

Modelação computacional de padrões variáveis de concordância em português¹

Computational modeling of variable agreement patterns in Portuguese

Leonel Figueiredo de Alencar
Universidade Federal do Ceará

Resumo

Este trabalho descreve a implementação de um *parser* para um fragmento do português no XLE, sistema que constitui o estado da arte para o desenvolvimento de gramáticas computacionais baseadas no modelo gerativo da LFG. O fragmento abrange os seguintes aspectos da concordância verbal, sujeitos à variação dialetal: (i) concordância sintática do chamado pronome *a gente*, com formas verbais de terceira pessoa do singular e semântica com a primeira e terceira do plural; (ii) concordância do pronome *tu*, com a segunda e a terceira do singular e do pronome *vocês*, com a segunda e terceira do plural; (iii) concordância semântica de um DP definido no plural, com a primeira pessoa do plural. Para derivar os diferentes padrões de concordância nesse domínio, constatados no português padrão e em alguns dialetos não padrão tanto do português do Brasil quanto do português europeu, são propostas estruturas de traços para os DPs e flexões verbais em que se distingue entre traços sintático-formais e traços semântico-discursivos. Mostra-se que as diferenças entre os dialetos investigados em relação a (i) – (iii) podem ser derivadas por meio da ativação ou desativação desses traços.

Palavras-chave

Linguística computacional, Gramática Léxico-Funcional, Concordância sintática, Concordância semântica, Sintaxe dialetal, Pronomes.

Abstract

This paper describes the implementation of a parser for a Portuguese fragment on XLE, the state of the art system in computational grammar development within the LFG generative framework. This fragment comprises the following subareas of verbal agreement that are subject to dialectal variation: (i) syntactic agreement of the so called pronoun *a gente* with 3rd person singular and semantic agreement with 1st and 3rd plural verb forms; (ii) agreement between the 2nd person pronoun *tu* with 2nd and 3rd person singular and between the 2nd person plural pronoun *vocês* with 2nd and 3rd person plural verb forms; (iii) semantic agreement between a definite plural DP and the 1st person plural. In order to derive the different patterns in this domain that can be observed in Standard Portuguese and in some non-standard dialects from both Brazilian Portuguese and European Portuguese, feature structures are proposed for DPs and verbal inflections in which a distinction is made between formal-syntactic features and discourse-semantic features. It is shown that the differences observed between the investigated dialects in respect to (i) – (iii) can be derived by activating or deactivating these features.

Keywords:

Computational linguistics, Lexical-Functional Grammar, Syntactic agreement, Semantic agreement, Dialect syntax, Pronouns.

1. Introdução

Conforme a gramática tradicional, o verbo, em português, “concorda com o seu sujeito em número e pessoa” (LUFT, 1986, p. 22). Desse modo, se a forma verbal em (1) é classificada como terceira pessoa do singular (doravante 3S), parece necessário atribuir esses traços também ao sintagma [a gente], para que a relação de concordância possa ser instanciada. Analogamente, em (2), o quantificador aparentemente necessita da especificação de terceira pessoa do plural (3P), de modo a poder concordar com o verbo. Uma comparação de (2) com (3a) e (3b) parece corroborar que [todos] esteja especificado como 3P.

- (1) A gente espera.
- (2) Todos esperam.
- (3) a. Os brasileiros esperam.
b. Todos os brasileiros esperam.

Na construção de um analisador sintático automático (*parser*) de uma língua natural como a portuguesa, a concordância é um dos fenômenos a ser modelados computacionalmente. Na ciência da computação, um *parser* de uma língua formal L é um programa capaz de determinar algoritmicamente, com base em uma gramática G , para uma cadeia s arbitrária, se s pertence ou não a L e, em caso afirmativo, atribuir a s uma estrutura, conforme G . Aplicando essa definição ao *parsing* sintático da linguagem natural, a função da concordância resume-se, à primeira vista, em excluir construções como (4) e licenciar sentenças como (1)–(3).

- (4) *Os brasileiros espero.

No âmbito do processamento automático da linguagem natural, os *parsers* sintáticos integram a arquitetura de sistemas de tradução automática e extração de informações, entre outros (NEDERHOF; SATTA, 2010). Do ponto de

vista linguístico, uma das aplicações mais importantes de um *parser* é a análise sintática de *corpora*, visando à construção de florestas sintáticas (*treebanks*), que têm sido cada vez mais usadas para testar a validade empírica de teorias gramaticais (PALMER; XUE, 2010; HAJIČOVÁ *et al.*, 2010). Nesse contexto particular, parece que a concordância não precisaria ser modelada, uma vez que, abstraindo de casos como erros relativos à performance, todas as sentenças de um *corpus* são gramaticais em relação à variedade da língua de que o *corpus* constitui uma amostra (LEMNITZER; ZINSMEISTER, 2006, p. 55-57).

Na verdade, o papel da concordância no *parsing* sintático é mais amplo, uma vez que constitui um dos meios de que uma língua como a portuguesa dispõe para restringir o mapeamento de funções sintáticas sobre papéis semânticos e determinar relações de escopo (KAPLAN, 2004). Por exemplo, é a concordância que determina a incidência do quantificador *todos* sobre [os meninos] ou [o bolo] em (5a) e (5b), ao mesmo tempo em que permite caracterizar (5c) e (5d) como agramaticais. Desse modo, trata-se de mecanismo indispensável na construção automática de representações semânticas para as sentenças analisadas sintaticamente.

- (5) a. Os meninos comeram todos o bolo.
 b. Os meninos comeram todo o bolo.
 c. *Os meninos comeram todas o bolo.
 d. *Os meninos comeram toda o bolo.

Por outro lado, a concordância, no *parsing* sintático, desempenha um papel importantíssimo na redução da ambiguidade. Por exemplo, um *parser* baseado no sistema de anotação sintática do *corpus* CORDIAL que não implementasse a concordância poderia atribuir a (4) as representações sintáticas (6) e (7), em que o sintagma nominal (NP) [os brasileiros] funciona, respectivamente, como sujeito (SBJ) e objeto direto (ACC) topicalizado, esse último coindexado ao traço de movimento A' simbolizado por *ICH*, abreviatura de “interpret constituent here” (CARRILHO *et al.*, 2011).²

(6) (IP-MAT (NP-SBJ (D-P os) (N-P brasileiros)) (V-P espero) (NP-ACC *))

(7) (IP-MAT (NP-1 (D-P os) (N-P brasileiros)) (NP-SBJ *pro*) (V-P espero) (NP-ACC *ICH*-1))

Cada uma dessas estruturas projeta uma representação semântica diferente. Evidentemente, (6) é implausível, dada a agramaticalidade de (4) enquanto instanciação do padrão canônico sujeito – verbo do português. A implementação da concordância na análise sintática automática, portanto, ao diminuir a quantidade de estruturas ambíguas, facilita a tarefa de compilação de uma floresta sintática, uma vez que os *parsers* de línguas naturais, por não incorporarem informações de natureza pragmático-discursiva, propendem naturalmente para a geração de um número excessivo de análises (NEDERHOF; SATTA, 2010).

A modelação computacional da concordância, porém, não é, como sugerem os exemplos de (1) a (3), uma tarefa trivial. Além das dependências de longa distância implicadas por exemplos do tipo de (5a), vários outros fatores contribuem para a complexidade desse fenômeno. No domínio da língua padrão, uma dificuldade é a chamada concordância ideológica ou silepse, como em (8), em que o verbo concorda não com a “forma gramatical” do sujeito, mas com o “sentido” ou “ideia” que este expressa (CUNHA; CINTRA, 1985, p. 614). Trata-se de um tipo de “figuras de sintaxe”, que discrepa das “estruturas frásicas tidas por modelares” em razão de um “empenho de maior expressividade” que substitui a “coesão gramatical” por uma “coesão significativa” (CUNHA; CINTRA, 1985, p. 602).

- (8) a. Todos esperamos.
- b. Todos os brasileiros esperamos.
- c. Os brasileiros esperamos.

Outra dificuldade decorre da variação dialetal, de que damos em (9) um primeiro exemplo, formalmente paralelo aos de (8):

- (9) A gente esperamos.

Na gramática tradicional, ocorrências com *a gente* do tipo de (9) não recebem o mesmo tratamento que exemplos do tipo de (8), como salienta Vianna (2006, p. 22). Cunha e Cintra (1985), por exemplo, no capítulo sobre figuras de sintaxe, apresentam exemplos análogos a (9) como instância de silepse de pessoa, ressaltando que se trata de característica comum da fala popular do Brasil, África e Europa. Em consonância com isso, no capítulo sobre pronomes, prescrevem que o verbo, com *a gente*, deve ficar na terceira pessoa do singular. Construções do tipo de (8), pelo contrário, são consideradas representativas da norma culta do

português e abonadas por citações de escritores como Almada Negreiros, Otto Lara Resende e Camilo Castelo Branco (CUNHA; CINTRA, 1985, p. 615-616).

Na gramática gerativa, a flexão de 1P do verbo em sentenças com *a gente* na posição de sujeito não constitui fenômeno de natureza estilística, externo, conseqüentemente, à gramática. Em vez disso, é analisada como resultado de um processo regular de concordância com os traços semânticos dessa expressão (COSTA; PEREIRA, 2005), a exemplo de construções análogas em outras línguas (VON HEUSINGER; KAYSER, 2011), ou com os traços sintáticos de um pronome NÓS fonologicamente nulo, núcleo de um DP em que *a gente* funciona como adjunto (TAYLOR, 2009). Nesse quadro, a flexão de 3S do verbo em (1) decorre da concordância com os traços sintáticos de *a gente*.

Dialetalmente, a concordância verbal com *a gente* apresenta um complicador adicional para a modelação computacional, uma vez que a flexão de 1P convive com a de 3S nos dialetos que licenciam a primeira e, no dialeto de São Miguel, nos Açores, essas duas estratégias de concordância alternam também com a flexão de 3P, como exemplificado em (10). A alternância, no dialeto micalense, entre as flexões de 3S, 1P e 3P com *a gente* na posição de sujeito é fato bem estabelecido na dialetologia portuguesa (PEREIRA, 2003) e se pode amplamente comprovar nos dados relativos à localidade de Ponta Garça do *corpus* CORDIAL (TAB. 1).

(10) A/D-F gente/N vem/VB-P-3S aqui/ADV ./, a/D-F gente/N tiram/VB-P-3P .../. [...] A/D-F gente/N ./, primeiro/ADV ./, a/D-F gente/N temos/TR-P-1P uma/D-UM-F urdideira/N donde/WADV a/D-F gente/N levam/VB-P-3P .../. (MIG)

A TAB. 1 sintetiza, a partir de dados do *corpus* CORDIAL, a variação nos dialetos do português europeu (doravante PE), no que tange à concordância verbal com *a gente*. A flexão de 3P praticamente se restringe à Ilha de São Miguel, onde constitui a estratégia dominante. Quanto à flexão de 1P, verificam-se dois extremos, com várias situações intermediárias. Em um extremo, temos o falar de Gião, onde só se verifica a concordância padrão com 3S. Noutro extremo está o dialeto de Moita do Martinho, que, com 24%, apresenta o índice mais alto de 1P com *a gente* no *corpus* CORDIAL.

TABELA 1
Estratégias de concordância verbal no corpus CORDIAL³

Localidade	Sigla	Tokens	3S	1P	3P
Covo (Aveiro)	COV	22.3K	65 (98.5%)	1 (1.5%)	0 (0 %)
Gião (Porto)	GIA	9.8K	71 (100.0%)	0 (0 %)	0 (0 %)
Moita do Martinho (Leiria)	MTM	12K	38 (76%)	12 (24%)	0 (0 %)
Ponta Garça (Ponta Delgada)	MIG	18.7K	120 (39.60%)	10 (3.30%)	173 (57.10)
Todo o corpus		604K	2816 (88.03%)	197 (6.16%)	186 (5.81%)

Finalmente, um *parser* do português, voltado para a análise de sentenças em diferentes dialetos, precisa lidar com a variação na concordância com pronomes de segunda pessoa,⁴ como, por exemplo, a concordância de *vocês* tanto com a 3P quanto com a 2P no dialeto de Covo (PEREIRA, 2003, p. 25), de que damos um exemplo extraído do CORDIAL:

(11) [...] **Olhai/VB-I-2P** lá/FP ./, que/WPRO é/SR-P-3S que/C **vocês/NPR-P andam/VB-P-3P** a/P fazer/VB ?/. !/. **Vocês/NPR-P são/SR-P-3P** uns/D-UM-P tolos/ADJ-P ./ **Vocês/NPR-P andais/VB-P-2P** a/P gastar/VB dinheiro/N [...] (COV)

Neste artigo, partimos do pressuposto de que esses fenômenos de variação na concordância verbal do português, incluindo os casos classificados tradicionalmente como *silepse*, não constituem desvios da “coesão gramatical”, como postulam Cunha e Cintra (1985), mas, sim, uma forma diferente de estabelecê-la. Dada a sua natureza gramatical, são passíveis, portanto, de sistematização por meio de princípios ou regras no âmbito de um determinado modelo formal.

Com isso, pretendemos, inicialmente, prestar uma contribuição para a descrição gerativa do português, uma vez que, pelo que pudemos apurar, inexistem estudos sobre a concordância verbal nesse arcabouço que abarquem a mesma gama de fenômenos nos dialetos que enfocamos. Outra lacuna para cujo preenchimento acreditamos vir a contribuir se refere ao tratamento computacional da variação dialetal e da concordância semântica, dado que os *parsers* que têm sido implementados no âmbito dos principais formalismos gramaticais para línguas

como inglês, alemão e francês, além do próprio português, privilegiam a linguagem padrão, em decorrência do foco em aplicações como tradutores e corretores gramaticais (BUTT *et al.*, 1999).

No contexto do desenvolvimento de um *parser* para um fragmento do português, focaremos as seguintes questões relacionadas à variação na concordância verbal:

i. Como modelar a concordância entre elementos de propriedades gramaticais aparentemente díspares como o pronome *vocês*, formalmente especificado como 3P, e a forma verbal de 2P?

ii. Como implementar computacionalmente a concordância semântica em um *parser* do português capaz de lidar com a variação dialetal dos padrões de concordância verbal?

iii. Como dar conta, de forma elegante, do fato de que, no dialeto mi-caelense, em construções com *a gente* na posição de sujeito, o verbo pode ficar tanto na 3S e na 1P quanto na 3P?

iv. Como gerar, a partir de uma gramática G capaz de analisar os dados dos diferentes dialetos ($D_1, D_2, D_3, \dots, D_n$), gramáticas ($G_1, G_2, G_3, \dots, G_n$) e correspondentes *parsers* ($P_1, P_2, P_3, \dots, P_n$) específicos a cada um desses dialetos, de modo que uma construção gramatical em D_i , mas agramatical em D_j , seja reconhecida por P_i mas não por P_j (para $i \neq j$ e $1 \leq i, j \leq n$)?

Na implementação desse *parser*, utilizaremos o modelo gerativo da Gramática Léxico-Funcional (LFG, do inglês *Lexical-Functional Grammar*) – um dos mais importantes atualmente no âmbito do *parsing* sintático – no ambiente *Xerox Linguistic Environment* (XLE), que representa o estado da arte no desenvolvimento de gramáticas computacionais (CROUCH *et al.*, 2011).

O restante do trabalho estrutura-se em cinco seções. A próxima seção resume as abordagens mais recentes sobre a expressão *a gente* e sobre os seus padrões de concordância, tanto no português europeu quanto brasileiro, a partir de perspectivas sociofuncionalistas e gerativistas. Algumas dessas abordagens exploram, igualmente, as implicações desse elemento para o sistema pronominal e para o próprio mecanismo de concordância em português. Constituem, portanto, excelente ponto de partida para a implementação de um *parser* capaz de lidar com a variação da concordância não só envolvendo *a gente*, mas também outras expressões nominais. Veremos, contudo, que falta às abordagens sociofuncionalistas o rigor conceitual necessário ao tratamento computacional, ao passo que as abor-

dagens gerativas não abarcam todo o espectro de variação que pretendemos modelar. Na seção 3, tratamos da teoria da LFG, focando, principalmente, os níveis de representação do formalismo, o mecanismo de unificação de estruturas de traços e as condições de boa formação, no âmbito dos quais a concordância é modelada. Esses diferentes aspectos são exemplificados por meio de um pequeno *parser* do latim implementado no sistema XLE. Depois de tratar da metodologia de desenvolvimento de gramáticas computacionais na seção 4, mostramos, na seção 5, a relevância de um *parser* do português, baseado na LFG e expomos as principais características de um fragmento de gramática nesse formalismo que desenvolvemos, visando às questões de (i) a (iv) acima sobre a variação da concordância. Nesse fragmento, merece destaque o tratamento do chamado pronome *a gente*, pelo fato de licenciar, no dialeto micaelense, tanto 3S quanto 1P e 3P. Propomos uma análise léxico-funcional desse elemento, no quadro da hipótese DP, que incorpora aspectos tanto dos estudos sobre gramaticalização de orientação sociofuncionalista quanto de abordagens mais recentes no Programa Minimalista. Adotando a análise minimalista de Taylor (2009), postulamos que *a gente* não constitui um núcleo D, tal como os pronomes pessoais *nós*, *vocês* etc., mas uma expressão idiomática com uma estrutura sintática interna de DP e traços de pronome pessoal.

