



## Exponência de gênero e classe temática em português brasileiro\*

### *Exponence of gender and thematic class in Brazilian Portuguese*

Luiz Carlos Schwindt  
(Universidade Federal do Rio Grande do Sul, Porto Alegre,  
Rio Grande do Sul, Brasil  
CNPq, Brasília, Distrito Federal, Brasil)

(...) o gênero abrange todos os nomes substantivos portugueses, quer se refiram a seres animais, providos de sexo, quer designem apenas “coisas”, como *casa*, *ponte*, *andaiá*, femininos, ou *palácio*, *pente*, *sofá*, masculinos. Explicar todas essas ocorrências pela metáfora, à maneira de um pansexualismo freudiano (...), não nos levaria muito longe. (CAMARA JR. 1970:78)

### RESUMO

*Neste texto, discutimos a interação de restrições fonológicas com restrições responsáveis pela exponenciação de morfemas de gênero e classe temática em português brasileiro numa perspectiva otimalista de base realizacional, a Optimal Interleaving, combinada ao Serialismo Harmônico (cf.: McCarthy 2011). Para isso, partimos de uma hierarquia dessas marcas na língua (cf.: Schwindt 2011) confrontada a dados de produtividade. A análise permitiu formalizar a preferência da realização de classe sobre gênero, no caso do masculino, e de gênero sobre classe,*

\* Este texto integra produção referente ao projeto de pesquisa *Exponência morfológica na fonologia do português brasileiro*, que conta com auxílio, sob forma de bolsas, do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Pró-Reitoria de Pesquisa da Universidade Federal do Rio Grande do Sul (PROPESQ-UFRGS). Participaram desta etapa da pesquisa os estudantes Camila Witt Ulrich, Rossana Saute Kolodny, Pedro Perini Surreaux e Júlia Ricardo.



This content is licensed under a Creative Commons Attribution License, which permits unrestricted use and distribution, provided the original author and source are credited.

*no caso do feminino, e também permitiu lidar com zeros alomórficos como produto de não realização de informação morfológica.*

**Palavras-chave:** *Gênero gramatical. Classe temática. Optimal Interleaving. Serialismo Harmônico.*

## ABSTRACT

*In this paper we discuss the interaction between phonological constraints and constraints responsible for the exponence of morphemes of gender and thematic class in Brazilian Portuguese in an optimalistic realizational approach, the Optimal Interleaving, combined with Harmonic Serialism (cf.: McCarthy 2011). For this purpose, we depart from a hierarchy of these marks in the language (cf.: Schwindt 2011) confronted with productivity data. The analysis allowed to formalize the preference of class over gender in the case of masculine and gender over class in the case of the feminine, and it also allowed to deal with allomorphic zeros as a product of non realization of morphological information.*

**Key-words:** *Grammatical gender. Thematic class. Optimal Interleaving. Harmonic Serialism.*

## 1. Introdução

Neste texto, partindo do entendimento de que os nomes portugueses são dotados de informação morfológica de gênero (GEN) e classe temática (CL) (cf.: Camara Jr. 1970), discutimos a realização desses morfemas através de unidades fonológicas na língua. Defendemos uma hierarquia de exponenciação<sup>1</sup> (cf.: Schwindt 2011) em que se alternam marcas de GEN e CL motivada a partir de dados de produtividade do português brasileiro (PB). A implementação dessas formas se define na base da competição entre restrições de fidelidade, responsáveis pela *realização* de morfemas, e restrições de marcação fonológica, que dão conta da boa formação da sílaba e da palavra prosódica. O modelo adotado rotula-se como HS-OI (McCarthy 2011) – uma perspectiva que, como veremos, combina o Serialismo Harmônico (McCarthy 2007) a uma abordagem realizacional da Teoria da Otimidade (OT) conhecida como Optimal Interleaving (Wolf 2008).<sup>2</sup>

1. O termo “exponenciação, que receberá maior detalhamento no texto, é usado aqui no sentido de Mathews (1974), segundo o qual traços morfossintáticos abstratos se associam a substância fônica, tornando-se *abertos*.

2. No âmbito dos modelos formais, optamos, em diferentes partes do texto, por manter siglas ou mesmo determinados termos em sua forma original, em inglês, quando essas expressões são amplamente

Para empreender essa discussão, o texto está organizado da seguinte forma: na seção 2, apresentamos o quadro geral do sistema de marcação de GEN e CL em português; na seção 3, trazemos alguns dos principais resultados de pesquisa sobre produtividade no emprego de marcas de GEN e CL em PB; nas seções 4 e 5, respectivamente, apresentamos a hierarquia que defendemos e a análise de restrições.

## 2. Sobre a marcação de gênero e classe temática

Os nomes portugueses, a exemplo de outras línguas românicas, comportam informação de gênero. Trata-se de informação categórica, isto é, todo o substantivo da língua vem acompanhado, em sua matriz lexical<sup>3</sup>, de traço valorado para gênero. Esse traço, a que rotulamos como “inerente” (cf.: Katamba; Stonham 2006), manifesta-se binariamente: os nomes são masculinos ou femininos, não havendo, em princípio, evidência de um gênero neutro em português. Assim, vocábulos como *olho*, *criança*, *ponte*, *menino* ou *menina* todos nascem designados quanto ao gênero. A não neutralidade ou ambiguidade do gênero do ponto de vista do falante se revela no processo de concordância, isto é, mesmo um substantivo como *estudante*, que não possui marca explícita ou aberta de gênero, ao se relacionar com outros elementos na sentença, deixará emergir seu gênero inerente: *o estudante dedicado* / *a estudante dedicada* (cf.: Schwindt; Wetzels 2016). É esse aspecto arbitrário e muitas vezes imprevisível do gênero que torna seu caráter flexional suspeito no entender de autores como Villalva (2007), que prefere tratá-lo como variação morfológica em lugar de flexão, reservando esta última designação para fenômenos de natureza morfossintática.<sup>4</sup>

