural	134	9,9
eles/ês	122	9,0
Ø+V3PS	122	9,0
Ø+V3PP	39	2,9
Ø+VINFINF	36	2,7
se	12	0,9
TOTAL	1357	100,0

Fonte: Oliveira (2018)

Os resultados encontrados alinham-se com vários estudos sobre o tema, em que as construções com nomes gerais, chamadas por outros autores de formas nominais, ocupam um terceiro lugar. Pela importância desses itens, pouco explorados até o momento nos estudos sobre indeterminação do sujeito, apresenta-se uma discussão a respeito dos resultados sobre a influência de fatores semânticos, tal como explicado anteriormente.

⁶ Para discussão completa, veja-se Oliveira (2018).

Resultados referentes à variável presença ou ausência de anáfora

De acordo com a metodologia apresentada, em que se explica a variável linguística ‘presença ou ausência de anáfora’, selecionaram-se três maneiras de se classificar os itens no *corpus* para este grupo: a) retomada anafórica com o mesmo núcleo; b) ausência de anáfora; c) retomada anafórica com outro elemento. Entretanto, quando os dados foram rodados no *software Goldvarb X*, houve um *knockout* nessa variável e, dessa forma, reagrupou-se a variável ‘retomada anafórica com o mesmo núcleo’ com ‘retomada anafórica com outro elemento’. Tendo isso em vista, optou-se pelo reagrupamento, porque o número de dados do que foi denominado de ‘retomada anafórica com outro elemento’ era relativamente alto, não sendo, portanto, interessante para a pesquisa fazer uma exclusão desses resultados. Feita essa junção, obtêm-se os seguintes resultados (Tabela 2):

Tabela 2. Influência do fator ‘presença ou ausência de anáfora’ na indeterminação do sujeito por meio de nomes gerais.

Fator	N	(%)	P.R.
Ausência de anáfora	174/949	18,3	0,62
Presença de anáfora	17/408	4,2	0,25
Total	191/1357	14,1/100	

Fonte: Oliveira (2018).

De acordo com a Tabela 2, observa-se que, no caso dos nomes gerais, predominantemente, quando um item é utilizado, não há uma retomada com esse mesmo item. Em Carvalho (2010), a presença é mais frequente que a ausência, mas nossa análise se diferencia da desse autor. No trabalho de Carvalho (2010), foi feita uma subdivisão dos dados em preenchimento de sujeito e preenchimento de sujeito pronominal explícito ou implícito. No primeiro caso, o autor contemplou apenas as variantes ‘a gente’, ‘eles’, ‘nós’, ‘eu’ e ‘você’. Segundo Carvalho (2010, p. 90), “[...] os demais casos não foram considerados, uma vez que são estratégias sem sujeito preenchido ou então não permitem a inclusão do sujeito implícito”. Já no segundo caso, isto é, preenchimento de sujeito pronominal explícito, “[...] a posição de sujeito é preenchida por um dos pronomes analisados como estratégias de indeterminação do sujeito” (Carvalho, 2010, p. 91) e, no caso do preenchimento de sujeito pronominal implícito, “[...] a posição de sujeito não é preenchida por um dos pronomes analisados, contudo, através do contexto, sabe-se que ele está oculto” (Carvalho, 2010, p. 91). Em nosso trabalho, não foi feita essa subdivisão. O que se observou para esta análise é se um item era retomado da mesma forma, de forma distinta ou se não era retomado.

A partir desses dados, o que se pode inferir é que, quando o falante utiliza o nome geral, ele está, predominantemente, em uma oração absoluta, não havendo, portanto, uma construção que retome a anterior. Isso parece ser, portanto, um grande influenciador do resultado. Nos exemplos (5), (6) e (7) abaixo, vê-se que a oração absoluta não apresenta uma subordinada. Também não há uma retomada anafórica.