2. O sistema pronominal e a concordância verbal em português

O sistema pronominal e a concordância de pronomes e expressões nominais com os predicados de que funcionam como sujeito constituem fenômenos gramaticais nucleares. Desse modo, as mudanças históricas e as variações sociais ou regionais nesse campo, constatadas no português, têm motivado uma vasta gama de trabalhos de cunho sociolinguístico e / ou funcionalista, dos quais uma parte considerável se debruça sobre os padrões de concordância com *a gente* (VIANNA, 2006), seu percurso de gramaticalização a partir do substantivo *gente* (LOPES, 2003, 2004) ou do sintagma nominal correspondente (ZILLES, 2005), passando a integrar, em concorrência com o pronome *nós*, o sistema pronominal e se cliticizar (MAIA, 2012). Outra linha de pesquisa bastante produtiva trata da alternância entre os pronomes *tu* e *ocê* e a variação na concordância do primeiro no português do Brasil (doravante PB) (LOPES; CAVALCANTE, 2011). Aparentemente restrita ao falar de uma localidade portuguesa, a concordância variável de *vocês* com a 3P e a 2P não parece ter sido alvo de abordagens fora da dialetologia.

Não obstante a produtividade de *a gente* tanto no PE quanto no PB, a alternância com o pronome *nós* aparentemente tem recebido mais atenção nesse último (ZILLES, 2005; MAIA, 2009; FRANCESCHINI, 2009; MENDONÇA, 2012), provavelmente devido tanto à grande difusão da sociolinguística quantitativa no Brasil quanto à alegada contribuição de *a gente* para uma simplificação ainda maior do paradigma flexional verbal dessa variedade (ZILLES, 2004). Fora do Brasil, especialmente em Portugal, ao contrário, parece ocupar o primeiro plano a investigação dos padrões de concordância de *a gente* a partir de uma perspectiva gerativa (PEREIRA, 2003; COSTA; PEREIRA, 2005), o que também se tem feito em relação ao PB (TAYLOR, 2009).⁵

Nesta seção, resumimos, dentre as abordagens mais recentes sobre a expressão *a gente*, aquelas que nos pareceram mais relevantes para a análise apresentada na seção 5. Dado o papel de *a gente* no sistema da língua, como alternativa frequente ao pronome *nós* para a expressão da primeira pessoa do plural, essas abordagens acabam tratando de aspectos mais gerais relacionados ao sistema pronominal como um todo e à concordância em português. Na subseção , focamos trabalhos alinhados à sociolinguística quantitativa em que predomina a visão funcionalista, não obstante eventuais aportes gerativistas. A subseção confronta duas explicações contrárias sobre a ocorrência de expressões nominais semanticamente de primeira pessoa do plural com formas verbais de 3P. Finalmente, na subseção , comparamos dois tipos de abordagens minimalistas sobre os padrões variáveis de concordância com *a gente*: o primeiro tipo pressupõe que essa expressão constitui um pronome pessoal, ao passo que o segundo nega o estatuto pronominal dessa expressão.

2.1. Abordagens variacionistas e funcionalistas

Conforme Lopes (2004, p. 52), no processo de gramaticalização do substantivo *gente* que resultou no pronome *a gente*, preservaram-se algumas das propriedades formais do primeiro, como o traço formal de 3ª pessoa, ao mesmo tempo em que o último não adquiriu todas as propriedades inerentes aos pronomes pessoais. Para descrever as propriedades de *a gente* que o indivíduo no quadro dos pronomes pessoais, essa autora distingue entre traços morfossintáticos e semântico-discursivos. No caso da categoria semântica de pessoa, ela toma como base o traço [EU] para caracterizar o locutor e o alocutário como [+EU] e [-EU], respectivamente, enquanto a chamada terceira pessoa é caracterizada como [ϕ EU],

em que ϕ denota a não atribuição de um valor ao traço. O QUADRO 1 apresenta os traços morfossintáticos (em caixa baixa) e semântico-discursivos (em caixa alta) do pronome *a gente* conforme a abordagem de Lopes (2004), uma reformulação da proposta de Lopes (2003, p. 32). A letra α é uma variável sobre os valores “+” e “-”. Segundo a autora, a concordância de *a gente* com a 1ª pessoa do plural, frequente no português não padrão, bem como sua utilização como antecedente do possessivo *nosso*, constituem indícios sintáticos da especificação semântica [+EU].

QUADRO 1
Traços de *a gente*, segundo Lopes (2004, p. 54)

Tracos	Morfossintáticos	Semântico-discursivos
Pessoa	[Φ eu]	[+EU]
Gênero	[Φ fem]	[α FEM]
Número	[$-p$]	[+PL]

Em sua investigação sobre *nós* e *a gente* na língua falada e escrita por falantes de baixa e média escolaridade (1º e 2º graus) nativos do Rio de Janeiro, Vianna (2006) se concentra, a exemplo do estudo de Pereira (2003) sobre o PE, na concordância desses pronomes com o adjetivo ou particípio em exemplos do tipo de (12). Ela conclui que, nessas estruturas, o predicado, com *a gente*, fica quase sempre no singular, ao passo que, com *nós*, é mais frequente no plural.

- (12) a. A gente está cansados.
 b. A gente está cansada.
 c. A gente está cansadas.
 d. A gente está cansado.

Não obstante o foco nesse tipo de construção, Vianna (2006, p. 46-47) faz um levantamento também das diferentes possibilidades de concordância verbal tanto de *a gente* quanto de *nós*, identificando, ao todo, cinco estratégias, exemplificadas em (13). Nos dois corpora investigados, respectivamente das décadas de 1980 e 2000, das 85 ocorrências de *a gente* analisadas, em 8 se verifica a concordância com a 1P, ao passo que em apenas 3 se constata a flexão de 3P. Quanto

à concordância com o pronome *nós*, observam-se somente 5 casos de 3S, em um total de 92 ocorrências. Comparando os dados obtidos por Vianna nas décadas de 1980 e de 2000, verifica-se um aumento na concordância de *nós* com a 3S, ao mesmo tempo em que diminui a concordância de *a gente* com a 1P e a 3P.

- (13) a. A gente vai ao cinema.
 b. A gente vamos ao cinema.
 c. A gente vão ao cinema.
 d. Nós vai ao cinema.
 e. Nós vamos ao cinema.

Para Vianna (2006, p. 54-55), ocorrências como (13c) constituem exemplos de concordância semântica, pois o pronome *a gente*, durante o processo de gramaticalização, teria conservado a noção semântica de plural do nome *gente*.

Combinando os diferentes padrões de concordância com o verbo, exemplificados em (13), com os padrões de concordância em estruturas predicativas de (12), obtemos um grande número de possibilidades, entre as quais destacamos em (14) e (15) algumas que realmente se verificam nos dados de Vianna (2006).

- (14) a. A gente nunca fomos assaltada, não. (dado 89, M2, 1º grau)
 b. ... a gente tão se sentindo sufocados, né? (dado 50, H4, 2º grau)⁶
- (15) a. A gente ficamos estressado (dado 135, M1, 5ª série)
 b. A gente ficam estressada (dado 492, M2, 5ª série)
 c. A gente viveríamos mais realizados (dado 15, M1, 3º ano)⁷

Uma outra importante contribuição de Vianna (2006) é a sistematização que, seguindo a proposta de Lopes (2003), apresenta para os traços formais e semânticos tanto dos chamados pronomes autênticos quanto das formas gramaticalizadas *a gente* e *você* e dos traços de pessoa das flexões verbais (QUADRO 2 e QUADRO 3).⁸

QUADRO 2

Traços formais e semânticos dos pronomes conforme Vianna (2006)

Pronomes	Pessoa	Número	Gênero
Eu	[+EU] [-eu]	[-PL] [-pl]	[αFEM] [Φfem]
Nós	[+EU] [+eu]	[+PL] [+pl]	[αFEM] [Φfem]
Tu	[-EU] [-eu]	[-PL] [-pl]	[αFEM] [Φfem]
Vós	[-EU] [-eu]	[+PL] [+pl]	[αFEM] [Φfem]
a gente	[+EU] [Φeu]	[+PL] [-pl]	[αFEM] [Φfem]
Você	[-EU] [Φeu]	[-PL] [-pl] ¹	[αFEM] [Φfem]

QUADRO 3

Traços formais e semânticos das flexões verbais conforme Vianna (2006)

Flexão	Pessoa
-o	[+eu]
-s	[-eu]
∅	[Φeu]
-mos	[+eu]
-is	[-eu]
-m	[Φeu]

Em um estudo sobre a variação entre *a gente* e *nós* no dialeto mineiro, Maia (2009) constata que ambos os pronomes deflagram a concordância com o verbo tanto na 3S quanto na 1P, como exemplificado em (16). No entanto, a flexão de 3S é bem mais frequente com *a gente* do que com *nós* (61% contra 38%, respectivamente). No caso da flexão de 1P, porém, a situação se inverte radicalmente, ocorrendo esta muito mais vezes com o pronome *nós* do que com *a gente* (96% contra 3%).

- (16) a. Verdura nós come sim senhora... (D.L., 64, f2, Pb)
 b. Nós começamo só com três. (P.S.,27,f1, BH)
 c. A gente fica amolado. (d.E., 76,f3, Pb)
 d. A gente duramo quase dois meses. (I.A.S., 22, f1, BH)¹⁰

Para Maia (2009, p. 55), formas do tipo de *vão* em exemplos como (17) não parecem constituir instâncias de 3P.

- (17) Nós vão lanchá num lugá lá. (R.A., 20, f1, Pb)¹¹

Trata-se-ia, em vez disso, segundo ela, de redução da forma de 1P *vamos* por meio das operações de (18), redução essa que produz a forma intermediária *vamo* (QUADRO 4), do tipo das que se verificam em (16b) e (16d). No *corpus* investigado por Maia, 100% das formas em *-ão* ocorrem com o pronome *nós*.

- (18) a. Queda da consoante *s* da flexão *-mos*
 b. Queda da vogal final
 c. Ditongação da sílaba final¹²

QUADRO 4

A redução fonética das formas de 1P no dialeto mineiro conforme Maia (2009)

Input	operação	output
<i>Vamos</i>	(18a)	<i>vamo</i>
<i>Vamo</i>	(18b,c)	<i>vão</i>

Com base nos dados de *corpus* relativos à interação entre a morfologia verbal e a realização de *nós* e *a gente*, Maia (2009) propõe, diferentemente de Zilles (2004) e muitos outros, que não foi a introdução de *a gente* no sistema pronominal que favoreceu o empobrecimento da morfologia verbal, mas este que levou àquela, conforme esquematizado em (19) (MAIA, 2009, p. 55).

- (19) *nós* V + *-mos* > *nós* V + *-mo*∅ > *nós* V+ *-ão* > *nós* V+ ∅ > *a gente*
 V + ∅

2.2. Concordância de *a gente* com a terceira pessoa do plural?

Comparando Vianna (2006) com Maia (2009), nos deparamos com duas possibilidades opostas de explicação para exemplos do tipo de (13c): enquanto a primeira autora classifica formas desse tipo como 3P, a última as analisa como formas subjacentes de 1P, de que constituem variantes foneticamente reduzidas (QUADRO 4).

Os dados do *corpus* CORDIAL sugerem fortemente que a hipótese de Maia (2009) a respeito do uso de formas aparentemente de 3P com pronomes de primeira pessoa não é válida para a variedade do português da Ilha de São Miguel, nos Açores, no qual essa estratégia flexional é a preferida quando o sujeito é *a gente*, como se pode constatar na TAB. 1. Em primeiro lugar, não há ocorrência de 3P com o pronome *nós*, diferentemente do que constatou Maia no dialeto mineiro, no qual essas formas somente ocorrem com esse pronome. Em segundo lugar, quando *a gente* concorda com a 1P no falar micaelense, se verificam apenas formas plenas como *chamamos*, *trazíamos* etc., mas não intermediárias como *chamamo*, *trazíamos* etc., resultantes da supressão do segmento final da flexão *-mos*, visto que não se atesta esse processo fonológico, produtivo no PB, no dialeto em questão. Finalmente, ainda que formas como *compram*, *compravam* e *vendem* em (20) possam resultar por meio de (17) a partir das formas correspondentes de 1P, uma forma do tipo de *porem* em (21) evidencia que se trata realmente de concordância de *a gente* com a 3P, como entende Pereira (2003), análise essa que se reflete na etiqueta 3P aplicada a essas formas verbais no *corpus* CORDIAL. De fato, aplicando (17) na forma de 1P correspondente *pormos*, obtemos primeiro *pormo* e, em seguida, a forma *porm*, a partir da qual seria necessário postular a inserção de um segmento *e* para se obter *porem*.

(20) E/CONJ eu/PRO fazia/VB-D-1S doze/NUM tigelas/N-P e/CONJ punha/VB-D-1S ali/ADV no/P+D chão/N e/CONJ fazia/VB-D-1S ./ Agora/ADV **a/D-F gente/N compram/VB-P-3P** ./, já/FP não/NEG é/SR-P-3S .../. Porque/CONJ a/D-F gente/N era/SR-D-3S meadas/N-P que/WPRO **a/D-F gente/N compravam/VB-D-3P** ./ **A/D-F gente/N levavam/VB-D-3P** à/P+D-F dobadoura/N que/CONJ era/SR-D-3S para/P dobar/VB .../. (MIG)

(21) Porque/CONJ se/CONJS **a/D-F gente/N porem/VB-F-3P** sempre/ADV no/P+D mesmo/ADJ tear/N ./, isso/DEM faz/VB-P-3S um/D-UM rolo/N assim/ADV (...) e/CONJ descabeça/VB-P-3S ./ (MIG)

Por outro lado, para derivar, a partir de *somos*, formas atestadas no dialeto carioca do tipo de *são* em (22), precisaríamos, igualmente, estipular um processo fonológico adicional: *som* > *sam* > *são*.

(22) ... fala que **a gente** (“**são**”) metida, **são** orgulhosa... (dado 90, Mulher, 1º grau)¹³

Do mesmo modo, no dialeto do Rio de Janeiro, a ocorrência de 3P apenas com *a gente*, embora represente apenas 3.5% dos dados, contra 9.4% e 87.1% de 1P e 3S, respectivamente, sugere que se trata, realmente, conforme propõe Vianna (2006), de concordância com o traço semântico de plural do substantivo coletivo *gente*. Vêm ao encontro dessa hipótese exemplos como (23), do século 15, em que a expressão nominal *a gente*, muito antes de iniciado o processo de gramaticalização que a levaria a funcionar analogamente ao pronome *nós*, deflagra a flexão de 3P no verbo. Em (24), temos exemplo análogo do século 18. Ambos os exemplos foram extraídos do Corpus Histórico do Português Tycho Brahe (GALVES; FARIA, 2010).

(23) **A gente** começou a se ajuntar a ele, e era tanta, não **cabiam** pelas ruas principais, e **atravessavam** lugares [...], desejando cada um der ser o primeiro, e perguntando uns aos outros, quem matou ao Mestre? (L_002).

(24) Não sabe de quem é o papel, e tudo concorre para que absolutamente se ignore que eu tenho parte em semelhante cousa. Sòmente [*sic*] **a gente de casa são** sabedores, incluindo Haller e Dorat, que são reputados como tais. (A_004).

2.3. Abordagens gerativas

Os dados levantados na literatura evidenciam que a concordância com *a gente* ocorre ora entre os traços morfossintáticos, ora entre os traços referencial-discursivos dos elementos envolvidos. Nesse contexto, exemplos como os de (12 a, c, d) são especialmente interessantes, uma vez que apresentam uma assimetria entre o verbo e o predicativo: enquanto o primeiro concorda com os traços formais de *a gente*, o último o faz com os traços semânticos.¹⁴

Sob uma perspectiva gerativa, as construções com *a gente*, por conta dessa complexidade, representam como que uma “prova de fogo” para verificar

a adequação de um determinado modelo da concordância (COSTA *et al.*, 2001) ou o papel da interface entre os níveis sintático e semântico-discursivo (ou pragmático) numa arquitetura particular de gramática (D’ALESSANDRO, 2008).

No âmbito do Minimalismo, diversas soluções foram propostas para a variação na concordância em construções com *a gente*. Considerando, na esteira de Menuzzi (2000), essa expressão como elemento do sistema pronominal, Costa e Pereira (2005) partem da distinção entre elementos gramaticais e referenciais na estrutura de traços dos pronomes para, por um lado, derivar, em PE, a gramaticalidade de exemplos como (12a,c) e (25a,b) e a agramaticalidade de exemplos como (25c), em que o plural se restringe ao verbo.¹⁵ Por outro lado, eles procuram, igualmente, derivar o fato de que, no PB, *a gente* nunca induz a concordância de plural com o predicativo em estruturas do tipo de (12a,c) e (25a,b), que seriam, portanto, agramaticais nessa variedade.

- (25) a. A gente estamos cansados.
b. A gente estamos cansadas.
c. *A gente estamos cansado.

Nessa abordagem, o predicativo, enquanto *Small Clause*, constitui uma fase forte, no âmbito da qual é deflagrada a concordância com os traços referenciais de *a gente*. Em seguida, após movimento do sujeito para a posição SpecIP, é realizada a concordância com os traços ainda não valorados, ou seja, os traços gramaticais. Eles procuram atribuir as diferenças entre PE e PB, no que tange ao comportamento de *a gente*, aos estatutos diferentes do morfema de número nas duas variedades, que seria um “morfema dissociado” na primeira variedade e um morfema “*singleton*” na última. Isso explicaria tanto a crescente preferência da concordância desse pronome com a 1P no PE quanto a despreferência desse padrão no PB e o não licenciamento do plural do predicado da *Small Clause* nessa variedade.

Detendo-se apenas sobre o tipo exemplificado por (12a), mas levando em conta, também, o comportamento análogo do pronome impessoal italiano *si*, D’Alessandro (2008) propõe uma ampliação do mecanismo de concordância do Programa Minimalista, distinguindo entre *Agree*, que opera sobre os traços phi, e *Concord*, que tem como objeto o que chama de traços sigma, representando informação pragmática codificada sintaticamente. No caso de (12), o predicativo concorda com os traços sigma, enquanto os traços phi determinam a concordância com o verbo.