Não há como se negar, entretanto, que a atribuição de gênero é, em grande medida, sistemática em português, como em outras línguas românicas de características semelhantes, no sentido definido por Corbett (1991:7): falantes

---

difundidas na área ou mesmo na falta de traduções consolidadas que as substituam. É, por exemplo, o caso de OI, para *Optimal Interleaving*, e de *input/output*, mas é também o caso de OT – aqui por paralelismo, já que se trata igualmente de nome de teoria. O mesmo vale para nomes de restrições no domínio da OT.

3. Usamos "matriz lexical" aqui para referir o conjunto de informações mais básicas de um nome da língua. Numa perspectiva não lexicalista, como a da Morfologia Distribuída, podemos entender que o gênero faz parte da matriz morfossintática, própria da lista 1 (ou lista de formativos).

4. Há, ainda, quem considere gênero uma categoria derivacional. Uma das evidências seria o fato de que a simples variação de gênero pode levar uma raiz a obter novo significado, como é o caso de *lagarta* / *lagarto*, *barca* / *barco* etc. Essa não nos parece, contudo, uma generalização adequada, pois, para além de se sustentar em um número muito limitado de exemplos passíveis da suposta derivação, trata-se de expediente licenciado também por categorias inquestionavelmente flexionais da língua, como o número, em exemplos como *as férias* / *a fêria* (a última designando a arrecadação diária de um estabelecimento comercial), *os óculos* / *o óculo* (este último um termo da arquitetura que designa um tipo de abertura em uma construção).

nativos têm a habilidade de depreender com grande margem de acerto o gênero dos nomes de sua língua.

Essa habilidade aludida por Corbett, que tomamos como pressuposto, mas que pode ser aferida via experimentação, tem a ver, em nossa compreensão, em primeiro lugar, com um sistema de marcação que integra nossa competência linguística. No caso do português, esse sistema, na ausência do feminino, parece convergir para o masculino (cf.: Camara Jr. 1970). Há diversas evidências em favor dessa hipótese de convergência, mas por ora queremos destacar duas: o comportamento dos falantes diante de palavras novas ou de estrangeirismos e o sistema de concordância, sobretudo numa situação de plural, quando estão envolvidos nomes masculinos e femininos na mesma sentença. Para exemplificar, substantivos recentemente emprestados, por exemplo, do inglês (língua que não possui marca aberta de gênero) e bem acomodados no léxico do PB, como *link*, *scanner*, *mouse* etc., foram todos automaticamente assimilados no masculino na língua sem que isso dependesse de qualquer recomendação formal ou mesmo de um acordo refletido entre os falantes. Também frases como *A gente continua super amigos*, proferida por um homem, ao se referir ao seu recente rompimento com sua parceira, nunca poderiam ser enunciadas no feminino e, neste caso em particular, nem mesmo permitiriam se acomodar a estrutura numa construção similar enunciando os dois gêneros separadamente (*\*A gente continua super amigo e amiga*).<sup>5</sup>

Em segundo lugar, nossa competência para atribuir gênero a palavras novas ou desconhecidas está na dependência, em certa medida, da forma e do significado dessas palavras. No caso da forma, como veremos adiante, quando tratarmos de produtividade, isso está relacionado à grande prevalência na língua de vocábulos masculinos terminados na vogal *o* e de vocábulos femininos terminados em *a*. No caso do significado, tem a ver com a relação entre gênero gramatical e sexo ou gênero social. Diferentemente da relação fonológica, como veremos na seção seguinte, esta última relação, de natureza semântica, atinge um conjunto pequeno de vocábulos do português. A combinação entre a relação fonológica e semântica, porém, ainda que cobrindo um subconjunto ainda menor de itens, por apresentar grande regularidade, parece balizar uma generalização que parecia, no inconsciente coletivo dos falantes, sexo ou gênero social masculino / gênero gramatical masculino / terminação *o*, de um lado, e sexo ou gênero social feminino / gênero gramatical feminino / terminação *a*, de outro. Não é tarefa de uma análise formal discutir o mérito desse senso comum, mas tentar entender em que medida ele reflete o modo como essas estruturas se configuram na gramática.

5. Frase enunciada por Thammy Miranda, artista / pessoa pública que se autodefine como “homem trans”, no programa TV Fama, da Rede TV, exibido em 15/11/2016. Na sequência da entrevista, registre-se, ela se refere à ex-parceira sempre no feminino.

Para fins da análise proposta neste texto, não discutiremos gênero semântico, ainda que voltemos a mencioná-lo ao trazer nossos dados de produtividade. Vamos nos debruçar sobre o sistema de marcação de gênero gramatical, particularmente a instanciada por vogais átonas ou por sua ausência (alomorfe zero), caso em que se estabelece íntimo diálogo com o que entendemos por classe temática – termo sobre o qual discorreremos a partir daqui.

Nomes portugueses terminam preferencialmente em vogais. Essas vogais, entretanto, nem sempre correspondem a gênero. Por equiparação com o sistema verbal, em que claramente se distinguem três classes, e porque essa exigência supostamente fonológica parece ser requerida pela flexão de número, é mais ou menos tácito o entendimento de que, ao lado dos marcadores de gênero, contamos com marcadores de classe (ou vogais temáticas) na língua. Segundo Camara Jr. (1970:76), essas vogais se classificam a partir de sua associação a três grupos de palavras: as de tema em -a (ex. *rosa*), em -o, muito comumente pronunciado [u] (ex. *livro*), e em -e, em geral pronunciado [i] (ex. *dente*).