(5) o **povo** fala me[s]mo... sabe que eu num sei (BAM 11)

(6) p[r]a joga[r] na internet ixatamente é/é/é isso aí até é chato né a/e/e/o tipo o **ser humano** num podia chega[r] a esse ponto não (BAM 19)

(7) intão é aonde o **pessoal** ainda brinca até hoje lá o pessoal/os minino ainda brinca na rua brinca de pique-isconde é um lugar que as crianças podem ficar solta na rua (BAM 21)

Outro fator que parece ter proporcionado tal resultado é que, no caso dos nomes gerais, por favorecerem uma indeterminação completa, parece não ser necessária uma retomada desse sujeito, visto que, quando se usa tal forma, é porque já se objetiva indeterminar totalmente, conforme exemplo (8).

(8) num tô falano o que o po/o que o **home[m]** fala eu tô falano o que Deus de[i]xô (BAM 5)

É importante salientar que, apesar de nosso resultado não ter se assemelhado com o de outros autores, o resultado obtido, isto é, de favorecimento da ausência de anáfora, é relevante para as pesquisas sobre pronominalização de nomes gerais, visto que Siewierska e Papastathi (2011) salientam que o pronome de terceira pessoa como forma de indeterminação do sujeito não possui um antecedente aberto no discurso e Gast e van der Auwera (2013) reforçam que uma das características mais marcantes dos pronomes de terceira pessoa (*human impersonal pronouns*, para os autores) é que eles não introduzem referentes discursivos e, portanto, não são retomados anaforicamente.

No caso dos resultados de Bambuí, o favorecimento da ausência de anáfora na indeterminação do sujeito por meio de nomes gerais também poderia estar relacionado a certo conservadorismo na fala da cidade, motivado pela caracterização sociolinguística acima. Esse argumento se apoia nos trabalhos mais recentes que têm indicado a anáfora como um indício importante no fenômeno de pronominalização dos nomes gerais (Mihatsch, 2017).

Grau de indeterminação

A segunda variável selecionada como favorecedora dos nomes gerais é o ‘grau de indeterminação’. De acordo com a metodologia apresentada, ‘a indeterminação parcial com referente explícito’ se dá quando existe alguma palavra, expressa pelo informante, que torna possível retomar de quem se está falando. A ‘indeterminação parcial com referente implícito’ ocorre, no entanto, quando se recupera de quem se está falando pelo contexto. A ‘total’, contrariamente às outras, não torna possível a recuperação. A Tabela 3 apresenta os resultados obtidos a partir do *corpus*:

Tabela 3. Influência do fator ‘grau de indeterminação’ na indeterminação do sujeito por meio de nomes gerais.

Fator	N	%	P.R
Indeterminação total	59/426	13,8	0,54
Parcial com referente implícito	112/673	16,6	0,53
Parcial com referente explícito	20/258	7,8	0,37
Total	191/1357	14,1/100	

Fonte: Oliveira (2018).

A partir da Tabela 3, nota-se um leve favorecimento da indeterminação total e da parcial com referente implícito nos nomes gerais. Salienta-se que Carvalho (2010) também encontrou que a indeterminação completa é favorecedora das formas nominais.

O que se observa, a partir dos dados, é que formas indeterminadoras como ‘a gente’, ‘você’ e ‘nós’ incluem mais o falante ou o interlocutor que os nomes gerais, conforme exemplos (9) e (10) abaixo:

(9) uai não a gente igual ês fala **a gente** morre... o que a gente vê vai vive[r] ot[r]a vida na frente eu acho que não morreu [a]cabô eu acho que num tem ot[r]a vida não (BAM 14)

(10) intão quer dizer ela é ruim e boa... ela pode algumas pessoas infruencia[r] mais num é tanto porque as pessoas que tem que infruencia igual cê vê os presi/os presídio o **cara** vai preso ele sai da cadeia pio[r] (BAM 08)

Vê-se que, com a forma ‘a gente’, na maioria dos casos, sempre haverá a inclusão de quem fala e, no caso dos nomes gerais, mesmo aqueles que representem um grupo de pessoas, tal como ‘povo’, ‘pessoa(s)’ ou ‘pessoal’ isso não tende a ocorrer. Apesar disso, Moltmann (2010) ressalta que em alguns casos o falante usa uma forma indeterminadora, tal como ‘pessoa’, por exemplo, para se projetar, a partir de suas experiências pessoais.