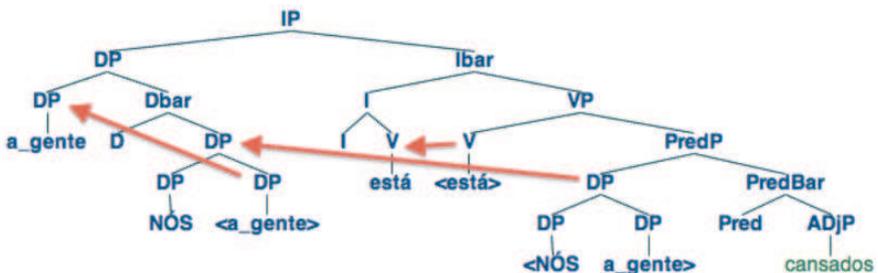
Radicalmente oposta às abordagens de Costa e Pereira (2005) e D'Alessandro (2008) é a abordagem de Taylor (2009) sobre a concordância com *a gente* em dois dialetos do PB de Piri-piri, Piauí: um que ele classifica como padrão e outro, como não padrão. Em primeiro lugar, contesta a análise dessa expressão como pronome, longamente estabelecida tanto na tradição sociofuncionalista quanto gerativa. Em segundo lugar, procura derivar os dados observados não pela distinção entre traços gramaticais e traços semânticos, mas pela postulação de duas estruturas sintáticas para *a gente*. Assumindo a análise dos pronomes pessoais como determinantes, postulada para o PE por Raposo (1998), Taylor postula, para o dialeto não padrão, a seguinte estrutura:

(26) [DP NÓS [DP *a gente*]]

Essa estrutura, onde NÓS é um pronome não realizado foneticamente, reflete a proposta de Den Dikken¹⁶ (2001 *apud* Taylor, 2009, p. 12) para substantivos do inglês como *committee*, chamados de *plurilinguals*, por serem gramaticalmente singulares mas deflagrarem a concordância de plural com o verbo. Na *Small Clause*, em exemplos como (12a, c) ou (25a, b), o predicado concorda com o pronome implícito NÓS. Na posição de SpecIP, porém, o DP de (26) integra, no dialeto padrão, uma estrutura ainda mais complexa. Nessa estrutura, *a gente* se encontra em Spec de um DP com um núcleo D foneticamente vazio e gramaticalmente de terceira pessoa do singular, de que o DP de (26) funciona como complemento (Figura 1). Essa configuração explica, conforme Taylor, a concordância com o verbo na terceira pessoa do singular.

FIGURA 1

Análise de (12a) conforme a abordagem de Taylor (2009, p. 23)



Nesse aspecto, a abordagem de Taylor baseia-se na teoria dos “impositores” de Collins e Postal (2008), DPs como *daddy* no inglês, em exemplos como (27), que são semanticamente de uma pessoa *n*, mas gramaticalmente de uma pessoa *m*, sendo $n \neq m$.

(27) PRO_i to keep myself_i/himself_i from getting sunburned, Daddy_i will put on suntan lotion.

No dialeto não padrão, a concordância com o verbo na primeira pessoa do plural é deflagrada pelo sintagma (26) quando ele se move, despojado da estrutura expandida, para a posição de SpecIP.

Taylor apresenta dados convincentes de que *a gente* não é um pronome, isto é, não constitui um núcleo D, pois, contrariamente aos pronomes de plural *nós* e *vocês*, que constituem núcleos D conforme Raposo (1998), não admite um NP como complemento nem coocorre com numeral (ou seja, também não admite um NumP, projeção máxima da categoria numeral, como complemento), o que explica o contraste entre (28) e (29).

- (28) a. [DP [D nós] [NP brasileiros]]
 b. [DP [D nós] [NumP três]]¹⁷
- (29) a. *[DP [D a gente] [NP brasileiros]]
 b. *[DP [D a gente] [NumP três]]

Para derivar os padrões de concordância, Taylor não recorre, portanto, nem a traços semânticos nem a operações de concordância distintas. Sua abordagem prescinde também da necessidade de explicar por que a concordância com os traços referenciais se restringe à fase forte representada pela *Small Clause*, o que, segundo nos parece, representa um problema na abordagem de Costa e Pereira (2005).

2.4. Conclusão

No contexto do desenvolvimento de um *parser* do português capaz de lidar satisfatoriamente com exemplos da língua não padrão envolvendo *a gente*, as abordagens sociolinguísticas e funcionalistas resumidas na subseção desempenham um importante papel, na medida que fornecem os dados relevantes e pro-

põem algumas explicações para os padrões de concordância observados, apontando caminhos para a sua formalização e modelação computacional.

Comparando os dados do PB dos dialetos carioca e mineiro, levantados por Vianna (2006) e Maia (2009), com os referentes ao PE na TAB. 1, sobressaem as seguintes diferenças entre as duas variedades principais do português, ressaltando-se o fato de que não se dispõe, em termos de *corpora* livremente acessíveis, de um painel tão abrangente do PB quanto o oferecido para o PE pelo CORDIAL:

i. em PB, a concordância de *a gente* com a 1P alcança, quando muito, 10%, ao passo que chega a atingir 24% no PE, o que sugere estar correta a predição de Costa e Pereira (2005) sobre a maior preferência desse padrão em PE do que em PB; e

ii. a concordância de *a gente* com a 3P parece constituir um fenômeno marginal em PB, contrariamente ao seu caráter plenamente produtivo no dialeto do PE da Ilha de São Miguel, em que constitui a estratégia predominante.

Outro aspecto muito relevante da contribuição das abordagens funcionalistas de Lopes (2004) e Vianna (2006) é a formalização inicial proposta para os traços formais e semânticos dos pronomes e das flexões verbais, o que, no caso específico da expressão *a gente*, permite modelar o processo de gramaticalização a partir do substantivo *gente* como uma sequência de alterações na estrutura de traços desse elemento. Como veremos na seção 3, são exatamente as estruturas de traços, formalizadas matematicamente como matrizes de atributos e valores, que constituem o fundamento das principais teorias gramaticais computacionalmente implementadas. No entanto, uma vez que, a essas propostas, não subjaz uma concepção de gramática enquanto formalização matemática de um procedimento recursivo para geração de sentenças (BUSSMANN, 2002, p. 242), elas deixam em aberto importantes questões, que são imprescindíveis para uma implementação computacional, por exemplo:

(30)

i. Como o mecanismo da concordância se insere na gramática enquanto dispositivo gerativo?

ii. Que tipo de estruturas de dados um mecanismo de concordância manipula?

iii. Que tipos de operações matemáticas estão envolvidas no(s) mecanismo(s) de concordância?

iv. Há um mecanismo de concordância semântico formalmente distinto de um mecanismo de concordância morfossintático?

v. Há mecanismos de concordância distintos para a concordância entre sujeito e verbo e a concordância entre o predicativo e o seu sujeito?

Outro problema das abordagens funcionalistas é a não definição rigorosa dos termos, como podemos constatar na passagem abaixo de Lopes (2004, p. 52-53), em que se usa o termo *traço* tanto para designar um valor de um atributo quanto um atributo:

Rooryck [...] defende que uma especificação [α] admite um valor positivo “+” ou negativo “-” para um dado atributo, enquanto um traço \emptyset (*default* ou neutro) encobre um atributo com nenhum valor. Procurando uma representação econômica em termos de sistema de traços, propõe-se, no caso do gênero, a distinção binária, partindo do traço “fem(inino)” como forma marcada [...].

As três abordagens minimalistas resumidas na subseção anterior têm a vantagem de lançar luzes sobre essas questões, embora não o façam com o grau de detalhamento necessário à implementação de um *parser*, dado o próprio caráter semiformal do modelo subjacente (FALK, 2001, p. 65). De um ponto de vista descritivo, contudo, essas abordagens são insatisfatórias, à medida que não permitem derivar todos os padrões de concordância verificados. Isso é mais evidente em relação à proposta de D’Alessandro (2008), restrita ao tipo (12a). Para Costa e Pereira (2005), no PB, em estruturas predicativas, o adjetivo ou particípio nunca vai para o plural quando o sujeito é *a gente*, contrariando os fatos observados por Vianna (2006) no dialeto carioca (ver exemplos (14b) e (15c)).

Os dados (12b, d) são problemáticos para a proposta de Taylor (2009), que não prevê uma assimetria entre os traços de gênero e os de número. Essa dificuldade é reconhecida pelo próprio autor, para quem a configuração da FIG. 1 induz tanto a concordância de gênero quanto de número entre SpecPredP e AdjP.

3. O modelo gerativo da LFG e o sistema XLE

Uma vez que a LFG, infelizmente, ainda não é uma teoria muito difundida no mundo de língua portuguesa, ao contrário do que ocorre nos países de língua inglesa, alemã e francesa, apresentamos, nesta seção, os aspectos principais

desse formalismo tal como implementado no XLE, a fim de que o leitor possa acompanhar a exposição das seções seguintes. Para tanto, recorreremos a uma minigramática computacional do latim, língua de ordem relativamente livre que, por dispor de um sistema rico de casos morfológicas, permite exemplificar diversos recursos avançados da LFG e do XLE.

Atualmente, no *parsing* sintático profundo, além da LFG, outros três formalismos gramaticais se destacam: a *Head-Driven Phrase Structure Grammar* (HPSG), a *Combinatorial Category Grammar* (CCG) e a *Tree-Adjoining Grammar* (TAG) (LJUNGLÖF; WIRÉN, 2010). Todos dispõem de sistemas bastante sofisticados para o desenvolvimento de gramáticas computacionais e têm sido usados na implementação de *parsers* de ampla cobertura para diversas línguas (MÜLLER, 2010).

Nesse quadro, quando consideramos tanto as implicações desses formalismos para a linguística teórica quanto as suas aplicações nas tecnologias da linguagem natural (em áreas como a tradução automática, a resolução de perguntas e a extração de informações), a HPSG e a LFG sobressaem. Uma extensa gramática do PE, a LXGram, desenvolvida no formalismo da HPSG, está disponível (BRANCO; COSTA, 2012). A LFG tem sido usada como base para a elaboração de gramáticas computacionais de mais de duas dezenas de línguas, especialmente no âmbito do Projeto ParGram, voltado para a descrição de línguas tipologicamente diversas como inglês, árabe, malgaxe, indonésio, húngaro, chinês e urdu, entre outras, a partir de um aparato conceitual comum (PARGRAM, 2012). O português, contudo, ainda não dispõe de um *parser* de maior extensão baseado na LFG (MÜLLER, 2010; MISTICA *et al.*, 2012). O presente trabalho representa, portanto, um esforço inicial para preencher essa lacuna.

A exemplo do termo HPSG, o termo LFG é ambíguo, uma vez que se refere tanto a um formalismo gramatical quanto a uma teoria sintática que, como a HPSG, filia-se à linguística gerativa (FALK, 2001; SAG; WASOW; BENDER, 2003). Sob esse último aspecto, a LFG e a HPSG diferem do Programa Minimalista, a atual vertente do gerativismo chomskyano, principalmente por negarem que as transformações sintáticas (por exemplo, as operações de movimento) integrem a faculdade da linguagem (FALK, 2001; BRESNAN, 2001). A relevância da implementação computacional da LFG envolve dois aspectos. De um lado, ela viabiliza a aplicação dos estudos gramaticais em áreas-chave da indústria de *software* como a tradução automática, a recuperação e a extração de informações etc. (BUTT *et al.*, 1999). De outro lado, ela possibilita a construção de modelos do processamento cognitivo da linguagem humana passíveis de serem testados de

forma automática no computador (FALK, 2001). Neste trabalho, concentramos nos aspectos formais da LFG, dado o nosso foco na análise sintática automática.

Enquanto formalismo, a LFG permite definir, por meio de uma gramática, uma *língua formal*, isto é, um conjunto de cadeias (*strings*) sobre um alfabeto Σ , por exemplo, a língua $L = \{\varepsilon, ab, aabb, aaabbb, \dots\}$, formada a partir do alfabeto $\{a, b\}$ tal que $a^n b^n$, com $n \geq 0$. Essa língua é gerada pela gramática livre de contexto (doravante CFG, do inglês *context-free grammar*) de (31).

$$(31) \quad \begin{aligned} S &\rightarrow A S B \\ S &\rightarrow \varepsilon \\ A &\rightarrow a \\ B &\rightarrow b \end{aligned}$$

O formalismo LFG é uma extensão da CFG em que se associam *anotações funcionais* a regras do tipo de (31). Seja, por exemplo, o fragmento de CFG de (32), que, provido das entradas lexicais relevantes, permite gerar, entre outras, as sentenças do latim de (34) e (35). Na regra (32), o licenciamento, no latim, de qualquer permutação entre verbo, sujeito e objeto é indicado por meio do operador de expressões regulares “,”, ao passo que os parênteses indicam a opcionalidade dos sintagmas nominais .

$$(32) \quad \begin{aligned} S &\rightarrow (NP), \\ &\quad V, \\ &\quad (NP). \\ NP &\rightarrow N. \end{aligned}$$

$$(33) \quad \begin{aligned} V &\rightarrow \text{amat} \\ N &\rightarrow \text{poeta} \mid \text{poetam} \mid \text{puella} \mid \text{puellam} \end{aligned}$$

$$(34) \quad \begin{aligned} \text{a.} &\quad \text{Poeta puellam amat.} \\ &\quad \text{Poeta.NOM menina.ACC ama} \\ &\quad \text{‘O poeta ama a menina.’} \\ \text{b.} &\quad \text{Amat puellam poeta.} \\ &\quad \text{‘O poeta ama a menina.’} \end{aligned}$$

$$(35) \quad \begin{aligned} \text{a.} &\quad \text{Poetam amat puella.} \\ &\quad \text{‘A menina ama o poeta.’} \end{aligned}$$

- b. Poetam puella amat.
'A menina ama o poeta.'

Em (36), temos uma versão de (32) no formalismo da LFG, usando a notação do XLE. De passagem, observe-se que nem (32) nem (36) poderiam ser expressas diretamente em um formalismo simples como a implementação da CFG (com ou sem traços) do NLTK (BIRD; KLEIN; LOPER, 2009), biblioteca em Python para o processamento automático da linguagem natural destinada, principalmente, ao ensino da linguística computacional, uma vez que só dispõe do operador de disjunção lógica “|”. A expressiva metalinguagem de expressões regulares do XLE, aliada a vários outros recursos, tornam esse sistema adequado para o desenvolvimento, e utilização em um contexto industrial, de gramáticas de ampla cobertura capazes de operar com léxicos de milhões de entradas sem sacrifício da eficiência computacional (BUTT *et al.*, 1999; CROUCH *et al.*, 2011).

As anotações funcionais acrescentadas, em (36), aos lados direitos das regras de (32) se destacam em **negrito**. Na FIG. 2 estão as entradas lexicais correspondentes às regras livres de contexto de (33).

- (36) S → (NP: (^ SUBJ)=! (! CASE)=**nom**),
V,
(NP: (^ OBJ)=! (! CASE)=**acc**).
NP → N.¹⁸

FIGURA 2

Entradas lexicais do latim na notação do XLE editadas no Aquamacs¹⁹

```
amat      V * (^ PRED)='amare<(^ SUBJ)(^ OBJ)>
          (^ SUBJ NUM)=sg
          (^ SUBJ PERS)=3
          (^ TNS-ASPECT MOOD)=indicative
          (^ TNS-ASPECT TENSE)=pres.

poeta     N * (^ PRED) = 'poeta'
          (^ GEND)=masc
          (^ NUM)=sg
          (^ PERS)=3
          (^ CASE)=nom.

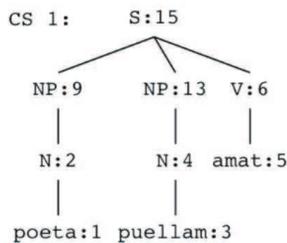
puellam  N * (^ PRED) = 'puella'
          (^ GEND)=fem
          (^ NUM)=sg
          (^ PERS)=3
          (^ CASE)=acc.
```

As anotações “(^ SUBJ)=!” e “(^ OBJ)=!” de (36) especificam que os NPs correspondentes funcionam como sujeito e objeto da sentença. As anotações “(! CASE)=nom” e “(^ OBJ)=! (! CASE)=acc”, por sua vez, impõem restrições sobre os casos do sujeito e do objeto direto, que devem estar no nominativo e no acusativo, respectivamente.

O formalismo da LFG possui uma arquitetura multiestratagemal paralela, que abrange níveis de representação correspondentes às diferentes dimensões da análise linguística, por exemplo, a estrutura de constituintes, a estrutura funcional, a estrutura semântica ou a estrutura fonológica (FALK, 2001). As regras livres de contexto atribuem representações arbóreas às sentenças gramaticais, chamadas na LFG de *c-structures* (estruturas de constituintes ou estruturas sintagmáticas). Na FIG. 3 apresentamos a primeira e única estrutura de constituintes (identificada por CS 1), gerada para a sentença (34a) pelo *parser* construído pelo XLE a partir das regras sintagmáticas de (36) e das entradas lexicais da FIG. 2. Os diferentes nós da árvore são numerados pelo sistema. Essa numeração é utilizada pelo algoritmo de projeção da estrutura de constituintes sobre a estrutura funcional, como veremos a seguir.

FIGURA 3

Estrutura de constituintes de (34a) gerada pelo XLE a partir das regras sintagmáticas de (36) e das entradas lexicais da FIG. 2



As anotações funcionais das regras sintagmáticas e do léxico possibilitam a construção de uma *f-structure* (estrutura funcional) para a sentença. A FIG. 4 exibe a estrutura funcional de (34a) gerada pelo *parser* a partir dessas informações e da estrutura de constituintes da Figura 3.