No quadro a seguir, nos quatro primeiros grupos relacionados na coluna da esquerda, apresentamos vocábulos definidos como *temáticos* e, no último grupo, vocábulos *atemáticos*. Para Camara Jr. (1970), são temáticas as palavras terminadas pelos fonemas /a/, /o/, /e/ átonos e também pelas consoantes aqui representadas pelos arquifonemas (ou formas subespecificadas) /L/, /R/, /S/, já que, segundo o autor, tais palavras também possuem uma forma teórica paroxítona terminada em *e*, que se instancia em formas flexionadas (ex. *papele+s* → *papéis*; *flore+s*; *luze+s*). Atemáticas são as palavras terminadas em vogal acentuada. A discussão promovida neste texto focaliza especialmente os três primeiros conjuntos de palavras temáticas, aquelas que são incontestavelmente fechadas por vogais átonas.

**Quadro 1** - Segmento terminal e gênero em português

Vogal terminal	Gênero	Exemplos
/a/	masculino	atleta, drama
	feminino	menina, mesa
/o/	masculino	menino, bolo
	feminino	-- <sup>6</sup>
/e/	masculino	estudante, abacate
	feminino	estudante, cidade
/L/, /R/, /S/	masculino	anzol(e), amor(e), giz(e)
	feminino	bacharel(a), professor(a), luz(e)
Outros	masculino	café, limão, homem
	feminino	lã, viagem

6. À exceção das palavras *tribo* e *libido*, não há registro de palavras em uso no PB genuinamente substantivas de gênero gramatical feminino terminadas em *o* átono (mesmo *libido* apresenta variante masculina). Estamos excluindo desta análise palavras que consideramos de natureza deadjetival, isto

Descomprometidos, em princípio, com a perspectiva de Item-e-Arranjo, que balizou a análise mattosiana, um rápido olhar sobre o Quadro 1 já nos permite levantar pelo menos três perguntas encadeadas: (i) há evidências para atribuição de *status* morfológico às vogais que fecham as palavras portuguesas?; (ii) uma vez se assumindo *status* morfológico para essas vogais, há que se falar em duas categorias distintas – GEN e CL – ou se trata de unidades de mesma natureza?; (iii) se categorias distintas na morfologia, seus expoentes fonológicos devem ser vistos conjunta ou separadamente?.

A pergunta (i) tem especial relevância num contexto em que se exija univocidade entre elementos do plano do conteúdo e do plano da expressão, como no estruturalismo. Por sua coincidência, ao menos aparente, com sufixos de gênero, o caráter morfológico de *o* e *a* terminais é pouco questionado. A vogal *e*, contudo, provoca maior debate. Isso levou, por exemplo, Camara Jr. a discutir uma possível alomorfa de plural contemplando *-s* e *-es*, para dar conta de formas como *flores*, e finalmente levando-o a assumir, como dissemos, que a vogal *e* era parte de um suposto tema teórico (*flore*), emergindo na flexão e sendo suprimida na forma simples.<sup>7</sup> A hipótese de univocidade, apesar de relativizada em perspectivas teóricas que admitem inserção de segmentos epentéticos *default* para resolução de problemas silábicos (ex. *\*flors* → *flores*), não perdeu completamente sua força nas perspectivas formalistas modernas: na OT, por exemplo, parte-se de ideia de violação de fidelidade para explicar a existência de segmentos de *output* sem filiação ao *input* correspondente e vice-versa; na Morfologia Distribuída (DM), a inserção de substância fônica sem correspondente morfossintático também é, em princípio, tratada como processo marginal à gramática. Assim que, salvaguardado melhor juízo acerca do *status* gramatical de vogais terminais em português, assumimos que todas têm afiliação morfológica. Essa pressuposição, todavia, não invalida a tese de que há, concomitantemente, motivação fonológica para a instanciação de vogais terminais nas palavras portuguesas, como veremos mais adiante.

Considerada a premissa de expectativa de correspondência entre o *output* fonológico e o *input* morfológico, abre-se espaço para as perguntas (ii) e (iii). Destacamos alguns significativos estudos que procuram dar conta de GEN e CL em português numa perspectiva formal, na direção de nossas questões:

---

é, nomes que caberiam numa paráfrase contendo a estrutura do *tipo X* (ex. *é uma cantora do tipo 'soprano'*). Também se excluem daqui, por ora, produtos de truncamento de palavras femininas (*foto*, *moto*) e nomes terminados em ditongo nasal (*a tradição*), porque, além de constituírem grupo pouco expressivo quantitativamente, a classificação mórfica do segmento que fecha esses vocábulos é altamente controversa.

7. O principal argumento para a proposição de formas teóricas fechadas por *e* é o da uniformidade. Se as demais vogais terminais têm evidência de filiação morfológica (ex. *bol+o* / *bol+a*), parece incoerente se tratar apenas *e* como meramente fonológico. Além disso, pares opositivos, como *gas* / *gaze*, e mesmo variações sincrônicas da língua, como *halter* ~ *haltere*, e ainda o processo produtivo de paragoge constatado em dialetos do português europeu (ex. *mar* ~ *mar[ɨ]*), parecem contribuir para a tese de que *-e* se configura de fato como um morfema.