Conclusão

Este artigo discutiu a relevância de dois fatores semânticos no emprego de nomes gerais como estratégia de indeterminação do sujeito: presença ou ausência de anáfora e grau de indeterminação. A partir dos resultados de Oliveira (2018), constatou-se como os nomes gerais cumprem um papel relevante para indeterminar o sujeito, pois, considerando sua frequência, ocupam a terceira posição entre as formas de indeterminar o sujeito, atrás somente de formas já pronominalizadas, como ‘a gente’ e ‘você/ocê/cê’.

Considerando o fato de que os nomes gerais têm sido incluídos como elementos entre o léxico e a gramática e que alguns deles estão passando por um processo de pronominalização, os resultados obtidos mostram como os dados próprios da fala semiespontânea de uma localidade do interior de Minas Gerais atestam algumas e contestam outras características já encontradas em estudos recentes.

Este trabalho contribui também para mostrar como fatores semânticos são relevantes no uso dos nomes gerais como estratégia de indeterminação. Se os nomes gerais têm potencial anafórico e muitas vezes encapsulador (Francis, 2003), quando empregados como forma de indeterminar o sujeito podem ou não funcionar como antecedentes anafóricos, o que constitui um fenômeno variável no português brasileiro. Isso se explica por sua proximidade com elementos pronominais – eles estão, conforme comentado no início, entre o léxico e a gramática. Esse resultado se alinha com o fato de que uma indeterminação total ou

parcial com referente implícito favorece a ocorrência do nome geral ao se indeterminar o sujeito. Em todos os casos, o sintagma em que ocorre com a função de sujeito delimita (um conjunto de) entidades humanas, mas com perda de referencialidade.