FIGURA 4

Estrutura funcional de (34a) gerada pelo XLE a partir das regras sintagmáticas de (36) e das entradas lexicais da FIG. 2

"poeta puellam amat"

PRED	'	amare<[1:poeta], [3:puella]>	']
SUBJ	1	[PRED 'poeta']
	2	CASE nom, GEND masc, NUM sg, PERS 3]]
	9	[]
OBJ	3	[PRED 'puella']
	4	CASE acc, GEND fem, NUM sg, PERS 3]]
5	13	[]
6	15	TNS-ASPECT	[MOOD indicative, TENSE pres

Matematicamente, uma estrutura funcional é uma matriz de atributos e valores (AVM, do inglês *attribute-value matrix*), também chamada de estrutura de traços (*feature structure*), conjuntos de pares do tipo da FIG. 5, em que o valor de um traço pode tanto ser atômico (como o valor “fem” do traço “GEND”) quanto constituir uma outra estrutura de traços (KOLB, 2004; BIRD; KLEIN; LOPER, 2009). Estruturas de traços têm, portanto, uma natureza recursiva, pelo que podem representar a estrutura funcional de construções sintáticas de qualquer nível de complexidade.

FIGURA 5

Exemplo de estrutura de traços

TRAÇO1	VALOR1
TRAÇO2	VALOR2
TRAÇO3	VALOR3

A estrutura funcional desempenha um papel fundamental na LFG, uma vez que, no seu âmbito, se definem diferentes condições de boa formação, de que trataremos mais adiante. Esse nível de representação constitui, também, o ponto

de partida para a construção de uma representação semântica da sentença e pode ser usado como interlíngua em um sistema de tradução automática baseado em transferência (BUTT *et al.*, 1999; CROUCH *et al.*, 2011).

Os símbolos “^” e “!” das anotações funcionais são chamados de *meta-variáveis*.²⁰ Na anotação funcional de um nó K, a metavariable “^” se instancia como uma variável que designa a estrutura funcional do nó pai de K, ao passo que “!” se instancia como uma variável que designa a estrutura funcional do próprio nó K. Nós como V e N em (32), que não exibem anotações funcionais, recebem, na verdade, a anotação funcional *default* “^=!”. Desse modo, V funciona como núcleo de S, uma vez que a sua estrutura funcional é identificada à estrutura funcional de S.²¹ Do mesmo modo, N funciona como núcleo do NP. A metavariable “^” possibilita, portanto, que informações do léxico percolem dos núcleos às suas projeções máximas.

A estrutura funcional de uma sentença é construída pelo *parser* a partir da estrutura de constituintes por meio da instanciação das metavariables nas anotações funcionais das regras sintagmáticas e das entradas lexicais. Exemplifiquemos esse processo por meio da análise da sentença (34a) no XLE. Para instanciar as metavariables, os nós da árvore gerada para a sentença pelo *parser* são, como vimos, numerados e esses índices identificam as estruturas funcionais projetadas por cada um dos nós. Por exemplo, o nó 15 projeta a estrutura funcional f15, o nó 6, a estrutura funcional f6 e assim por diante. Inserindo, na árvore, as anotações funcionais das regras sintagmáticas e das entradas lexicais e instanciando as metavariables, obtemos as equações de (37). Na notação da LFG, uma equação como (f9 CASE)=nom equivale à equação f9(CASE)=nom na notação matemática usual, em que f9 é uma função que, aplicada ao argumento CASE, retorna o valor nom.

$$\begin{aligned}
 (37) \quad & (f15 \text{ SUBJ})=f9 \\
 & (f9 \text{ CASE})=\text{nom} \\
 & f15=f6 \\
 & (f15 \text{ OBJ})=f13 \\
 & (f13 \text{ CASE})=\text{acc} \\
 & f5=f6 \\
 & f2=f1 \\
 & f4=f3 \\
 & f2=f9 \\
 & f4=f13
 \end{aligned}$$

As equações (f15 SUBJ)=f9 e (f9 CASE)=nom se parafraseiam informalmente como “o sujeito de f15 é igual a f9” e “o caso de f9 é nominativo”, ou seja, o sujeito da sentença é a estrutura funcional do sintagma nominal que tem caso nominativo. Formalmente, porém, não temos, nessas equações, uma relação de igualdade em sentido estrito. O símbolo “=” nessas equações expressa a operação matemática de unificação. Da resolução das equações de (37) resulta, via unificação, a estrutura funcional da FIG. 4.

Para explicar a operação de unificação, consideremos um exemplo bastante simples. Seja o NP em latim:

- (38) ille poeta
‘aquele poeta’

Para gerar esse sintagma, alteramos a regra do NP de (36) para (39), introduzindo um determinante facultativo. No léxico, incluímos demonstrativos como (40), que, na notação do Projeto ParGram, são caracterizados pelo valor “demon” atribuído ao traço DET-TYPE (KING, 2004). Esse traço constitui, tal como o traço DEIXIS, valor do traço DET (determinante), que, por sua vez, é valor do traço SPEC (especificador).

- (39) NP \rightarrow (D) N.
(40) ille D * (^ GEND)=masc
 (^ NUM)=sg
 (^ CASE)=nom
 (^ SPEC DET DET-TYPE)=demon
 (^ SPEC DET DEIXIS)=distal.

Com essas modificações na nossa minigramática do latim, o *parser* gera, para (38), a árvore da FIG. 6 e a estrutura funcional da Figura 7.

FIGURA 6
Estrutura de constituintes do exemplo (38) gerada pelo XLE

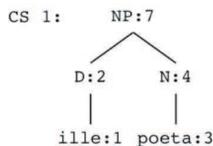


FIGURA 7
Estrutura funcional do exemplo (38) gerada pelo XLE

```
"ille poeta"
3[PRED 'poeta'
4 1[ SPEC [DET [DEIXIS distal, DET-TYPE demon]]
2 2[ CASE nom, GEND masc, NUM sg, PERS 3
7 ] ] ]
```

Com base na árvore da FIG. 6, obtemos, a partir de (39) e (40), as equações de (41), de que resulta $f2=f4$, o que, em termos informais, podemos parafrasear como “a estrutura funcional do determinante é igual à estrutura funcional do substantivo”. No entanto, é fácil verificar que não se tem, nesse caso, a rigor, uma igualdade entre as duas estruturas, como mostram a FIG. 8 e a FIG. 9.

- (41) $f1=f2$
 $f2=f7$
 $f3=f4$
 $f4=f7$

FIGURA 8
Estrutura funcional do nó 4 da árvore da FIG. 6

```
3[PRED 'poeta'
4 4[ CASE nom, GEND masc, NUM sg, PERS 3 ] ]
```

FIGURA 9
Estrutura funcional do nó 2 da árvore da FIG. 6

```
"ille poeta"
3[PRED 'poeta'
4 1[ SPEC [DET [DEIXIS distal, DET-TYPE demon]]
2 2[ CASE nom, GEND masc, NUM sg, PERS 3
7 ] ] ]
```

Na LFG, em equações como (41), o operador binário “=” pode ser interpretado como uma instrução ao *parser* para unificar os dois operandos. Da unificação de duas estruturas funcionais *f1* e *f2* resulta uma estrutura funcional *f3* com todos os traços de *f1* e *f2*, desde que os valores desses traços não sejam contraditórios (BIRD; KLEIN; LOPER, 2009; FRANCEZ; WINTNER, 2012). Por exemplo, os valores dos traços CASE, GEND e NUM nas estruturas funcionais da FIG. 8 e da FIG. 9 são compatíveis, pelo que são incorporados no resultado da unificação dessas duas estruturas na FIG. 7. Do mesmo, a estrutura funcional resultante da unificação incorpora o traço PRED, que só consta na FIG. 8, e o traço SPEC, existente apenas na FIG. 9.

Seja agora o seguinte sintagma, que é agramatical por violar a concordância nominal, uma vez que o valor do traço de gênero do demonstrativo *illa* é feminino, enquanto o valor do traço de gênero do substantivo latino *poeta* é masculino:

- (42) *illa poeta
aquela poeta.MASC

Nesse caso, as especificações de traço de gênero do demonstrativo e do substantivo, respectivamente [GEND fem] e [GEND masc], não unificam, pois os valores do traço GEND são diferentes. O *parser* constrói uma árvore para o sintagma, mas, como essa árvore não gera uma estrutura funcional bem formada, é declarada inválida (FIG. 10). Na FIG. 11 vemos que a agramaticalidade desse sintagma resulta de inconsistência entre os valores do traço GEND.

FIGURA 10

Árvore inválida gerada pelo XLE para o sintagma agramatical (42)

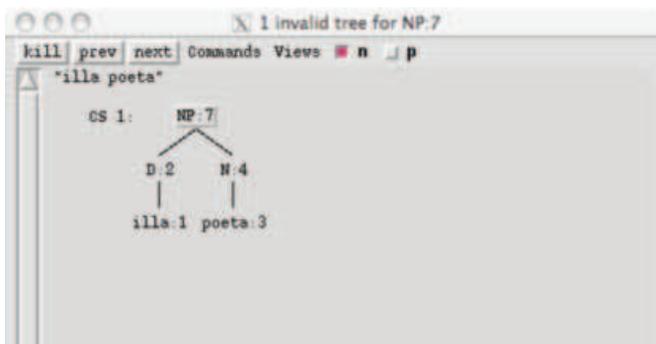
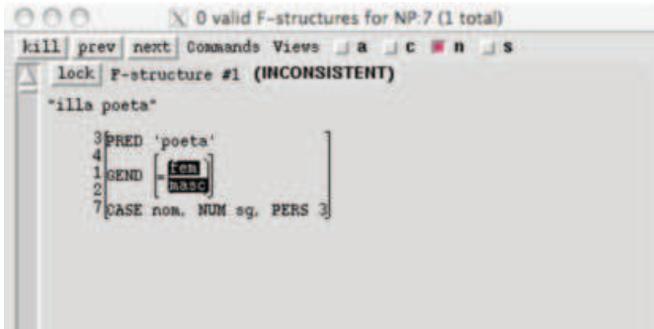


FIGURA 11
Estrutura funcional inconsistente gerada pelo XLE
para o sintagma agramatical (42)



A consistência (ou unicidade) é, portanto, uma das condições de boa formação postuladas pela LFG para as estruturas funcionais (FALK, 2001, p. 64). As outras duas condições são a completude e a coerência. Essas duas últimas condições explicam a agramaticalidade, respectivamente, de (43) e (44), conforme um *parser* construído pelo XLE a partir do nosso fragmento de gramática do latim.

- (43) *Poeta amat.
o poeta.NOM ama
- (44) *Puella poetam saltat.
a menina.NOM o poeta.ACC dança

Como evidencia a FIG. 2, a entrada lexical do verbo *amar* não contempla o uso intransitivo desse verbo, que é, na verdade, também licenciado em latim (FARIA, 1982). Por outro lado, a minigramática contempla apenas o uso intransitivo do verbo *dançar*, que, na acepção de “representar”, admite, igualmente, um uso transitivo (FARIA, 1982). Naturalmente, o XLE admite a especificação de múltiplas valências para um determinado verbo e permite modelar regras lexicais que derivam as diferentes diáteses verbais (BUTT *et al.*, 1999). Nosso objetivo aqui, no entanto, é explicar como informações que constam no léxico de uma gramática no formalismo da LFG são utilizadas pelo *parser* para determinar a agramaticalidade das sentenças.

FIGURA 12
Estrutura funcional do verbo do exemplo (43) gerada pelo XLE

```

< F-structure #1 (INCOMPLETE)
[PRED      'amare<[3-SUBJ], [3-OBJ]>' ]
[SUBJ      [CASE nom, NUM sg, PERS 3] ]
[OBJ       [CASE      acc ]
           [HAS_PRED  OBJ ] ]
3
4 [INS-ASPECT [MOOD indicative, TENSE pres] ]

```

A condição da completude exige que todas as funções argumentais de um predicado devem constar da estrutura funcional desse predicado (FALK, 2001, p. 63), o que não ocorre no exemplo (43), em que falta o objeto direto (FIG. 12). A condição da coerência determina que todas as funções argumentais de uma estrutura funcional devem ser licenciadas pelo respectivo predicado (FALK, 2001, p. 63), pelo que (44) é agramatical, uma vez que o objeto direto [_{NP} poetam] não consta da valência do verbo *saltar*, tal como especificada na respectiva entrada lexical. Nesse caso, o XLE também gera uma estrutura funcional em que o problema da incoerência é apontado.

Não há, portanto, um mecanismo na LFG específico para a modelação da concordância, uma vez que ela é tratada no âmbito da unificação, que, por sua vez, resulta da composição das estruturas de traços dos itens lexicais, concatenados por meio das regras de estrutura sintagmática. Essa composição deve obedecer à condição da consistência, que, ao lado da completude e da coerência, constitui uma das três condições de boa formação das estruturas funcionais. Devido ao caráter recursivo das estruturas de traços, a concordância pode ocorrer tanto local quanto não localmente, nos contextos de dependência de longa distância do tipo de (5a) e (5b).

Outro aspecto extremamente relevante do tratamento da concordância por meio da unificação é que um termo A pode concordar com um termo B em relação a um traço C sem que A (ou B) possua o traço C. Desse modo, a distinção de Lopes (2003, 2004) e Vianna (2006) entre os valores Φ e α de um traço não parece relevante computacionalmente, pelo menos nos formalismos de unificação como a LFG. Por exemplo, uma entrada lexical do pronome *eu* na notação do XLE, adotando os traços propostos por Vianna (2006) (QUADRO 2), poderia ser reformulada como (45), sem a inclusão das especificações de traço de gênero [α FEM] e [Φ fem].

- (45) eu D * (^ EU)= +
 (^ eu)= +
 (^ PL)= –
 (^ pl)= –.

Pressupondo a entrada (46) para o adjetivo *cansado*, obtemos, na estrutura funcional de uma sentença como (47), pela unificação da estrutura funcional do sujeito com a do predicativo, a especificação do traço de gênero semântico do sujeito como masculino.

- (46) cansadoA * (^ PRED)= 'cansado<(^ SUBJ)>'
 (^ PL)= –
 (^ pl)= –
 (^ FEM)= –
 (^ fem)= –.

- (47) Eu estou cansado.

Na seção 5, veremos que a entrada de (46) é insatisfatória por uma série de razões e proporemos uma alternativa que sana essas deficiências.

4. Metodologia

A técnica de *design* de algoritmo conhecida como desenvolvimento em espiral consiste em implementar um protótipo simples capaz de lidar com uma parte do problema a ser resolvido e, em seguida, em sucessivos ciclos de implementação e testagem, expandir progressivamente a sua cobertura, de modo a lidar com todo o problema (ZELLE, 2004, p. 286).

De forma análoga, a complexidade da sintaxe de uma língua natural impõe a definição de um fragmento dessa língua como objeto de uma formalização inicial e a ampliação da cobertura dessa formalização por meio de uma sucessão de fragmentos cada vez mais amplos (FRANCEZ; WINTNER, 2012, p. 6). Adotando essa técnica, o ponto de partida para o desenvolvimento de uma gramática no sistema XLE é a elaboração de uma lista inicial de construções que devem ser geradas pela gramática, abrangendo um leque bem definido de fenômenos. Implementam-se, então, com base em sucessivos fragmentos de gramática,

parsers visando à análise dessas construções. Uma vez alcançado o resultado esperado, amplia-se o leque de construções a serem analisadas, introduzem-se as correspondentes alterações na gramática e constrói-se um novo *parser*. Caso esse *parser* consiga analisar essas novas construções, repete-se o ciclo constituído das três etapas: (i) ampliação da lista de construções, (ii) alteração da gramática e (iii) testagem do *parser* correspondente.

A geração das construções de um determinado fragmento não deve ser o único objetivo de uma gramática. É preciso, igualmente, modelar as restrições que permitem caracterizar determinadas sequências de palavras como agramaticais. Desse modo, para modelar a variação da concordância em um grupo de dialetos do português, construímos um arquivo de teste, nomeado *testfile.lfg*, conforme as convenções do XLE, contendo dois blocos de construções: um bloco de construções gramaticais em pelo menos um dos dialetos e outro, de construções agramaticais em todos os dialetos.²² Excertos dos dois blocos do arquivo de teste que utilizamos na fase final de desenvolvimento de nosso *parser* estão em (48) e (49). As linhas precedidas de “#” constituem comentários e são ignoradas pelo *parser*.

(48) Exemplos do bloco de construções gramaticais

```
#=====
# sentenças gramaticais
#=====
# dialeto padrão
# sentenças com um verbo de ligação e predicativo do sujeito
eu estou cansado
eu estou cansada
# sentenças com um verbo intransitivo
os cavaleiros esperam
# exemplos com quantificador universal
todos os cavaleiros esperam
todos os cavaleiros esperamos
```

(49) Exemplos do bloco de construções agramaticais

```
#=====
# sentenças agramaticais
#=====
```

```
# exemplos com numeral
a gente três espera
a gente três esperamos
a gente três esperam
```

O XLE oferece uma interface bastante amigável na linguagem de *scripts* Tcl. Pressupondo que uma gramática do português foi salva com o nome `basic-port.lfg` e o conjunto de sentenças para testagem da gramática está no arquivo `testfile.lfg`, podemos construir um *parser* a partir dessa gramática e aplicá-lo ao arquivo de teste por meio dos dois simples comandos de (50) e (51), respectivamente. Nesses dois exemplos, os comandos estão destacados em negrito e os respectivos *output* são apresentados de forma resumida.

```
(50) % create-parser basic-port.lfg
loading /Users/user/gramaticas/basic-port.lfg...
0.00 CPU seconds
/Users/user/gramaticas/basic-port.lfg loaded
Grammar last modified on Nov 20, 2012 15:32.

(51) % parse-testfile testfile.lfg
parsing sentences 1 to end in testfile.lfg...
((1) (1 0.00 10) (3 words))
((2) (1 0.02 10) (3 words))
[...]
((95) (0 0.00 14) (4 words))
done parsing testfile.lfg
Grammar = /Users/user/gramaticas/basic-port.lfg.
Grammar last modified on Nov 20, 2012 15:32.
95 sentences, 0 errors, 0 mismatches
50 sentences had 0 parses
0.35 CPU secs total, 0.02 CPU secs max
```

Como podemos constatar no output dos comandos de (50) e (51), o XLE é extremamente rápido. O *parser* é compilado em menos de um centésimo de segundo, ao passo que as 95 sentenças do arquivo de teste são analisadas em 0.35 segundo.