- Camara Jr. (1970) assume que GEN e CL se distinguem maximamente enquanto categorias morfológicas (itens) e nascem combinadas a substância fônica – caso que não permite uma leitura de cumulação ou sincretismo, mas apenas de homofonia;
- Alcântara (2010) propõe que se distinguem relativamente na morfologia, sendo CL a categoria geral à qual GEN se subordina;<sup>8</sup>
- Schwindt (2011) admite que se distinguem maximamente na morfologia e que têm sua realização limitada pela interação entre uma hierarquia baseada em restrições de marcação fonológica e fidelidade morfológica (ponto de vista que se preserva neste texto);
- Armelin (2014) propõe que não se distinguem, já que CL não existe enquanto categoria gramatical – o que se percebe como GEN *versus* CL é produto da exponenciação de um único núcleo funcional, GEN.

Por limite de espaço e de foco, não problematizaremos os pormenores de cada uma dessas abordagens. Limitamo-nos a assumir na presente análise que há categorias morfológicas para CL e GEN independentes descombinadas de fonologia enquanto *input*. Admitimos também que a língua dispõe de sons listados para a realização de cada um desses morfemas. A principal evidência para isso reside na realização de GEN por zero alomórfico, que pode ser identificada no mecanismo de concordância, como em *o linguista atento*, em que o *a* final do núcleo do sintagma corresponde inegavelmente à CL, mas a marca de GEN, mesmo não aberta na superfície, é informação essencial para a concordância no masculino, revelada no artigo e no adjetivo.<sup>9</sup> Essa análise que desvincula morfologia de substância fônica também permite, como veremos adiante, dar conta de processos pouco rentáveis no estágio estável da língua mas relativamente disponíveis na fala da criança, como a realização de um *o* de GEN e não de CL, em expressões como *um crição* ou *o formigo*. Nossa hipótese, então, é a de que morfemas, que nascem abstratos, estão sujeitos a *exponenciação* (cf.: Mathews 1974), isto é, são *realizados* fonologicamente, como preconizado, por exemplo, pela DM ou pela HS-OI. No percurso da exponenciação (ou do mapeamento entre morfologia e fonologia), operam restrições de fidelidade e de marcação fonológica.

8. Alcântara (2010) segue, em linhas gerais, a análise de Harris (1991, 1996), proposta para o espanhol, segundo a qual GEN e CL se distinguem minimamente na morfologia (sob a hipótese, entre outras, de que o único traço lexical é feminino) e tem a exponenciação garantida por regra, num mecanismo que vai do mais restrito ao mais geral (*elsewhere condition*).

9. Para Camara Jr., contudo, neste caso, essa marca, para além de se caracterizar como um alomorfe zero (*realização por zero*), correspondia também a um morfema zero, já que ao masculino o autor não atribuía qualquer substância fônica. Como morfemas nasciam já combinados a fonemas em sua análise, para dar conta do feminino era preciso se entender a realização do morfema *-a*, marcador de gênero no nome (ex. *a linguista atenta*), como alomorfe zero ou como produto de degeminação, dada a sequência de duas vogais idênticas (linguist+a<sub>CL</sub> +a<sub>GEN</sub>).

### 3. Produtividade de gênero e classe temática

A modelagem gramatical no âmbito da morfologia se qualifica quando não é contradita por dados de produtividade, tanto na perspectiva da disponibilidade de regras e de representações na língua – o que vimos discutindo desde a introdução – quanto da perspectiva de sua rentabilidade, isto é, do ponto de vista do quanto essas regras e representações são efetivamente empregadas na comunicação (cf.: Corbin 1987). Um exame que considere essas duas dimensões, no caso de GEN e CL em português, deverá responder, entre outras questões, se há predominâncias no emprego de masculino ou de feminino em PB e, neste caso, se essas predominâncias estão efetivamente expressas por marcas fonológicas e se manifestam em contextos específicos.<sup>10</sup>

Com o objetivo de mapear de forma ampla o emprego de GEN e CL em PB, de modo a contribuir para a proposição de uma hierarquia de seus marcadores (cf.: Schwindt 2011), empreendemos uma pesquisa da qual recortamos alguns resultados para descrever em linhas gerais neste texto. Para discutir a produtividade de GEN e CL no léxico e seu uso em contexto comunicativo, valemo-nos das seguintes amostras:

- (i) léxico dicionarizado – Dicionário Aurélio Eletrônico (17.049 *types*) – todos os substantivos, excetuando-se compostos, derivados de outros nomes e elementos de locuções;
- (ii) uso vernacular – Projeto VARSUL<sup>11</sup> (4.800 *tokens* e 1.266 *types* extraídos de 24 entrevistas, 8 de cada uma das 3 capitais que integram o banco de dados – Porto Alegre, Florianópolis e Curitiba –, estratificadas por sexo, faixa etária e escolaridade);
- (iii) corpus de referência – Projeto ASPA<sup>12</sup> (14.539 *types* – vocábulos não oxítonos terminados nas vogais [a], [u] e [i], divididos em alta e baixa frequência a partir de uma escala logarítmica, cruzados aos dados de (ii)).

Entre os aspectos quantificados, de acordo com as possibilidades de cada uma das amostras, estão *tipologia de GEN (uniforme, biforme, comum-de-dois), segmento terminal, correspondência com sexo, animacidade, concretude, antecedente no sintagma e frequência lexical*.

Por não ser objetivo deste texto pormenorizar os resultados quantitativos, e porque as frequências para as amostras (i) e (ii) apresentaram considerável

10. Defensores de uma política de gênero linguístico partem, muitas vezes, da intuição coletiva de que o português é uma língua que promove um pareamento cem por cento isomórfico entre *o* final e masculino, de um lado, e *a* final e feminino de outro. Tendem também a entender que *e* final tem potencial para expressar neutralidade no idioma. Uma análise de produtividade pode contribuir em grande medida para verificar tal hipótese (cf.: Collischonn; Schwindt 2015).