Referências

- Amaral, E. T. R. (2017). Estudio contrastivo de nombres generales para humanos en español y en portugués. *Lingüística y literatura*, 72(1), 54-79.
- Amaral, E. T. R., & Ramos, J. (2014). *Nomes gerais no português brasileiro*. Belo Horizonte, MG: Faculdade de Letras da UFMG.
- Barbosa, A. B. (2016). *A indeterminação do sujeito no falar ludovicense* (Dissertação de Mestrado). Universidade Federal do Paraná, Curitiba.
- Carvalho, V. (2010). *Você, a gente et alia indeterminam o sujeito em Salvador* (Dissertação de Mestrado). Universidade do Estado da Bahia, Salvador.
- Coelho, I. L., Görski, E. M., Nunes de Souza, C. M., & May, G. H. (2015). *Para conhecer sociolinguística*. São Paulo, SP: Contexto.
- Cunha, C., & Cintra, L. F. L. (2008). *Nova gramática do português contemporâneo* (5a ed.). Rio de Janeiro, RJ: Lexicon.
- Francis, G. (2003). Rotulação do discurso: um aspecto da coesão lexical de grupos nominais. In M. C. Cavalcante, B. B. Rodrigues, & A. Ciulla (Ed.), *Referenciação* (p. 191-238). São Paulo, SP: Contexto.
- Fronek, J. (1982). 'Thing' as a function word. *Linguistics*, 20(9-10), 633-654.
- Gast, V., & van der Auwera, J. (2013). Towards a distributional typology of human impersonal pronouns, based on data from European languages. In D. Bakker, & M. Haspelmath (Ed.), *Languages across boundaries: studies in memory of Anna Siewierska* (p. 119-158). Berlin, DE: Gruyter Mouton.
- Godoy, M. A. M. (1999). *A indeterminação do sujeito no interior paranaense: uma abordagem sociolinguística* (Dissertação de Mestrado). Universidade Federal do Paraná, Paraná.
- Halliday, M., & Hasan, R. (1976). *Cohesion in english*. London, NY: Longman.
- Havers, W. (1931). *Handbuch der erklärenden Syntax: ein Versuch zur Erforschung der Bedingungen und Triebkräfte in Syntax und Stilistik*. Heidelberg, DE: Carl Winters Universitätsbuchhandlung.
- Instituto Brasileiro de Geografia e Estatística [IBGE]. (2017). Recuperado em 10/11/2018 de <https://cidades.ibge.gov.br/v4/brasil/mg/bambui/panorama>
- Koch, P., & Oesterreicher, W. (2007). *Lengua hablada en la Romania: español, francés, italiano*. Madrid, ES: Gredos.
- Labov, W. (2008). *Padrões sociolinguísticos*. São Paulo, SP: Parábola Editorial.
- Lavandera, B. R. (2014). *Variación y significado. Y discurso*. Ciudad Autónoma de Buenos Aires: Paidós.
- Menon, O. P. S. (1994). *Analyse sociolinguistique de l'indétermination du sujet dans le portugais parlé au Brésil, à partir des données du NURC/SP* (Tese de Doutorado). Université de Paris VII, Paris.
- Mihatsch, W. (2002). De plantas, animales y (otros) objetos: lexemas cultos genéricos entre léxico y gramática. In A. Veiga, G. Pereira, M. González Pereira, & M. Souto Gómez (Eds.), *Léxico y gramática* (p. 237-248). Lugo, ES: Tris Tram.
- Mihatsch, W. (2006). *Kognitive Grundlagen lexikalischer Hierarchien: untersucht am Beispiel des Französischen und Spanischen*. Tübingen, DE: Max Niemeyer.
- Mihatsch, W. (2017). Les noms d'humains généraux aux limites de la grammaticalisation. *Syntaxe e sémantique*, 1(18), 67-99.
- Milanez, W. (1982). *Recursos de indeterminação do sujeito* (Dissertação de Mestrado). Universidade Estadual de Campinas, Campinas.
- Moltmann, F. (2010). Generalizing detached self-reference and the semantics of generic 'one'. *Mind and language*, 25(4), 440-473.
- Neves, M. H. M. (2000). *Gramática de usos do português*. São Paulo, SP: Unesp.
- Oliveira, K. R. (2006). *'Nós', 'a gente' e o clítico se como estratégias de indeterminação do sujeito no português* (Dissertação de Mestrado). Universidade Federal de Uberlândia, Uberlândia.
- Oliveira, F. C. (2018). *Nomes gerais e outras formas de indeterminação do sujeito no português oral de Bambuí-MG* (Dissertação de Mestrado). Universidade Federal de Minas Gerais, Belo Horizonte.

- Ponte, V. (2008). *A indeterminação do sujeito no português rural do interior do estado da Bahia* (Dissertação de Mestrado). Universidade Federal da Bahia, Salvador.
- Ramos, J. M. (2013). O surgimento de um nome geral: a lexia 'trem' no dialeto mineiro. In J. M. Ramos, & S. M. Coelho. *Português brasileiro dialetal: temas gramaticais* (p. 137-147). Campinas, SP: Mercado de Letras.
- Santana, N. M. O. (2006). *A indeterminação do sujeito no português rural do semi-árido baiano* (Dissertação de Mestrado). Universidade Federal da Bahia, Salvador.
- Setti, A. C. R. (1997). *A indeterminação do sujeito nas três capitais do Sul do Brasil* (Dissertação de Mestrado). Universidade Federal do Paraná, Curitiba.
- Siewierska A., & Papastathi, M. (2011). Towards a typology of third person plural impersonals. *Linguistics*, 49(3), 575-610.
- Silva, V. L. P. (2004). Relevância das variáveis linguísticas. In M. A. Mollica, & M. L. Braga, (Org.), *Introdução à sociolinguística: o tratamento da variação* (p. 67-71). São Paulo, SP: Contexto.
- Souza, S. S. (2014). *Formas de indeterminação do sujeito: uma comparação entre as comunidades rurais Martinha e Casinhas (BA)* (Dissertação de Mestrado). Universidade Estadual de Feira de Santana, Feira de Santana.
- Taglimonte, S. A. (2017). *Goldvarb X*. Recuperado de <http://individual.utoronto.ca/tagliamonte/goldvarb.html>