Para cada construção analisada, o XLE apresenta cinco informações diferentes, conforme se exemplifica para a primeira sentença de (48).

QUADRO 5
 Informações fornecidas pelo XLE para a primeira sentença de (48)

1	1	0,00	10	3 words
número de ordem da sentença no conjunto de teste	quantidade de análises	tempo de processamento em "CPU seconds"	quantidade de subárvores processadas	quantidade de palavras

Ao analisar um arquivo de teste, o XLE automaticamente gera um outro arquivo com o mesmo nome e extensão "new", em que acrescenta o resultado da análise de cada sentença, no formato explicado no Quadro 5.

(52) Excerto inicial do arquivo testfile.lfg.new

```
#=====
# sentenças gramaticais
#=====
# dialeto padrão
# sentenças com um verbo de ligação e predicativo do sujeito
eu estou cansado (1 0.01 10)
eu estou cansada (1 0.00 10)
```

A facilidade de uso do XLE se torna ainda maior, por um lado, pela possibilidade de automatizar a compilação e testagem de *parsers* por meio de *scripts* em Tcl, por outro, pela possibilidade de utilizar a interface gráfica do editor de texto *emacs* para executar os principais comandos desse sistema, o que é especialmente interessante para linguistas sem conhecimentos de programação.²³ Por exemplo, no *emacs*, para visualizar a estrutura de constituintes e a estrutura funcional de uma sentença em um arquivo de teste, basta posicionar o cursor ao lado da sentença e digitar a sequência de teclas *escape*, *control* e *p*. O sistema gera, então, janelas como a da FIG. 11, que não apenas exibem gráficos mas constituem também interfaces gráficas (implementadas na linguagem Tcl/Tk) por meio das quais se podem inspecionar as estruturas funcionais dos nós individuais de uma árvore, salvar os gráficos em arquivos de imagem etc.

5. Um fragmento de gramática computacional do português no XLE

Nesta seção, apresentamos, inicialmente, a motivação para o desenvolvimento de uma gramática do português no modelo da LFG. Em seguida, detalhamos a implementação de um fragmento de gramática nesse formalismo capaz de lidar com padrões variáveis de concordância verbal, envolvendo tanto a concordância semântica quanto diferenças entre diversos dialetos do português. No leque de fenômenos modelados, destacam-se os seguintes:

concordância de *a gente* com 3S, 1P e 3P;
casos tradicionalmente considerados silepse de pessoa; e
concordância de *vocês* com 2P e 3P e de *tu* com 2S e 3S.

Graças ao sistema XLE, os diferentes aspectos do fragmento são exemplificados por meio de estruturas de constituintes e estruturas funcionais geradas automaticamente pelo correspondente *parser* para construções extraídas do arquivo de teste.

Diferentemente da LXGram, implementada com base na HPSG pelo Grupo de Fala e Linguagem Natural (NLX), do Departamento de Informática da Universidade de Lisboa, o Curupira, o principal *parser* de ampla cobertura do português desenvolvido no Brasil, não se fundamentou na gramática gerativa, mas na gramática tradicional. Antes que esse *parser*, voltado para análise da variedade padrão do PB, pudesse ter aumentadas as suas taxas de cobertura e precisão, consideradas insatisfatórias, o projeto foi descontinuado em 2006 pela instituição responsável, o Núcleo Interinstitucional de Linguística Computacional (NILC) (MARTINS; OTHERO, 2012).

Independentemente disso, porém, havia, a nosso ver, uma falha no desenho arquitetural do Curupira, que revela o descompasso, ainda hoje vigente no Brasil, entre as pesquisas de gramática gerativa, tipicamente direcionadas para a descrição do português e a teorização linguística, e as de linguística computacional, desenvolvidas principalmente não por linguistas, mas informatas e engenheiros, sem embasamento nas modernas teorias gramaticais, diferentemente da realidade do *parsing* sintático nos EUA e na Europa. Para citar apenas um exemplo, inicialmente a Xerox e, depois, a Microsoft adotaram o modelo gerativo da LFG para a elaboração de diversas gramáticas computacionais no âmbito do Projeto ParGram e, para tanto, investiram no desenvolvimento do sistema XLE, base para a implementação

de vários sistemas de processamento automático da linguagem natural (PARGRAM, 2012).

Vejamos um exemplo concreto de uma deficiência do Curupira decorrente dessa opção de *design*. As análises gerativas do chamado pronome *a gente* no âmbito da teoria da ligação estabelecem que essa expressão, ao contrário do DP [o povo] em (54), não pode constituir antecedente de um pronome possessivo de terceira pessoa do singular (MENUZZI, 2000). Essa informação sintática, essencial para o processamento semântico da sentença, integra as diferentes teorias gerativas, inclusive aquelas mais vinculadas ao processamento computacional da gramática, como a LFG e a HPSG (FALK, 2001; SAG; WASOW; BENDER, 2003).

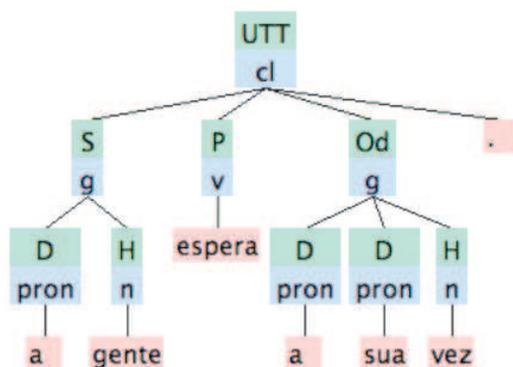
(53) [A gente]_i; espera a [sua]_j; vez. ($i \neq j$)

(54) [O povo]_i; espera a [sua]_i; vez.

A gramática tradicional, contudo, não possui um componente que trate diretamente da determinação sintática das relações de correferência. Esse tipo de informação, portanto, não é codificado pelo Curupira, cuja análise de (53) apresenta uma série de problemas, não obstante se tratar de uma sentença bastante trivial (FIG. 13). Em primeiro lugar, o *parser* não reconhece a interpretação mais usual de *a gente* como forma de expressão da primeira pessoal do plural. Em segundo lugar, o sintagma [a sua vez] foi analisado não como complemento do verbo, mas, erroneamente, como adjunto adverbial oracional, realizado por sintagma preposicional encabeçado pela preposição *a*. Finalmente, o Curupira não explicita que *a gente*, diferentemente de [o povo] em (54), é inviável como antecedente do possessivo.

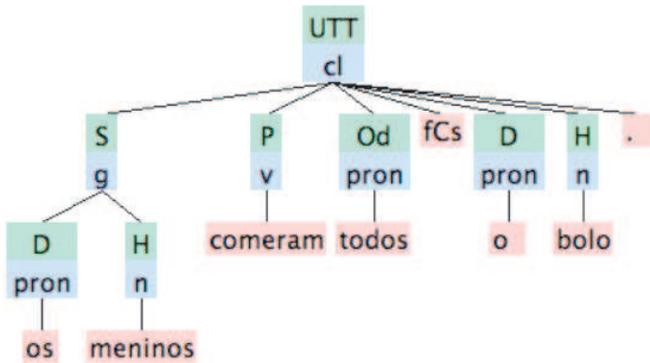
Aparentemente, o NILC cessou o investimento no Curupira porque, nos projetos que necessitam de *parsing* sintático, como o analisador discursivo DiZer (MAZIERO; PARDO; NUNES, 2007), passou a utilizar o *parser* PALAVRAS. Atualmente, o PALAVRAS integra o analisador sintático automático do português do Projeto VISL, desenvolvido pela Universidade do Sul da Dinamarca. Embora muito robusto, esse analisador também se ressentia da falta de entrosamento com as pesquisas descritivas do português de orientação gerativa. De fato, com exceção da segunda deficiência, os problemas apontados na análise do Curupira reproduzem-se na análise realizada pelo *parser* do Projeto VISL (FIG. 14).

FIGURA 14
Análise de (53), pelo *parser* do Projeto VISL²⁵



Outro problema do *parser* do Projeto VISL é que não parece modelar a concordância. Por exemplo, para a sentença agramatical (4), gera árvore em que o sintagma [os brasileiros] funciona como sujeito do predicador realizado pela forma verbal de 1S, sem nenhuma indicação do conflito entre as propriedades gramaticais desses elementos. Dessa forma, esse *parser* é incapaz de reconhecer que, em (5a), o quantificador *todos* tem escopo sobre o sujeito da sentença. Em vez disso, o quantificador, e não o sintagma [o bolo], é analisado como objeto direto do verbo (FIG. 15). Nessa sentença relativamente trivial para a gramática gerativa, uma vez que exemplifica o muito estudado fenômeno da flutuação de quantificadores, o *parser* ainda comete o erro adicional de desmembrar o sintagma *o bolo* em dois constituintes imediatos da oração.

FIGURA 15
Análise de (5a), pelo *parser* do Projeto VISL²⁶



Não obstante os problemas apontados, é inegável a utilidade do analisador VISL para aplicações como a anotação de *corpora*, uma vez que uma análise parcialmente correta é melhor do que análise nenhuma. Segundo parece, um dos preços que esse *parser* paga por essa robustez é não incluir no modelo subjacente nenhuma restrição relativa à concordância, que, como mostramos na introdução, é um fenômeno complexo, sobretudo quando se leva em conta a variação dialetal. Esse *parser* não enfrenta problemas com exemplos de concordância semântica do tipo de (8b) ou de língua não padrão do tipo de (9) porque admite a combinação de qualquer sujeito pré-verbal com qualquer verbo finito, como em (4).

Independentemente das qualidades e defeitos do analisador VISL e de outros *parsers* do português implementados noutros países, consideramos imprescindível, para o desenvolvimento da área de PALN no Brasil e a própria indústria brasileira de *software*, a formação de estudantes e pesquisadores tanto de linguística quanto de ciência da computação que dominem a tecnologia de *parsing*, o que, acreditamos, somente ser possível por meio do envolvimento com a construção de um *parser* de ampla cobertura. Por outro lado, há que ressaltar a importância da familiarização com diferentes abordagens e formalismos. Desse modo, dada a existência de uma ampla gramática do português no formalismo da HPSG, decidimos fundamentar o nosso fragmento de gramática do português na teoria da LFG e implementar o *parser* correspondente no sistema XLE, que representa o estado da arte no desenvolvimento de gramáticas computacionais. Este

trabalho representa, portanto, um passo na direção de uma ampla gramática do português no formalismo da LFG.

Como a literatura da LFG sobre o português é escassa e, pelo que pudemos apurar, não existem análises, nesse arcabouço teórico, dos padrões variáveis de concordância em português (MISTICA *et al.*, 2012), tomamos como ponto de partida abordagens sobre o chamado pronome *a gente* no quadro do Programa Minimalista, as quais apontam caminhos a seguir na implementação de nosso *parser*. Destacamos três aspectos dessas análises:

i. no âmbito da hipótese DP, se pressupomos que pronomes pessoais como *nós* e *vocês* são núcleos D que podem tomar um NP ou um NumP como complemento, e que a expressão *a gente* concorda com verbos e outros elementos no plural, então nos parece forçoso concordar com Taylor (2009) que essa expressão não seja um núcleo D, dada o contraste entre (28) e (29);

ii. uma vez que a LFG não admite operações de movimento, a que as análises minimalistas recorrem para explicar os padrões de concordância com *a gente*, faz-se necessário reconstruir o efeito dessas transformações por meio da unificação de estruturas traços; e

iii. a distinção entre traços semânticos (σ) e sintáticos (ϕ) é facilmente implementável na LFG. No entanto, não há necessidade de postular, com D'Alessandro (2008), duas operações distintas para a concordância com cada tipo de traço, uma vez que se podem derivar os dados relevantes por meio de um único mecanismo, que é a unificação de estruturas funcionais.

No que tange às análises funcionalistas, não só adotamos a sugestão de Lopes (2003, 2004) e Vianna (2006) de distinguir entre traços semânticos e sintáticos no âmbito de todo o quadro pronominal, como também a estendemos às flexões verbais.

A seguir, apresentamos os aspectos principais de nossa minigramática do português no formalismo da LFG, referida doravante por *basic-port.lfg*, o nome do respectivo arquivo-fonte dado como argumento do comando *create-parser* do XLE (ver (50)).

Pela técnica de desenvolvimento em espiral, essa gramática resultou de sucessivas repetições do seguinte ciclo: (i) (re)elaboração da gramática, (ii) aplicação do *parser* ao arquivo de teste, (iii) ampliação do arquivo de teste. No final desse processo, conseguimos que a gramática gerasse os padrões de concordância dos sete dialetos do QUADRO 6, excluindo, ao mesmo tempo, padrões agramaticais em todos esses dialetos.

QUADRO 6
Padrões de concordância analisados

Padrões de concordância	Dialectos						
	RJM	POA	COV	MIG	GIA	LER	PP
Tu vai	+	+	-	-	-	-	-
Tu vais	-	+	+	+	+	+	+
A gente vai	+	+	+	+	+	+	+
A gente vamos	+	-	+	+	-	+	-
A gente vão	-	-	-	+	-	-	-
Vocês vão	+	+	+	+	+	+	+
Vocês ides	-	-	+	-	-	-	-

No QUADRO 6, distinguimos, além do português padrão (PP), dois dialetos não padrão do PB (RJM e POA) e quatro dialetos não padrão do PE (COV, MIG, GIA e LER). No que tange à concordância com *a gente*, RJM, o primeiro dialeto do PB, corresponde, *grosso modo*, aos dialetos carioca e mineiro descritos, respectivamente, por Vianna (2006) e Maia (2009), com a diferença de que a flexão de 3P não é considerada, dado o seu caráter marginal nesses dialetos. Quanto à realização da segunda pessoa do singular, RJM instancia o que Lopes e Cavalcante (2011, p. 38), em sua classificação da variação dialetal dos pronomes *tu* e *você*, denominam subsistema 5, verificado no Rio de Janeiro e em parte de Minas Gerais, no qual *você* e *tu* alternam e, com esse último pronome, o verbo sempre exibe a flexão de 3S. Em POA,²⁷ o segundo dialeto do PB, as flexões de 2S e de 3S alternam-se com o pronome *tu*, analogamente aos subsistemas 2 e 3 da referida classificação, localizados em partes do Sul, Norte e Nordeste (LOPES; CAVALCANTE, 2011, p. 38). No dialeto POA, *a gente* não ocorre com a 1P, como relatado por Zilles (2005) para a cidade de Porto Alegre.²⁸

É evidente que o QUADRO 6 não abarca toda a variação da concordância no PB. Fenômenos como a concordância de DPs no plural com a 3S, bastante difundida nessa variedade, como relatado por Zilles (2005), Vianna (2006), Maia (2009) etc., serão contemplados apenas em uma etapa posterior de desenvolvimento da gramática.

Os dialetos COV, MIG e GIA do QUADRO 6 têm como modelos os falares de Covo (Aveiro), São Miguel (Açores) e Gião (Porto), respectivamente. O

dialecto LER é baseado no falar de Moita do Martinho (Leiria) e outros dialetos do PE onde a concordância de *a gente* com a 1P é bastante produtiva, não obstante a predominância da 3S.

Para as expressões nominais do português, seguindo em linhas gerais a abordagem de Othero (2009), que, por sua vez, reflete análises no âmbito da hipótese DP (BERNSTEIN, 2003), propomos a seguinte regra:

$$(55) \quad DP \longrightarrow \{ (QP) \text{ Dbar} \mid QP: \wedge \neq! (\wedge \text{ PRED}) = \text{'PRO'} \}.$$

Conforme (55), há duas possibilidades de realização de um DP em português, exemplificadas em (56): (i) como um Dbar (isto é, um D') com um sintagma quantificador (QP) facultativo na posição de especificador ou (ii) como um sintagma quantificador "nu". Na análise de (56), divergimos de Othero (2009, p. 76), que considera sintagmas com quantificadores do tipo de [todos os alunos] como projeções máximas do núcleo funcional Q, uma vez que essa análise resulta, no âmbito da própria CFG, utilizada por Othero, bem como da LFG, que é uma extensão desse formalismo, em uma gramática computacional menos elegante. De fato, em todas as regras em que aparece um DP, seria preciso construir uma disjunção entre DP e QP, porque esses sintagmas têm praticamente a mesma distribuição. As exceções a isso incluem contextos em que opera a chamada restrição do pronome pleno (OTHERO, 2009, p. 77). Na LFG, porém, esse tipo de restrição pode ser reconstruído por meio de anotações funcionais.

- (56) a. [DP [QP Todos]] esperam.
 b. [DP [QP Todas]] esperamos.
 c. [DP [QP Todos] [Dbar nós]] esperamos.

Na LFG, todo sintagma que funciona como argumento de algum predicador (tipicamente um verbo) e realiza uma função sintática regida por esse predicador (SUBJ, OBJ etc.) deve possuir um atributo PRED em sua estrutura funcional. Desse modo, é necessário incluir, no segundo membro da disjunção de (55), a anotação funcional " $(\wedge \text{ PRED}) = \text{'PRO'}$ ", em que se atribui a PRED o valor 'PRO', como se convencionou representar semanticamente os pronomes na LFG (BUTT *et al.*, 1999).

A projeção intermediária de D é gerada pela regra (57), que postula duas alternativas: (i) núcleo D facultativo, seguido de um NumP ou NP como complemento ou (ii) núcleo D "nu".²⁹ Essas duas possibilidades estão exemplificadas em (58).

- (57) $Dbar \rightarrow \{ (D) \{NumP \mid NP\} \mid D: \wedge \neq! (\wedge PRED) = 'PRO' \}$.
 (58) a. Os três cavaleiros esperam.
 b. Três cavaleiros esperam.
 c. Eles esperam.