11. Variação Linguística na Região Sul do Brasil (<http://www.varsul.org.br>).

12. Avaliação Sonora do Português Atual (<http://www.projetoaspa.org>).

proximidade, sintetizamos a seguir um recorte de nossos resultados numa grande tabela, cujos principais aspectos comentamos na sequência. Na discussão que a segue, trazemos resultados adicionais de nossa análise quantitativa, não expressos nesta tabela por limitações de espaço, sempre que se fizerem necessários.

**Tabela 1** - Produtividade de GEN e CL no léxico dicionarizado e no uso no PB

Grupo de fatores	% Dicionário 17049 types	% Uso 4800 tokens	Exemplos
<b>Tipologia de GEN</b>			
Uniforme feminino	51,9	49,4	porta/tribo
Uniforme masculino	43,1	41,1	bolo/enigma
Biforme	<b>2,6</b>	<b>8</b>	menino/a
Comum de dois	2,3	1,6	atleta/estudante
<b>Segmento terminal por tipo</b>			
Uniforme a átono	39,3	32,6	porta/enigma
Uniforme o átono	23	23,6	bolo/tribo
e átono	14,8	15,1	penete/ponte
Biforme a/o átono	<b>1,5</b>	<b>3,6</b>	menino/a
Outros átonos	2,9	4,7	móvel/éter/bônus
Tônicos	18,7	17	jabá/ofurô/cafê/anel/amor/luz
<b>Sexo, animacidade e concretude</b>			
Sexuado	<b>5,5</b>	<b>13,4</b>	homem/mulher/menino/a
Animado	<b>11,9</b>	<b>21,4</b>	criança/lobo/raposa/menino/a
Concreto	90	85,6	porta/bolo/penete/ponte/menino/a

Como se pode observar na Tabela 1, os índices relativos ao que está institucionalizado no dicionário e aquilo que se constatou no uso nas amostras estudadas apresenta grande proporcionalidade, independentemente de a comparação se dar entre *types* numa amostra e *tokens* na outra. Isso está de acordo com nossa expectativa: se entendemos que GEN e CL são unidades morfológicas, e, portanto, parte da gramática do PB, devemos esperar também significativa estabilidade dessas formas, refletindo o caráter coletivo daquilo que tradicionalmente se rotula como língua.<sup>13</sup>

Uma generalização importante a se fazer sobre esses resultados, em primeiro lugar, é a de que há mais palavras terminadas em *a* do que em *o*

13. Isso também se confirma pelo fato de que fatores sociais medidos na amostra de uso – *sexo, idade, escolaridade e localização geográfica* – não se mostraram significativos para a escolha de uma ou outra das categorias investigadas.

na língua, se consideradas as duas amostras estudadas. Em segundo lugar, observa-se um importante equilíbrio entre palavras masculinas e femininas, com leve vantagem, no dicionário e no uso, para as femininas. Em terceiro lugar, caracterizando-se como um dos argumentos para justificar a intuição dos falantes sobre o pareamento perfeito entre *o*/masculino e *a*/feminino, observou-se que, de fato, no léxico dicionarizado e no uso, respectivamente, 95,1% e 89,6% das palavras terminadas em *a* são femininas e 99,9% e 100% das palavras terminadas em *o* são masculinas, ainda que isso não tenha relação robusta com sexo biológico ou gênero social, como mostraremos logo a seguir. Antes, porém, ainda sobre as vogais terminais, cabe registrar que as palavras de gênero único terminadas em *e* estão perfeitamente distribuídas entre femininas e masculinas no léxico dicionarizado do PB (52,8% e 47,2%, respectivamente), ainda que, interessantemente, no uso predomine o feminino (72,7%). Isso, como veremos, está em certa medida refletido nos índices de frequência lexical.

As diferenças numéricas importantes entre os resultados apurados para as duas amostras, destacadas em negrito na Tabela 1, parecem estar todas relacionadas: substantivos bifformes, apesar de lexicalmente pouco representativos, têm relativa frequência no uso, repercutindo em alguma diferença entre *types* e *tokens*; nomes bifformes, em grande extensão, além de se oporem em termos das terminações *a/o*, também são sexuados, animados e concretos. Essas três categorias estão numa relação de subconjunto na língua, e, aparentemente, o traço relevante para a oposição explorada neste estudo é o mais restrito, sexo. No sentido dessas correlações, mas sem a pretensão de esgotá-las, dados os limites e os objetivos deste texto, destacamos um cruzamento que consideramos muito informativo para o debate sobre o padrão de marcação do PB, o que contrasta segmento terminal e correspondência com sexo. No léxico dicionarizado, dos 5,5% dos substantivos sexuados da língua, aproximadamente um terço, 1,5%, diz respeito a substantivos bifformes que fazem a oposição *a/o*. No uso, dos 13,4% de nomes sexuados verificados, palavras que opõem *a* e *o* correspondem a aproximadamente um quarto dos dados, 3,6%. Os demais dados se distribuem entre formas de gênero único e supletivas ou possuem outras terminações.

Um detalhe importante, e que possivelmente seja um dos argumentos em favor da tese de masculino como forma menos marcada da língua, e também para o senso comum de que o português seria uma língua com preferência pelo masculino, é o resultado encontrado para os nomes bifformes e comuns de dois gêneros na amostra de uso: 69,1% e 85,5%, respectivamente, do total de cada uma dessas subcategorias, foram empregados no masculino.

Na Tabela 2, a seguir, apresentamos os dados para *antecedente no sintagma*, isto é, contextos em que o substantivo é precedido por um artigo, pronome ou equivalente. É mensurável apenas na amostra de uso.

**Tabela 2** - Antecedente no sintagma em dados de uso do PB  
(VARISUL – 4800 tokens)

Grupos de fatores	%	Exemplos
Com antecedente	73,1	chamaram [a atendente]
Sem antecedente	26,8	quem é [atendente] aqui

Embora não seja nosso objetivo empreender uma análise morfossintática da marcação de gênero e classe, cabe registrar a preferência generalizada de sintagmas formados por nomes precedidos por determinante na língua. A distribuição entre os diferentes segmentos terminais é bastante equilibrada (em torno de 70% para todos os tipos), com alguma predominância de antecedentes para /e/ (80,9%), o que pode sugerir que se trate de contexto mais propício à desambiguação via antecedente sintático.

Por fim, examinamos o papel da frequência lexical no emprego de gênero na amostra de uso, usando como referência dados do Projeto ASPA. O objetivo desta análise, ainda em andamento, é falsear a hipótese de que o emprego de masculino ou de feminino se associe de modo particular ao uso de palavras muito ou pouco frequentes na língua. Cruzamos os resultados de frequência com masculino/feminino, com tipologia de GEN e com segmento terminal. Como se vê na Tabela 3, a seguir, os resultados se apresentaram muito regulares, com palavras frequentes correspondendo a 80-95% dos dados de todas as categorias pesquisadas, o que vai ao encontro da tese de imprevisibilidade de GEN e CL, sugerindo algum envolvimento de memória no emprego dessas formas, mas sem relevância de modo geral para distingui-las. O único dado que, por ora, destaca-se na análise de frequência, diz respeito ao emprego de formas terminadas em /e/. Nesse contexto, observa-se inversão em relação às palavras terminadas em /a/ e /o/: enquanto no domínio das de alta frequência predominam, mesmo que timidamente, as femininas, no das de baixa frequência, percentualmente as masculinas representam aproximadamente o triplo das femininas. Ainda que ensaisticamente, podemos dizer que isso é coerente com a observação que fizemos para *antecedente no sintagma*, de que esse contexto converge para o não marcado, o masculino, a menos que seja impulsionado por um expediente de desambiguação – aqui a memória facilitada pela alta frequência lexical.

**Tabela 3** - Frequência lexical, gênero e segmento terminal  
(VARISUL codificado a partir do ASPA)

	Alta frequência		Baixa frequência	
	/a/ feminino	88,4%	1345/1521	11,6%
/a/ masculino	92,8%	155/167	7,2%	12/167
/o/ feminino	--	--	--	--
/o/ masculino	89,8%	1157/1288	10,2%	131/1288
/e/ feminino	95,6%	517/541	4,4%	24/541
/e/ masculino	87,2%	171/196	12,8%	25/196

Os resultados sobre produtividade de GEN e CL apresentados nesta seção não permitem sustentar, em números brutos, uma correlação entre masculino e não marcado em português brasileiro. Pelo contrário, de modo geral o feminino apresenta frequências levemente superiores no léxico dicionarizado e no uso. Por outro lado, como vimos, se observados subconjuntos de dados, como o dos poucos nomes que estabelecem pareamento com sexo biológico ou gênero social, e mesmo o do uso efetivo dos biformes e comuns de dois, apesar de sua frequência também limitada, podemos afirmar que as formas masculinas são mais frequentes do que as femininas. Isso, entendemos, numa visão de produtividade que relaciona disponibilidade com rentabilidade, é evidência adicional favorável à tese de masculino como forma não marcada na gramática do português.

#### 4. Hierarquia e exponenciação de gênero e classe temática

Em posse do mapeamento da produtividade da marcação de GEN e CL em PB, nesta seção recuperamos a hierarquia de realização dos expoentes desses morfemas na língua, defendida em Schwindt (2011).

A escala de que nos valem combina expoentes fonológicos para GEN e CL de modo a assegurar as preferências de realização dessas marcas. Centramos, neste ponto, em palavras temáticas e, de modo particular, em palavras que se superficializam com uma vogal em posição final, ainda que acreditemos que a mesma hierarquia seja, de modo geral, adequada para o tratamento das demais palavras temáticas do português. Assumimos uma perspectiva otimalista em nossa análise. Fazemos isso mais por conta da crença na superioridade explicativa de restrições violáveis sobre regras do que por conta de uma hipótese de paralelismo estrito – menos relevante para os fins desta proposta. Além disso, assumimos nesta análise uma perspectiva realizacional da morfologia, isto é, de que morfemas abstratos dependem de atribuição de substância fônica no curso de sua exponenciação (Wolf 2008), o que detalharemos na próxima

seção. Nesses termos, a hierarquia em questão se licencia a partir da restrição de fidelidade *input/output* MAX-M(F), a seguir definida.

(1) MAX-M(F) (McCarthy 2011:4)

Para cada ocorrência  $\phi$  do traço F, se  $\phi$  está na estrutura sintática e não tem nenhum correspondente na estrutura fonológica, assinale uma marca de violação.

Em nossa proposta, M corresponde aos morfemas de GEN, com especificação para masculino ou feminino, e de CL, com especificação para as classes 1, 2 ou 3. A cada um desses morfemas corresponde um segmento sonoro independente.<sup>14</sup> Realizar o morfema significa não violar MAX-M(F) e, portanto, atribuir substância fônica à informação morfológica.