As projeções NumP e QP são geradas, respectivamente, por (59) e (60). Na primeira regra, contemplamos a possibilidade de NumP se realizar como um numeral “nu”, em exemplos do tipo de (61), cuja representação arbórea se encontra na Figura 16.

- (59) $NumP \rightarrow \{ Num \ NP \mid Num: \wedge \neq! (\wedge PRED) = 'PRO' \}$.
 (60) $QP \rightarrow Q$.
 (61) Três esperam.

FIGURA 16
 Estrutura de constituintes de (61) produzida pelo XLE a partir da gramática basic-port.lfg

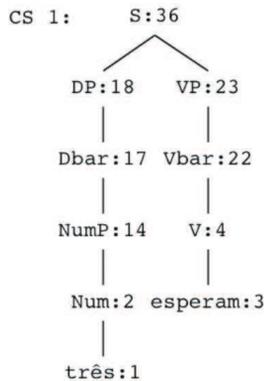
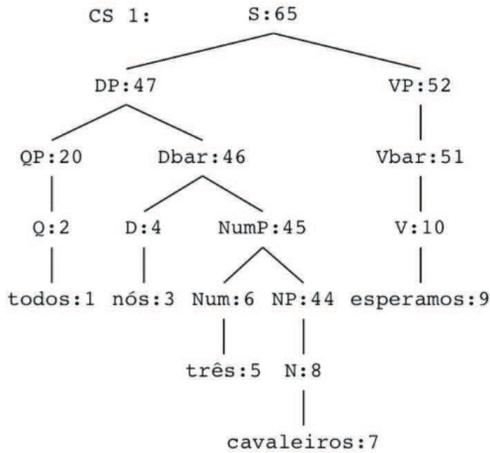


FIGURA 17
Estrutura de constituintes de (62) produzida pelo XLE a partir
de gramática basic-port.lfg



Na árvore da FIG. 16, Dbar não tem um núcleo D, o que viola a teoria X-barra tradicional. Na LFG, contudo, isso não constitui um problema, não sendo necessário (nem mesmo recomendável) postular um D foneticamente vazio nesse caso. De uma maneira geral, o uso de categorias vazias é parcimonioso na LFG, dado o caráter não configuracional e multiestratal da teoria (BRESNAN, 2001). A informação semântica correspondente à que é contribuída para a forma lógica da sentença por um nó D vazio em teorias configuracionais como o Programa Minimalista é projetada na estrutura funcional pela anotação (\wedge PRED)=’PRO’ no segundo membro da disjunção de (59).

Para o exemplo mais complexo de (62), em que um D pronominal rege um complemento e se instanciam três categorias internas ao DP, a saber QP, NumP e NP, são geradas as representações da FIG. 17 e da FIG. 18:

(62) Todos nós três cavaleiros esperamos.

FIGURA 18
Estrutura funcional de (62) produzida pelo XLE a partir
de gramática basic-port.lfg

"todos nós três cavaleiros esperamos"

9	PRED	'	esperar<	[1:cavaleiro]>	']
10		7	PRED	'	cavaleiro	'
51		8	PHI	[GEND masc, NUM pl, PERS 1]
52		44	PHI	[GEND masc, NUM pl, PERS 1]
		5	SIGMA	[GEND masc, NUM pl, PERS 1]
		6		[NUMBER [NUMBER-TYPE card]]
		45	SPEC	[QUANT [QUANT-TYPE universal]]
	SUBJ	3		[QUANT [QUANT-TYPE universal]]
		4	HUMAN	+	, PRON-FORM nós, PRON-TYPE pers]
		46		[HUMAN +, PRON-FORM nós, PRON-TYPE pers]
		1		[HUMAN +, PRON-FORM nós, PRON-TYPE pers]
		2		[HUMAN +, PRON-FORM nós, PRON-TYPE pers]
		20		[HUMAN +, PRON-FORM nós, PRON-TYPE pers]
		47		[HUMAN +, PRON-FORM nós, PRON-TYPE pers]
65	MOOD	indicative, TENSE	PRES			

A estrutura funcional da FIG. 18 exemplifica os aspectos principais de nossa proposta para o tratamento da concordância semântica e da variação nos padrões de concordância em português. Na esteira de D'Alessandro (2008), distinguimos, no âmbito do DP, dois tipos de traços que desempenham um papel na concordância: traços PHI (sintático-formais) e traços SIGMA (semânticos). Tanto os traços PHI quanto os traços SIGMA podem comportar traços de gênero, número e pessoa, mas esses traços não precisam necessariamente constar. Por exemplo, o pronome *nós*, como os demais pronomes que possuem um valor positivo ou negativo para o traço EU na abordagem de Vianna (2006) (QUADRO 2), é especificado apenas para traços de pessoa e número (ver (63)), ao passo que os pronomes de terceira pessoa *ele*, *ela* etc. são especificados também para gênero (ver (64)).³⁰

(63) nós D * (^ PHI NUM)=pl
 (^ PHI PERS)=1
 (^ SIGMA NUM)=pl
 (^ SIGMA PERS)=1
 (^ HUMAN)=+
 (^ PRON-TYPE)=pers
 (^ PRON-FORM)=%stem.

- (64) elas D * (^ PHI NUM)=pl
 (^ PHI PERS)=3
 (^ PHI GEND)=fem
 (^ SIGMA NUM)=pl
 (^ SIGMA PERS)=3
 (^ PRON-TYPE)=pers
 (^ PRON-FORM)=%stem.

Outra diferença entre esses dois grupos de pronomes é que só os membros do primeiro grupo são especificados para o traço humano. Essa diferença desempenha um papel importante na interpretação do gênero semântico do sujeito de sentenças como (65).

- (65) a. A vítima está cansada.
 b. Ela está cansada.
 c. A menina está cansada.
 d. Tu estás cansada.
 e. Nós estamos cansadas.

Diferentemente de (65a) e (65b), em que, à falta de informações contextuais, não podemos inferir o gênero do referente do sujeito, essa inferência é licenciada em (65c) – (65e). Em (65c), a informação de gênero semântico provém do DP [a menina], enquanto que, em (65d) e (65e), é o predicativo que contribui com essa informação. Como podemos verificar pelo contraste entre a FIG. 19 e a FIG. 20, nosso fragmento de gramática reflete corretamente essa distinção, ao atribuir gênero semântico ao sujeito de um predicativo (função XCOMP na LFG) apenas quando esse sujeito portar a especificação [HUMAN]=+.

FIGURA 19
Estrutura funcional de (65a) produzida pelo XLE a partir da gramática basic-port.lfg

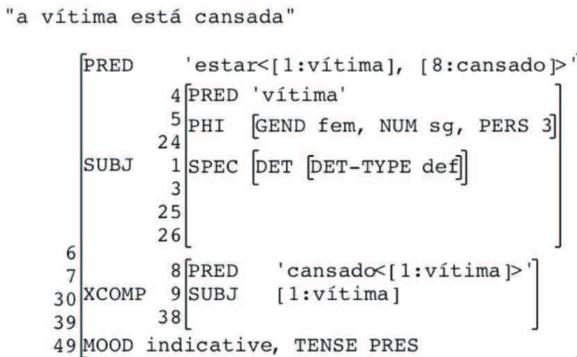
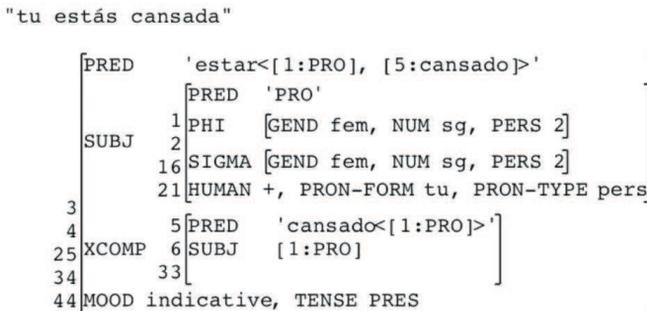


FIGURA 20
Estrutura funcional de (65c) produzida pelo XLE a partir da gramática basic-port.lfg



No nosso fragmento, utilizamos, em vez dos traços [eu] e [EU] propostos por Lopes (2003, 2004), os valores 1, 2 e 3 para o traço de pessoa, conforme o sistema de traços do Projeto ParGram (KING, 2004), visando a um futuro compartilhamento de nossa gramática com essa comunidade de pesquisadores. Isso não significa, porém, ignorar o fato de que a chamada terceira pessoa constitui,

na verdade, uma “não-pessoa”, por se encontrar fora do eixo formado pelo locutor e o alocutário, como enfatiza Lopes (2003, 2004). No entanto, sob uma perspectiva matemática, no âmbito da gramática de unificação, um sistema que distingue entre $[eu]=+$, $[eu]=-$ e $[eu]=\Phi$ é equivalente a um sistema que distingue entre $[PHI PERS]=1$, $[PHI PERS]=2$ e $[PHI PERS]=3$, uma vez que “ Φ ”, não obstante a interpretação que Lopes lhe confere, representa um valor atômico tal qual “+” e “-” e esses três valores diferem entre si tanto quanto os valores “1”, “2” e “3”.

Para que a concordância funcione na LFG (e isso é igualmente válido para outros formalismos baseados na unificação), no sentido de evitar construções agramaticais como as de (66), extraídas de nosso arquivo de teste, é preciso que não só os DPs, mas também as flexões verbais sejam especificadas em termos de traços semântico-discursivos (elementos de SIGMA) e traços sintático-formais (elementos de PHI), embora um ou outro traço possa não ser especificado em um determinado caso individual, como veremos mais adiante.

- (66) a. *a gente espero
 b. *a gente esperas
 c. *a gente esperais
 d. *você espero
 e. *você esperas
 f. *você esperamos
 g. *você esperais
 h. *vocês esperamos

No XLE, podemos modelar as informações relativas à concordância (bem como outros tipos de informação, como o tempo e o modo) contribuídas pelas flexões verbais por meio dos chamados moldes (*templates*), os quais também permitem expressar diversas generalizações linguísticas (BUTT *et al.*, 1999). Para a flexão de 1S, construímos, inicialmente, o seguinte molde:

- (67) V1PS=(^ SUBJ PHI NUM)=sg
 (^ SUBJ PHI PERS)=1
 (^ SUBJ SIGMA NUM)=sg
 (^ SUBJ SIGMA PERS)=1.

O molde (67), com outros moldes, permite elaborar entradas lexicais para as formas verbais de forma muito mais econômica, como podemos constatar

em (68), em que explicamos a função de cada molde por meio de comentários, delimitados, no XLE, por meio de aspas.

(68) espero V * @(OPT-TRANS esperar) “verbo transitivo ou intransitivo”

@V1PS “primeira pessoa do singular”

@VPRES “especifica o tempo verbal”

@IND. “especifica o modo verbal”

Os moldes constituem um recurso muito poderoso no XLE, devido ao seu caráter recursivo, isto é, na definição de um molde, podemos utilizar um molde pré-definido cuja definição, por sua vez, pode recorrer a um outro molde. Para simplificar ainda mais a especificação das propriedades de concordância das formas verbais finitas, criamos o seguinte molde:

(69) SUBJ-AGR(N P) = (^ SUBJ PHI NUM)=N
 (^ SUBJ PHI PERS)=P
 (^ SUBJ SIGMA NUM)=N
 (^ SUBJ SIGMA PERS)=P.

Com (69), a definição de (67) passa a resumir-se a um única linha, como podemos verificar no QUADRO 7. Esta notação compacta tem a vantagem adicional de explicitar que PHI e SIGMA se referem à concordância do sujeito e que, por *default*, os valores de PHI e SIGMA coincidem.

QUADRO 7

Traços PHI e SIGMA das flexões verbais em português

Forma verbal	Pessoa	Especificação de PHI e SIGMA
Espero	1S	V1PS=@/(SUBJ-AGR sg 1).
Esperas	2S	V2PS=@/(SUBJ-AGR sg 2).
Espera	3S	V3PS=(^ SUBJ PHI NUM)=sg (^ SUBJ PHI PERS)=3.
esperamos	1P	V1PP=@/(SUBJ-AGR pl 1).
esperais	2P	V2PP=@/(SUBJ-AGR pl 2).
Esperam	3P	V3PP=(^ SUBJ PHI NUM)=pl (^ SUBJ PHI PERS)=3 "(^ SUBJ SIGMA PERS)=3" (^ SUBJ SIGMA NUM)=pl.

O QUADRO 7 apresenta sinoticamente as especificações de PHI e SIGMA para todo o paradigma de desinências número-pessoais. Nas quatro linhas em destaque, encontram-se as formas com comportamento *default*, em que os valores de PHI e SIGMA coincidem, o que modelamos por meio do molde (69). A forma de 3S não é especificada para o traço SIGMA, uma vez que uma tal especificação conflitaria com as especificações correspondentes de *a gente* (primeira pessoa) e *você* (segunda pessoa), com os quais a forma de 3S concorda sintaticamente. Finalmente, a forma de 3P tem o traço semântico de pessoa (negritoado no QUADRO 7) entre aspas, fazendo com que ele seja ignorado pelo *parser*, o qual passa a analisar como bem formadas sentenças como (70), típicas do dialeto da ilha de São Miguel, nos Açores. Para gerar gramáticas específicas dos demais dialetos do QUADRO 6, basta ativar essa especificação, removendo as aspas.

(70) A gente esperam.

Uma vez especificados os traços de concordância das flexões verbais, vejamos quais os traços correspondentes dos diferentes núcleos funcionais que integram o DP. Para as categorias cujos traços sintáticos de número e pessoa refletem os respectivos traços semânticos, construímos o molde de (71).

(71) AGR(N P) = (^ PHI NUM)=N
 (^ PHI PERS)=P
 (^ SIGMA NUM)=N
 (^ SIGMA PERS)=P.

O QUADRO 8 apresenta as especificações de PHI e SIGMA de núcleos funcionais D, Q e Num em português, destacando em cinza os casos envolvidos na variação da concordância exemplificados em (72). Essas especificações permitem derivar esses dados por meio de um *parser* construído no XLE a partir da gramática basic-port.lfg.

QUADRO 8
Traços PHI e SIGMA de alguns núcleos funcionais do português

Núcleo Funcional	Categoria	Especificação de PHI e SIGMA
Eu	D	@(AGR sg 1)
Nós	D	@(AGR pl 1)
Tu	D	(^ PHI NUM)=sg "(^ PHI PERS)=2" (^ SIGMA NUM)=sg (^ SIGMA PERS)=2
Ele	D	@(AGR sg 3) (^ PHI GEND)=masc
Você	D	(^ PHI NUM)=sg (^ PHI PERS)=3 (^ SIGMA NUM)=sg (^ SIGMA PERS)=2
Vocês	D	(^ PHI NUM)=pl "(^ PHI PERS)=3" (^ SIGMA NUM)=pl (^ SIGMA PERS)=2
O	D	(^ PHI GEND)=masc (^ PHI NUM)=sg (^ PHI PERS)=3
Os	D	(^ PHI GEND)=masc (^ PHI NUM)=pl
Todos	Q	(^ PHI GEND)=masc (^ PHI NUM)=pl
Três	Num	(^ PHI NUM)=c pl ¹

- (72) a. Tu esperas.
 b. Tu espera.
 c. Você espera.
 d. Vocês esperam.
 e. Vocês esperais.
 f. Os brasileiros esperam.
 g. Os brasileiros esperamos.
 h. Os três esperam.
 i. Os três esperamos.
 j. Todos esperam.
 k. Todos esperamos.

Observe que o padrão de concordância típico do pronome *tu* no dialeto POA (e em vários outros dialetos do PB), exemplificado em (72a) e (72b), decorre da desativação do traço formal de pessoa (entre aspas e negrito no QUADRO 8) desse pronome, o que permite que ele concorde tanto com 2S quanto com 3S. Postulamos que a concordância de *tu* exclusivamente com a 3S no dialeto RJM deve-se à inexistência das formas verbais de 2S nesse dialeto.

O padrão típico do dialeto COV, por sua vez, exemplificado em (72d) e (72e), resulta da desativação do traço formal de pessoa do pronome *vocês*. Analogamente, os exemplos (72f) – (72i) e (72j) e (72k), tratados como silepse na gramática tradicional, são derivados na nossa gramática pela inexistência do traço formal de pessoa do plural do artigo definido e do quantificador *todos*. Nesse caso, pressupomos que não se trata de aspecto sujeito a variação dialetal, constituindo, antes, uma característica geral da língua portuguesa, pelo que o traço em questão simplesmente não consta da entrada lexical, não podendo ser ativado por meio da remoção das aspas como nos outros casos de desativação de traço. Postulamos, portanto, uma assimetria entre o singular e o plural do artigo definido, uma vez que apenas o primeiro possui, na nossa gramática, o traço formal de pessoa.

Ao mesmo tempo em que geram os padrões de (72), entre outros, as especificações do QUADRO 7 em interação com as do QUADRO 8, por meio das regras de estrutura sintagmática funcionalmente anotadas no âmbito da gramática basic-port.lfg, excluem as seguintes construções agramaticais, entre outras:

(73) *Você esperas.

(74) *Nós esperam.