A escala que adotamos expressa uma relação de *elsewhere*, isto é, parte da fidelidade à realização de morfemas mais gerais à fidelidade da realização de morfemas mais específicos. Essa hierarquia de restritividade é o que a literatura denomina comumente *marcação morfológica*<sup>15</sup> – aqui com a vantagem de associar GEN e CL numa só escala.

$$(2) \begin{array}{cccccc} \text{MAX-M} & >> & \text{MAX-M} & >> & \text{MAX-M} & >> & \text{MAX-M} & >> & \text{MAX-M} \\ (\text{CL3}) & & (\text{FEM}) & & (\text{CL2}) & & (\text{CL1}) & & (\text{MASC}) \\ e & & a & & o & & a & & o \\ & & & & & & & & (\text{Schwindt 2011:271}) \end{array}$$

Entre as generalizações cobertas pela escala em (2) estão as que seguem.

- Os segmentos que correspondem às especificações morfológicas nessa escala podem coincidir fonologicamente, mas não se confundem morfológicamente, isto é, estão sujeitos apenas a homofonia acidental. O fato de não haver cem por cento de isomorfismo entre GEN e CL com nenhuma das três vogais envolvidas constitui-se como o principal argumento para se rejeitar uma hipótese de sincretismo na análise dessas categorias.
- O fato de *e* ser a vogal menos previsível em termos de GEN coloca CL3 no topo da hierarquia (ex. *o pente / a ponte / o, a agente*). Isso se reflete nos dados, pois o léxico do português, dicionarizado e em uso, aponta

14. A especificação para segmentos poderia ser substituída por matrizes de traços, ficando na dependência da concepção sobre primitivos fonológicos que subjaz à análise. Essa escolha nos parece irrelevante para fins desta proposta. Elegemos segmentos em função da simplicidade de exposição.

15. Neste texto, o termo *marcação* é utilizado em dois sentidos. O primeiro, morfológico, se refere à oposição ‘mais específico’ versus ‘mais geral’ (ou *default*). O segundo sentido é o assumido pela Teoria da Otimidade, em que restrições que não se referem ao input, as de marcação, podem ser dominadas por ou dominar as de fidelidade, que mapeiam *input-output*. O primeiro sentido se revela na escala de MAX, porque ali, onde todas as restrições são de fidelidade, está em jogo a ordem de preferência de realização de morfemas (do mais ao menos marcado). O segundo sentido se revela quando restrições fonológicas – de marcação, portanto – intervêm nesse ranking, interagindo com as restrições de fidelidade.

para equilíbrio entre palavras masculinas e femininas nesta classe. Assim, GEN em palavras masculinas ou femininas de CL3 é sempre produto de não realização fonológica (ou zero alomórfico). Ainda do ponto de vista da rentabilidade, *e* é também a terminação menos recorrente entre as três vogais, provando seu caráter mais específico ou mais marcado morfológicamente. Sustentamos que, ao lado desta gramática mais geral, a língua dispõe de outra, com ela concorrente, que permite que palavras de CL3 realizem o expoente de feminino, como em *presidenta* ou *gerenta*. Nesta gramática alternativa, há uma inversão entre as duas restrições que encabeçam a hierarquia em (2), isto é, MAX-M(FEM) passa a dominar MAX-M(CL3).<sup>16</sup>

- A posição alta de feminino, na segunda posição da hierarquia, permite lidar com o padrão mais distintivo de marcação de GEN em português. Em termos de disponibilidade, isso parece bastante justificado: qualquer substantivo sexuado pode ser flexionado no feminino pela adição de *a* depois da raiz (ex. *o, a hétero* → *a hétera*; *o, a tenente* → *a tenenta*), mas o contrário não é verdadeiro, ou seja, a adição de *o* não forma masculinos na gramática padrão (*o, a tenente* → *o \*tenento*; *o, a linguista* → *o \*linguisto*). Em termos de rentabilidade, apesar de haver pequena vantagem de nomes femininos sobre masculinos em termos absolutos na língua – o que poderia conduzir à ideia de que o primeiro grupo é mais geral do que o segundo –, nomes masculinos, como dissemos, superam os femininos no conjunto dos biformes e comuns de dois, revelando que FEM é, de fato, mais específico, ou mais marcado, do que masculino.
- Os próximos dois pontos da escala trazem os marcadores de CL2 e CL1 respectivamente. A ordem entre eles é claramente sustentada, também, da ótica da produtividade. De modo geral, nas duas amostras estudadas até aqui, encontramos mais palavras terminadas em *a* do que em *o* em PB. Às palavras femininas terminadas em *a*, que superam em termos gerais as masculinas em *o*, acrescenta-se que a língua conta com palavras masculinas terminadas em *a* (ex. *problema, enigma*), o que não acontece com *o* (em que palavras genuinamente femininas são absolutamente excepcionais). Assim, pertencer à CL2, que se substancia pela vogal terminal *o*, é mais específico, ou mais marcado

16. Palavras terminadas em *e* não devem ser todas tratadas da mesma forma em português no que diz respeito à sua possibilidade de exponenciar feminino, em função de sua relação mais ou menos estreita com resíduos de participio presente na língua, manifestos por vezes no sufixo *-nte*. Assim, *presidente* é candidato mais provável à flexão com *-a* (*presidenta*) do que *amante* (*\*amanta*), uma vez que esta última palavra tem maior vinculação sincrônica com o verbo de que se origina do que a primeira, apresentando, portanto, maior resistência a processos de natureza flexional. Mais no centro desse *continuum*, localizam-se, possivelmente, palavras como *estudante*.

em termos morfológicos, do que pertencer à CL1, que se substancia por *a*, justificando a hierarquia.