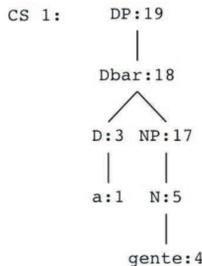
O exemplo (73) é especialmente interessante, uma vez que expõe a dificuldade principal de uma abordagem não completamente formalizada, problema para o qual Falk (2001) e Müller (2010), entre outros, chamam atenção. Como vimos na seção 2.1, Vianna (2006) atribui ao pronome *ocê* a especificação de traço [Φ eu] (QUADRO 2) e, à flexão de 2S, a especificação [–eu] (QUADRO 3). Se “ Φ ” representa um valor neutro ou a ausência de valor na abordagem de Lopes (2004), que subjaz a Vianna (2006), a questão que se coloca é se [Φ eu] pode estabelecer ou não uma relação de concordância com [–eu]. Um linguista falante do português com certeza negará essa possibilidade, mas o fará possivelmente não calculando se o resultado dessa operação de concordância é bem formado ou não, mas com base na sua intuição de falante. Pelo contrário, em uma teoria formal e computacionalmente implementada como a LFG, a resolução do problema de determinar se um dado modelo de gramática do português nesse formalismo gera ou não uma determinada construção é computada algorítmicamente com base nas informações das regras sintagmáticas anotadas funcionalmente e das entradas lexicais, levando em conta as condições de boa formação postuladas pela teoria. Em resumo, para verificar as predições de um dado modelo no contexto de um

sistema como o XLE, basta aplicar um *parser* aos exemplos relevantes (como (73)) ou construir, a partir da gramática, um gerador (*generator*), que gerará a língua definida pela gramática. Se o *parser* reconhecer a construção como bem formada, isso implica que o modelo de gramática subjacente está empiricamente incorreto. No caso da análise de Vianna do pronome *você* e da flexão de 2S, não podemos afirmar com certeza que tipo de representação resulta da combinação dos traços desses dois elementos e se é bem formada ou não. Tudo indica, porém, que essa abordagem não faz uma predição correta a respeito de (73), uma vez que também *a gente* porta a especificação [Φ_{eu}] e concorda com as formas de 1P, especificadas como [+eu]. Analogamente, a especificação [Φ_{fem}], presente em todo o quadro pronominal, deve ser compatível com a especificação [-fem] de uma forma como *cansado* (ver (46)), em exemplos do tipo de (47).

Para concluir esta seção, apresentamos nossa implementação do chamado pronome *a gente*, visando gerar os dados do QUADRO 6, excluindo, ao mesmo tempo, construções agramaticais como as de (49) e (66). Adotando a proposta de Taylor (2009), não analisamos essa expressão como um núcleo D, mas como um DP com uma estrutura sintática interna. Dessa forma, *a gente* não integra o quadro dos pronomes pessoais do português, constituído pelos núcleos D *eu*, *tu*, *você* etc. Diferentemente de Taylor, porém, não postulamos, no âmbito da estrutura sintática de *a gente*, um pronome NÓS foneticamente nulo, uma vez que essa categoria vazia tem como única função fornecer as especificações de traço de pessoa e número manifestas nas construções em que *a gente* concorda com a 1P. Como teoria não configuracional, na LFG esse tipo de informação é representado não na estrutura de constituintes, mas na estrutura funcional e deve ser fornecido pelas anotações funcionais.

FIGURA 21

Estrutura de constituintes da expressão idiomática *a gente*, gerada pelo XLE a partir da gramática basic-port.lfg



Propomos que, no português atual, *a gente* é uma expressão idiomática com a estrutura de constituintes da FIG. 21. Naturalmente, essa estrutura não é a única maneira de implementação, no XLE, de uma análise de *a gente* como expressão idiomática. Outra possibilidade seria tratar essa expressão como item lexical multpalavras. Deixamos para determinar em um trabalho futuro qual das diferentes possibilidades seria a mais elegante e adequada empiricamente.

Que *a gente* possui uma estrutura sintática interna, não obstante a perda do seu significado composicional, é um fato reconhecido por Menuzzi (2000, p. 2004) no âmbito de uma abordagem gerativa anterior à hipótese DP, argumentando em prol do STATUS dessa expressão como “forma pronominal”: “[...] despite having the internal structure of a full NP (i.e., [Det N]), just like *a menina* ‘the girl’, *a gente* cannot be altered by any compositional operation such as adjective modification”. Analogamente, Zilles (2005), em uma abordagem sociolinguística da variação entre *a gente* e *nós*, ora trata a primeira expressão como NP em processo de gramaticalização que estaria adquirindo características de um pronome pessoal, ora como um novo pronome.

No que concerne aos traços de *a gente*, propomos, para os dialetos que realizam a concordância dessa expressão tanto com a 3S quanto com a 1P, a estrutura funcional da FIG. 22, resultado da perda dos traços sintático-formais tanto do determinante *a* quanto do substantivo *gente*, por conta de um desenvolvimento, em linhas gerais, nos moldes do que propôs Lopes (2003, 2004) para a gramaticalização dessa expressão, ao longo do qual *a gente* adquiriu o traço PRED com o valor ‘PRO’, o que lhe confere um significado pronominal, e valores para o traço SIGMA.

FIGURA 22

Estrutura funcional da expressão idiomática *a gente*, gerada pelo XLE a partir da gramática basic-port.lfg³²

"a gente"

4	PRED	'PRO'	
5	SIGMA	[NUM p1, PERS 1]	
17	DFORM	A, HUMAN +, NOUN-FORM GENTE	
1			
3			
18			
19			

Para a concordância de *a gente* com a 3P no dialeto MIG, não precisamos estipular nenhuma alteração na estrutura da FIG. 22, uma vez que decorre da supressão do traço semântico de pessoa da flexão de 3P (QUADRO 7). A análise produzida pela nossa gramática para o exemplo (70) é mostrada na FIG. 23 e na FIG. 25.

FIGURA 23
Estrutura de constituintes de (70),
gerada pelo XLE a partir da gramática basic-port.lfg

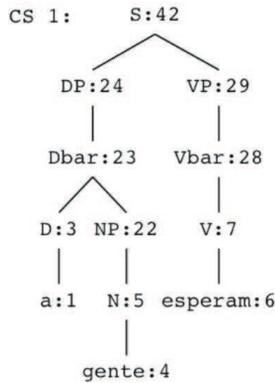
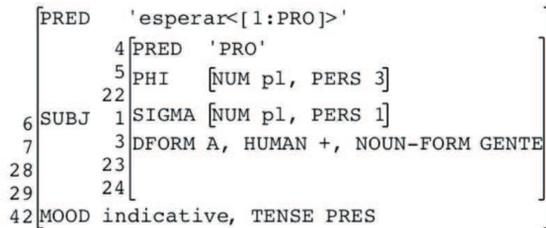


FIGURA 24
Estrutura funcional de (70), gerada pelo XLE
a partir da gramática basic-port.lfg

"a gente esperam"



No dialeto padrão, bem como no dialeto POA, no qual *a gente* somente concorda com 3S, essa expressão possui também a especificação (^ PHI NUM)=sg, a qual impede a geração de exemplos com 1P ou 3P do tipo de (9) e (70). Para testar o efeito de mudanças em uma gramática, como a ativação ou desativação de um determinado traço, o XLE oferece um recurso muito útil, que é a aplicação de um *parser*, compilado a partir dessa nova versão da gramática, sobre o arquivo *testfile.new*, no qual está o resultado da análise da versão anterior. O XLE gera, então, um arquivo *testfile.new.errors* com as discrepâncias entre as duas versões na análise do arquivo de teste, como em (75). Nesse exemplo, podemos constatar que as sentenças de nº 48 e nº 52 do arquivo teste deixam de ser reconhecidas quando introduzimos a especificação (^ PHI NUM)=sg na estrutura funcional de *a gente*.

```
(75) Excerto de arquivo testfile.new.errors
a gente esperamos (1 0.01 11)
# MISMATCH ON: 48 (1 -> 0)
a gente esperam (1 0.01 11)
# MISMATCH ON: 52 (1 -> 0)
```

Como vimos na seção 2, o chamado pronome *a gente* é fruto, conforme Lopes (2003, 2004), da gramaticalização do substantivo *gente*. Na nossa abordagem, a expressão idiomática *a gente* constitui-se de variantes do artigo *a* e do substantivo *gente* desprovidas do traço PHI que só se manifestam nessa construção. Evidentemente, tanto essa forma do artigo definido quanto o substantivo *gente*, nas suas acepções não idiomáticas, continuam produtivos no português, pelo que são gerados exemplos como *toda a gente*, *a gente pobre*, *a sua gente* etc., como se pode verificar, por exemplo, no Corpus do Português (DAVIES; FERREIRA, 2006), do qual extraímos a ocorrência (76). Compare-se a estrutura funcional da expressão idiomática *a gente* na FIG. 22 com a estrutura funcional do DP não idiomático *a sua gente* na FIG. 25. Nessa última, entre outras diferenças, consta a especificação do traço PHI para gênero, número e pessoa, ausente na primeira, e o valor do traço PRED é ‘gente’ e não ‘PRO’.

(76) Era preciso entrar na roda, e rodar, e bambear, mostrar sua figura, arrastar-se no chão, levantar-se do chão, entregar o chapéu a outro quem não fosse capaz de sustentar o ritmo. **A sua gente** não precisava passar o chapéu a ninguém: no segundo acampamento, “os grevistas” já trabalhavam há meia-hora. (LB-GD)

FIGURA 25
Estrutura funcional do DP *a sua gente*, gerada pelo XLE
a partir da gramática basic-port.lfg

"a sua gente"

6	[PRED	'	gente	']						
7	[PHI	[GEND	fem,	NUM	sg,	PERS	3]			
21	[4	[DET	[DET-TYPE	def]]				
5	[22	[PHI	[PERS	3]]				
1	[3	[PRON-TYPE	poss]						
23	[24	[SPEC	[POSS	[PHI	[PERS	3]]

6. Considerações finais

Neste trabalho, procuramos, inicialmente, contribuir para a descrição do português sob uma perspectiva gerativa. Para tanto, elaboramos um fragmento de gramática dessa língua no formalismo da LFG capaz de lidar com a variação dialetal na concordância verbal, fenômeno estreitamente relacionado, nos dialetos que investigamos, à concordância semântica, uma vez que parte dessa variação decorre do licenciamento ou não, por parte de um determinado dialeto individual, da concordância da expressão *a gente* com a 1P ou a 3P.

Um dos atrativos da LFG é a existência de um sistema como o XLE, ao mesmo tempo sofisticado e amigável, representando o estado da arte em termos de ambiente de desenvolvimento de gramáticas computacionais, que permite compilar um *parser* para uma dada gramática elaborada nesse formalismo. Esse *parser* pode ser aplicado, então, a dezenas (ou milhares, se for o caso) de exemplos, tanto gramaticais quanto agramaticais, pelo que se pode verificar, em segundos, a cobertura da gramática em relação aos dados que se pretendeu modelar. Desse modo, inadequações empíricas são facilmente detectadas, que, sem o recurso do computador, poderiam passar despercebidas ao linguista. Após esse processo, uma vez feitas as necessárias modificações na gramática, um novo *parser* pode ser rapidamente compilado e aplicado aos exemplos anteriores ou a novos exemplos. Várias repetições desse ciclo resultam em gramáticas cada vez mais robustas.

Com base no XLE têm sido desenvolvidas, no âmbito do Projeto ParGram, gramáticas no formalismo da LFG para um grande grupo de línguas, tipologicamente tão distintas quanto inglês, urdu, árabe, indonésio e húngaro, para citar apenas alguns exemplos. Por outro lado, algumas dessas gramáticas têm sido utilizadas por empresas líderes em tecnologia da informação em aplicações de processamento automático da linguagem de ampla envergadura, nas áreas de tradução automática, extração e recuperação de informações, resolução de perguntas etc.

Visando a contribuir, em um futuro próximo, para o Projeto ParGram, implementamos um *parser* para o português a partir de um fragmento que abrange a concordância verbal variável de *a gente*, *tu*, *vocês* e DPs do tipo de [todos os brasileiros] em dois dialetos não padrão do PB e quatro dialetos não padrão do PE, além do português padrão. Do mesmo modo que os casos envolvendo *a gente*, a concordância de DPs no plural, encabeçados por determinantes definidos com formais verbais de 1P, é classificada pela gramática tradicional como silepse, tratada não como um processo gramatical regular, mas como figura de sintaxe.

O foco na variação da concordância decorreu da constatação de que o desenvolvimento de *parsers* de línguas como inglês, alemão e francês tem visado mais as variedades padrão, dada a relevância do *parsing* sintático para aplicações como a tradução automática. A análise sintática de *corpora* em linguagem não padrão, contudo, é também relevante, haja vista, por exemplo, a necessidade de compilar florestas sintáticas para esse tipo de texto.

Na elaboração da nossa minigramática, levamos em conta análises recentes de vieses teóricos diversos a respeito dos fenômenos investigados. Tanto estudos de cunho sociofuncionalista quanto outros de orientação gerativista apontam para a necessidade de distinguir entre traços sintático-formais e traços semântico-discursivos para explicar os padrões de concordância com *a gente*. Por outro lado, as análises gerativas mostram que a concordância semântica é um processo gramatical regular e não uma exceção às regularidades da gramática. Essas análises, porém, não abrangem a concordância de *a gente* com a 3P nem tratam de outros fenômenos como a concordância de *vocês* com a 2P. As abordagens funcionalistas, por sua vez, não são formuladas com a precisão e o rigor matemáticos que permitam extrair todas as consequências das representações postuladas para pronomes e flexões verbais, até porque a questão de como funciona o mecanismo da concordância não é nem abordada. Essas deficiências são sanadas na nossa abordagem, cujas predições a respeito das construções gramaticais e agramaticais nos sete dialetos investigados, na subárea da concordância verbal a que o trabalho

se restringe, podem ser verificadas por meio de um *parser* construído no XLE a partir da minigramática proposta. Mostramos que as diferenças entre esses dialetos, relacionadas a esses fenômenos gramaticais específicos, podem ser explicadas principalmente pela ativação ou desativação de traços formais (PHI) ou semânticos (SIGMA) dos núcleos D (e outros núcleos internos ao DP) ou das flexões verbais. Desse modo, para gerar um *parser* específico para um desses dialetos, basta ativar ou desativar o traço ou conjunto de traços que o distingue dos demais. Ao sediar no léxico a fonte da variação da concordância, nossa abordagem faz jus a um dos pilares da LFG, o lexicalismo. Sob essa perspectiva, há um só processo responsável pela concordância: a operação matemática de unificação de estruturas de traços.

Contrariando uma já longa tradição, adotamos a sugestão de Taylor (2009), proposta no quadro da hipótese DP, de que *a gente* não é um pronome, ou seja, não constitui um núcleo D, tal como os pronomes pessoais *nós*, *vocês* etc., mas um DP com uma estrutura sintática interna. Para Taylor, a concordância com a 1P decorre de um núcleo D NÓS foneticamente nulo. Como essa abordagem, desenvolvida no âmbito do Programa Minimalista, depende de operações de movimento para explicar os dados observados, tratamos *a gente*, na nossa implementação de gramática no formalismo não derivacional da LFG, como um DP que constitui uma expressão idiomática, cujo significado pronominal (de primeira pessoa do plural) decorre do valor ‘PRO’ atribuído ao traço PRED, bem como da especificação de traço SIGMA. O comportamento dessa expressão idiomática, no que diz respeito à concordância verbal, decorre, portanto, da sua estrutura de traços em interação com as estruturas de traços das formais verbais, por meio da unificação. Nos dialetos em que *a gente* concorda também com uma flexão verbal no plural, propomos que essa expressão não possui especificação para o traço PHI, o que a diferencia dos núcleos D, que contêm essa especificação.

Para concluir, podemos dizer que o modelo funciona, na medida que gera um subconjunto importante dos padrões de concordância verbal observados e exclui padrões agramaticais. Fica, porém, para pesquisas futuras determinar se a descrição proposta é psicolinguisticamente plausível ou não. Seja como for, a nossa minigramática constitui um embrião de um futuro *parser* do português, baseado na LFG, e esperamos que venha a contribuir para fomentar a investigação dessa língua – que, com a LXGram, já possui uma gramática de ampla cobertura baseada na HPSG – a partir de modelos gerativos não transformacionais, que, segundo acreditamos, podem ajudar a resolver muitos dos problemas do analisar sintático automático do Projeto VISL.

APÊNDICE A

Corpora de onde foram extraídos exemplos

Siglas	Referências
A_004	Corpus Histórico do Português Tycho Brahe, Marquesa de Alorna, <i>Cartas</i>
COV	Corpus Cordial, Inquérito relativo à localidade Covo, Aveiro
LB-GD	Corpus do Português, Luiz Beltrão, <i>A Greve dos Desempregados</i> (1984)
L_002	Corpus Histórico do Português Tycho Brahe, Fernão Lopes, <i>Crónica del-Rei Dom João I</i>
MIG	Corpus Cordial, Inquérito relativo à localidade Ponta Garça (Ponta Delgada), Ilha de São Miguel, Açores

Notas

¹ Este artigo resultou de uma estada como professor visitante na Universidade de Constança, Alemanha, no mês de maio de 2012, proporcionada por bolsas da Fundação CAPES e do Serviço Alemão de Intercâmbio Acadêmico (DAAD), pelas quais agradecemos. Também somos gratos aos participantes do Colóquio de Romanística do Departamento de Linguística da Universidade de Constança, especialmente a Georg A. Kaiser, Miriam Butt e Christoph Schwarze, pelos valiosos comentários e sugestões a respeito de uma primeira versão deste trabalho.

² As demais abreviaturas utilizadas nos exemplos (6) e (7) expandem-se desta forma: IP-MAT = sintagma flexão independente declarativo, D-P = determinante no plural, N-P = substantivo no plural, V-P = verbo no presente do indicativo, NP-ACC * = objeto direto nulo não específico, *pro* = sujeito nulo referencial (CARRILHO *et al.*, 2011).

³ Os dados desta tabela representam uma primeira aproximação em relação a uma análise quantitativa exaustiva dos fenômenos em questão no *corpus* CORDIAL, uma vez que foram utilizadas expressões regulares para detectar sequências de etiquetas ou pares de *token* e etiqueta. Uma verificação manual de parte dos dados nos leva a supor que os valores mais altos apresentados devam ter uma margem de erro de até 10%. Os valores mais baixos (até 12 ocorrências) foram checados manualmente.

⁴ Neste trabalho, usaremos as designações tradicionais *primeira*, *segunda* e *terceira* pessoas para nos referir às três especificações do traço semântico de pessoa [+EU], [-EU] e [ΦEU] (no qual Φ indica a não atribuição de valor ao traço EU), respectivamente, na abordagem de Lopes (2004, p. 54-55). Seguindo o sistema de anotação do CORDIAL, as abreviaturas 1S, 2S, 3S, 1P, 2P e 3P constituem especificações dos traços formais de pessoa (1, 2 e 3) e número (singular e plural, representados por S e P, respectivamente).

⁵ Tenório (2008) constitui exemplo de trabalho brasileiro que trata da concordância entre *a gente* e predicativo sob um enfoque gerativo, na esteira de Pereira (2003).

⁶ Exemplos de língua falada extraídos de Vianna (2006, p. 51).

⁷ Ocorrências obtidas em testes escritos, extraídas de Vianna (2006, p. 77-79).

⁸ Vianna se refere à tese de doutorado, do ano de 1999, correspondente a Lopes (2003). Aparentemente, esse último trabalho constitui uma reprodução da tese.

⁹ Vianna (2006) não explicita quais seriam os traços de número do pronome *você*, mas cremos que esses valores podem ser deduzidos a partir do que ela atribui aos demais pronomes.

¹⁰ Exemplos extraídos de Maia (2009, p. 47).

¹¹ Exemplo extraído de Maia (2009, p. 55).

¹² Maia (2009, p. 55) descreve desta forma (18b) e (18c): “formação de uma só sílaba, em vez de duas: *vamo* > *vão*, resultando um ditongo nasal”.

¹³ Exemplo extraído de Vianna (2006, p. 54).

¹⁴ É possível interpretar (12b) como instância de concordância semântica com um referente de sexo feminino (VIANNA, 2006).

¹⁵ Maia (2012) resenha trabalho inédito de J. Costa e S. Pereira, datado de 2010, em que retomam a questão do estatuto pronominal e da concordância de *a gente*. Como esse artigo ainda não foi publicado e não está disponível nas *homepages* dos autores, preferimos não citá-lo de segunda mão.

¹⁶ DEN DIKKEN, M. “Plurilinguals”, Pronouns and Quirky Agreement. *The Linguistic Review*, vol. 18, p. 19-41, 2001.

¹⁷ Na análise desses exemplos, seguimos, em grande medida, Othero (2009). Para mais detalhes sobre isso, ver seção 5.

¹⁸ Neste trabalho, utilizamos, nos exemplos de gramáticas no formalismo da LFG, a notação do XLE (CROUCH *et al.*, 2011) e, sempre que possível, a nomenclatura gramatical das gramáticas do projeto ParGram, baseada na língua inglesa (KING, 2004). Acreditamos que essa prática, internacionalmente bastante difundida, facilita a leitura do código e da documentação de gramáticas em diferentes línguas.

¹⁹ O Aquamacs, disponível gratuitamente na URL <<http://aquamacs.org/>>, é uma versão, para o sistema operacional Mac OS X, do *emacs*, um dos mais populares editores de texto distribuídos como *software* livre e de código aberto. Esse editor, que oferece suporte para diversas linguagens de programação, pode ser customizado com uma interface amigável para a edição e testagem de gramáticas no XLE, a qual integra a distribuição desse sistema.

²⁰ Os símbolos “^” e “!” correspondem, respectivamente, à seta para cima “↑” e à seta para baixo “↓” da literatura da LFG, que não integram o conjunto de caracteres ASCII, que se podem digitar diretamente a partir de um teclado no padrão norte-americano.

²¹ Na LFG, não se considera a endocentricidade como uma característica universal das línguas. Falk (2001), por exemplo, em seu fragmento de gramática do inglês, adota a teoria X-barra para as projeções das categorias funcionais C, I e D e das categorias lexicais N, V, A e P. No entanto, postula uma categoria exocêntrica S para as sentenças matrizes do inglês. Por outro lado, nega que línguas não configuracionais como o malaiala tenham um VP.

²² A relação de dialetos e fenômenos de variação modelados estão no QUADRO 6, p. 30.

²³ Sobre o *emacs*, ver nota 18. Existe também uma interface gráfica do XLE em Eclipse (CROUCH *et al.*, 2011).

²⁴ Agradecemos à equipe desenvolvedora do Curupira, por nos terem cedido cópia do programa para testes.

²⁵ URL: <<http://beta.visl.sdu.dk/visl/pt/parsing/automatic/trees.php>>. Análise realizada em 26/11/2012.

²⁶ URL: <<http://beta.visl.sdu.dk/visl/pt/parsing/automatic/trees.php>>. Análise realizada em 31/10/2012.

²⁷ A designação POA tem apenas um caráter mnemônico, não devendo se confundir com a sigla homônima da cidade de Porto Alegre.

²⁸ Uma hipótese que talvez mereça ser investigada é que ocorrências de *a gente* com a 1P, em dialetos como o de Porto Alegre, não são produzidas devido ao estigma dessa construção, inculido pela escola, uma vez que a concordância semântica é um processo geral da língua.

²⁹ Deixamos de lado, para não sobrecarregar a exposição, a possibilidade de inclusão de outras categorias internas ao DP, como o PossP (OTHERO, 2009).

³⁰ Na atual fase de desenvolvimento de nossa gramática, não explicitamos que SIGMA e PHI constituem traços relacionados à concordância, como se pode verificar na entrada lexical de (63). No entanto, no formalismo da LFG, dado o caráter recursivo das estruturas de traços, é fácil explicitar essa relação entre os dois traços: basta incluí-los como valores de um traço AGR (de *agreement*, ‘concordância’ em inglês).

³¹ O símbolo “c” à esquerda do sinal de igualdade significa que se trata de uma *constraining equation*, e não de uma *defining equation*, como nos demais casos. No primeiro tipo de equação, é necessário que o valor do traço seja especificado em um outro local da estrutura funcional da sentença (FALK, 2001, p. 78). Com isso, Num é forçado, por exemplo, a combinar-se com um D no plural, o que exclui construções do tipo de **você três*.

³² Os traços DFORM e NOUN-FORM servem apenas para assegurar que a variante do artigo *a* que integra a expressão idiomática se combine apenas com a respectiva variante do substantivo *gente* e vice-versa.

Referências

BERNSTEIN, J. B. The DP Hypothesis: identifying clausal properties in the nominal domain. In: BALTIN, M.; COLLINS, C. (Ed.). *The handbook of contemporary syntactic theory*. Malden: Blackwell, 2003. p. 536-561.

BIRD, S.; KLEIN, E.; LOPER, E. *Natural language processing with Python: analyzing text with the Natural Language Toolkit*. Sebastopol: O’Reilly, 2009.

BRANCO, A. H.; COSTA, F. HPSG: Arquitectura. In: ALENCAR, L. F. de; OTHERO,

- G. A. (Org.). *Abordagens computacionais da teoria da gramática*. Campinas: Mercado de Letras, 2012. p. 219-235.
- BRESNAN, J. *Lexical-Functional Syntax*. Malden: Blackwell, 2001.
- BUSSMANN, H. (Ed.). *Lexikon der Sprachwissenschaft*. 3. ed. Stuttgart: Kröner, 2002.
- BUTT, M. et al. *A grammar writer's cookbook*. Stanford: CSLI, 1999.
- CARRILHO, E. et al. *CORDIAL-SIN – Syntax-oriented Corpus of Portuguese Dialects: Syntactic Annotation System Manual*. Lisboa: Centro de Linguística da Universidade de Lisboa, 2011. Disponível em: <<http://www.clul.ul.pt/cordial-sam/>>. Acesso em: 10 out. 2012.
- COLLINS, C.; POSTAL, P. M. *Imposters*. Nova Iorque: New York University, 2008. (Manuscrito).
- CORDIAL Corpus. Lisboa: Centro de Linguística da Universidade de Lisboa, 2012. URL: Disponível em: <<http://www.clul.ul.pt/en/resources/411-cordial-corpus>>.
- COSTA, J. et al. Concordância com a gente: um problema para a teoria de verificação de traços. In: ENCONTRO NACIONAL DA APL, 16, 2000. *Actas...* Coimbra, 2001. p. 637-657.
- COSTA, J.; PEREIRA, S. Phases and autonomous features: a case of mixed agreement in European Portuguese. In: MCGINNIS, M.; RICHARDS, N. (Ed.). *Perspectives on Phases*. Cambridge: MITWPL, 2005. Disponível em: <http://www.clul.ul.pt/sectores/variacao/cordialsin/pdfs_publicacoes/costa_pereira_2005.pdf> Acesso em: 30 jan. 2012.
- CROUCH, D. et al. *XLE Documentation*. Palo Alto: Palo Alto Research Center, 2011. Disponível em: <http://www2.parc.com/isl/groups/nltx/xle/doc/xle_toc.html>. Acesso em: 5 nov. 2012.
- CUNHA, C.; CINTRA, L. *Nova gramática do português contemporâneo*. 2. ed. Rio de Janeiro: Nova Fronteira, 1985.
- D'ALESSANDRO, R. Syntaktische und Pragmatische Merkmale: eine Fallstudie. In: REMBERGER, E.-M.; MENSCHING, G. (Ed.). *Romanistische Syntax – minimalistisch*. Tübingen: Narr, 2008. p. 277-294.
- DAVIES, M.; FERREIRA, M. *Corpus do Português: 45 million words, 1300s-1900s*. 2006. Disponível em: <<http://www.corpusdoportugues.org.2006>>.
- FARIA, E. *Dicionário escolar latino-português*. Rio de Janeiro: FENAME, 1982.
- FALK, Y. N. *Lexical-functional grammar: an introduction to parallel constraint-based syntax*. Stanford, CSLI Publications, 2001.
- FRANCESCHINI, L. O uso dos pronomes pessoais *nós/a gente* em concórdia – SC. In: SIMPÓSIO NACIONAL E INTERNACIONAL DE LETRAS E LINGÜÍSTICA – SILEL, Vol. 1, n. 1, 2009, Uberlândia. *Anais...* Universidade Federal de Uberlândia: EDUFU, 2009.
- FRANCEZ, N.; WINTNER, S. *Unification grammars*. Cambridge: CUP, 2012.
- GALVES, C.; FARIA, P. *Tycho Brahe Parsed Corpus of Historical Portuguese*. 2010. Disponível em: <<http://www.tycho.iel.unicamp.br/~tycho/corpus/en/index.html>>.
- HAJÍČOVÁ, E. et al. Treebank annotation. In: INDURKHAYA, N.; DAMERAU, F. J.

- (Ed.). *Handbook of Natural Language Processing*. 2. ed. Boca Raton, FL: Chapman & Hall/CRC, 2010. p. 167-188.
- KAPLAN, R. M. Syntax. In: MITKOV, R. (Ed.). *The Oxford handbook of computational linguistics*. Oxford: OUP, 2004. p. 70-90.
- KING, T. H. *Starting a ParGram Grammar*. 2004. Disponível em: <<http://www2.parc.com/isl/groups/nlft/xle/doc/PargramStarterGrammar/starternotes.html>>. Acesso em: 10 nov. 2012.
- KOLB, P. Graphentheorie und Merkmalsstrukturen. In: CARSTENSEN, K.-U. et al. (Ed.). *Computerlinguistik und Sprachtechnologie: eine Einführung*. 2. ed. Heidelberg: Elsevier, 2004. p. 91-110.
- LEMNITZER, L.; ZINSMEISTER, H. *Korpuslinguistik: eine Einführung*. Tübingen: Narr, 2006.
- LJUNGLÖF, P. ; WIRÉN, M. Syntactic parsing. In: INDURKHYA, N.; DAMERAU, F. J. (Ed.). *Handbook of Natural Language Processing*. 2. ed. Boca Raton, FL: Chapman & Hall/CRC, 2010. p. 59-91.
- LOPES, C. R. S. *A inserção de 'a gente' no quadro pronominal do português*. Frankfurt am Main/Madrid: Vervuert/Iberoamericana, 2003.
- LOPES, C. R. S. A gramaticalização de *a gente* em português em tempo real de longa e de curta duração: retenção e mudança na especificação dos traços intrínsecos. *Fórum Linguístico*, v. 4, n. 1, p. 47-80, Florianópolis, 2004.
- LOPES, C. R. S.; CAVALCANTE, S. R. O. A cronologia do voçamento no português brasileiro: expansão de *você*-sujeito e retenção do clítico-*te*. *Linguística*, v. 25, p. 30-65, Madrid, 2011.
- LUFT, C. P. *Moderna gramática brasileira*. 7. ed. Porto Alegre: Globo, 1986.
- MAIA, F. P. S. A variação *nós / a gente* no dialeto mineiro: investigando a transição. *Revista da ABRALIN*, v.8, n.2, p. 45-70, 2009.
- MAIA, F. P. S. *Investigando as formas reduzidas de a gente no dialeto mineiro*. 2012. Tese (Doutorado em Estudos Linguísticos) – Faculdade de Letras, UFMG, Belo Horizonte, 2012.
- MARTINS, R.; OTHERO, G. A. *Parsing do português*. ALENCAR, L. F. de; OTHERO, G. A. (Org.). *Abordagens computacionais da teoria da gramática*. 1. ed. Campinas: Mercado de Letras, 2012, p. 99-126.
- MAZIERO, E.G.; PARDO, T.A.S.; NUNES, M.G.V. *Identificação automática de segmentos discursivos: o uso do parser PALAVRAS*. Universidade de São Paulo: São Carlos, 2007.
- MENDONÇA, A. K. *Nós e a gente* na cidade de vitória: análise da fala capixaba. *PER-cursos Linguísticos*, Vitória, v. 2, n. 4, p. 1-18, 2012.
- MENUZZI, S. First Person Plural Anaphora in Brazilian Portuguese: chains and constraint interactino in binding. In: COSTA, J. (Ed.). *Portuguese Syntax: New Comparative Studies*. Oxford: OUP, 2000. p. 191-240.
- MISTICA, M. et al. LFG Bibliography. [S.l.]: [s.n.], 2012. Disponível em:

- <<http://ww2.cs.mu.oz.au/~mmistica/bibliography.html>> Acesso em: 21. nov. 2012.
- MÜLLER, S. *Grammatiktheorie*. Tübingen: Stauffenburg, 2010.
- NEDERHOF, M.-J.; SATTA, G. Theory of parsing. In: CLARK, A; FOX, C.; LAPPIN, S. (Ed.). *The handbook of computational linguistics and natural language processing*. Malden: Wiley & Blackwell, 2010. p. 105-130.
- OTHERO, G. A. *A gramática da frase em português: algumas reflexões para a formalização da estrutura frasal em português*. Porto Alegre: Edipucrs, 2009. Disponível em: <<http://www.pucrs.br/edipucrs/gramaticadafrase.pdf>>. Acesso em: 02.08.2010.
- PALMER, M.; XUE, N. Linguistic annotation. In: CLARK, A; FOX, C.; LAPPIN, S. (Ed.). *The handbook of computational linguistics and natural language processing*. Malden: Wiley & Blackwell, 2010. p. 238-270.
- PARGRAM / ParSem: An international collaboration on LFG-based grammar and semantics development. [S.l.]: [s.n.], 2012. Disponível em: <<http://pargram.b.uib.no/>> Acesso em: 2 nov. 2012.
- PEREIRA, S. *Gramática Comparada de a gente: variação no Português Europeu*. 2003. Dissertação (Mestrado em Gramática Comparada) – Faculdade de Letras, Universidade de Lisboa, Lisboa, 2003.
- RAPOSO, E. P. Some Observations on the Pronominal System of Portuguese. *CatWPL*, v. 6, p. 59-93, 1998.
- SAG; WASOW; BENDER. *Syntactic theory: a formal introduction*. 2. ed. Stanford: CSLI, 2003.
- TAYLOR, M. On the pronominal status of Brazilian Portuguese *a gente*. In: IRWIN, P.; MALDONADO, V. V. R. (Ed.). *NYU Working Papers in Linguistics*. Vol. 2: Papers in Syntax. New York: New York University, 2009. p. 1-36. Disponível em: <http://linguistics.as.nyu.edu/docs/CP/2345/taylor_09_a_gente_nyuwpl2.pdf> Acesso em: 30 jan. 2012.
- TENÓRIO, T. S. *A concordância de número e de gênero entre o DP pronominal a gente e o predicativo: uma comparação entre o português brasileiro e o português europeu*. 2008. Dissertação (Mestrado em Letras e Linguística) – Faculdade de Letras, UFAL, Maceió, 2008.
- VIANNA, J. B. S. *A concordância de nós e a gente em estruturas predicativas na fala e na escrita carioca*. 2006. Dissertação (Mestrado em Língua Portuguesa) – Faculdade de Letras, UFRJ, Rio de Janeiro, 2006.
- VON HEUSINGER, K.; KAISER, G. A. Mismatching the first person in Romance. In: POMINO, N.; STARK, E. (Ed.). *THE NEREUS INTERNATIONAL WORKSHOP “MISMATCHES IN ROMANCE”*, 5, Konstanz. *Proceedings...* Fachbereich Sprachwissenschaft, Universität Konstanz, 2011. p. 95-111.
- ZELLE, J. M. *Python programming: an introduction to computer science*. Wilsonville: Franklin, Beedle & Associates, 2004.
- ZILLES, A. M. S. Grammaticalization of *a gente* as a cluster of changes: evidence from apparent and real time studies. *Fórum Linguístico*, Florianópolis, v. 4, n. 1, 13-46, 2004.

ZILLES, A. M. S. The development of a new pronoun: The linguistic and social embedding of *a gente* in Brazilian Portuguese. *Language Variation and Change*, Cambridge, vol. 17, n. 1, p. 19-53, 2005.

Data de submissão: 29/11/2012

Data de aprovação: 29/01/2013