- A escala poderia ser fechada no penúltimo ponto, considerando que parece haver pareamento total entre MASC e CL2, o que permite tratar todo *o* final como marcador de classe. Formalizar, contudo, uma hipótese de sincretismo, ou mesmo de cumulação neste ponto da escala inviabilizaria, como antecipamos na seção 1, uma explicação para formas aparentemente disponíveis na gramática da criança e até mesmo numa fala criativa de adultos, como *crianço*, *formigo*, *bacharelo* etc. Além disso, supondo que essa hierarquia possa ser válida em alguma medida para lidar com pronomes, para os quais não se espera necessariamente marcador de CL, haveria evidência para a proposição de um morfema independente de MASC na língua. MASC, portanto, no fim da hierarquia, dá conta de uma potencialidade (disponibilidade) da gramática do português e não se submete, neste caso, a medidas objetivas de rentabilidade.

A hierarquia de marcação morfológica que apresentamos nesta seção, expressa por restrições de fidelidade, relaciona-se com restrições de marcação fonológica, como a exigência, comum em diversas línguas românicas, de que as palavras terminem em vogal, e a evitação ao hiato. Esta última restrição está crucialmente implicada na análise que apresentamos na próxima seção.

## 5. Abordagem realizacional de restrições para GEN e CL

Como prenunciamos, a hierarquia de representações aqui proposta se coaduna com uma perspectiva formal de gramática de base realizacional, isto é, um modelo em que categorias morfológicas ou morfossintáticas adquirem substância fônica no curso da derivação (no sentido de processamento da gramática). Isso permite, como dissemos, não apenas dar conta da exponenciação prevista em (2), mas lidar de modo formalmente adequado com a não realização fonológica de um morfema – o que a tradição convencionou chamar de zero alomórfico.

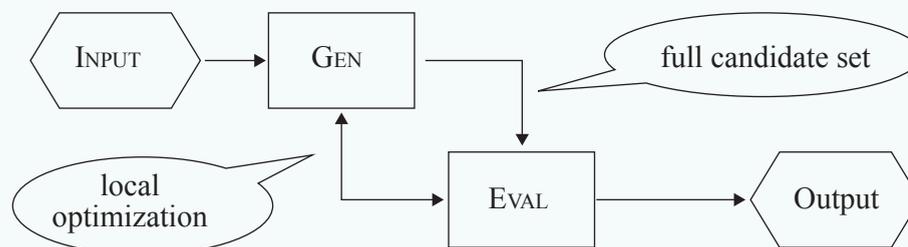
Como mencionado na introdução, adotamos um modelo de restrições rotulado como HS-OI (McCarthy 2011): uma perspectiva que combina o Serialismo Harmônico (HS, de *Harmonic Serialism*), proposto por McCarthy (2007), a uma abordagem realizacional da Teoria da Otimidade, conhecida como Optimal Interleaving (OI), proposta por Wolf (2008).<sup>17</sup>

17. Como nos interessa discutir a interação entre morfologia e fonologia neste texto mais do que o possível papel (morfo)sintático das categorias GEN e CL, optamos por uma perspectiva paralela de restrições, que confere a esse mapeamento *status* legitimamente gramatical. Salvaguardadas particularidades analíticas, não vislumbramos contradição significativa, contudo, no casamento entre essa proposta e a abordagem assumida pela Morfologia Distribuída.

De acordo com a HS, teoria proposta inicialmente sob o rótulo de OT-CC (de *candidate chains*) e baseada na ideia de serialismo harmônico proposta por Prince e Smolensky (2004[1993]), os candidatos em OT não apenas incluem formas de superfície, mas também uma série de formas intermediárias. Cada uma dessas formas é minimamente distinta da forma que a precede imediatamente. Tal modelo pressupõe, para tanto, uma interação entre o gerador (GEN) e o avaliador (EVAL). Isso o diferencia de modo importante da OT *standard*.

- (3) Interação de GEN / EVAL em HS
  - a. GEN realiza uma só operação de cada vez.
  - b. Cada operação deve resultar em melhora harmônica, de acordo com EVAL.

(4) Arquitetura da HS-OI



(McCarthy 2007:63)

A Optimal Interleaving (OI), proposta por Wolf (2008), é uma teoria realizacional, que concebe que formas fonológicas de morfemas são resultado de processos que realizam (*spell-out*) os traços morfossintáticos abstratos que os caracterizam. Aqui, apesar da aparência, o expoente fonológico não substitui literalmente o morfema, mas estabelece com ele uma relação de correspondência, o que permite que a estrutura de traço do *input* esteja acessível durante toda a derivação. Na OI, GEN é limitado a fazer uma mudança de cada vez, o que equivale a dizer que se pode inserir somente um morfema de cada vez. O *spell-out* da raiz tem precedência sobre o *spell-out* dos demais morfemas. Não há uma ordem pré-determinada de *spell-out* dos demais morfemas, nem mesmo qualquer prioridade relacionada com sua proximidade da base, como nas teorias derivacionais – pelo contrário, uma das vantagens do modelo está justamente no fato de adotar uma versão de *look-ahead* que permite que um alomorfe mais periférico seja inserido antes de um menos periférico, podendo, inclusive, o primeiro influenciar o segundo. Nesse quadro, alomorfes, diferentemente da forma como são tratados na OT *standard*, competem no ponto de *spell-out*, não na estrutura de superfície.

As restrições de fidelidade que dão conta da hierarquia de realização dos morfemas competem com condições de boa formação de unidades prosódicas, que se instanciam sob a forma de restrições de marcação.

No caso de GEN e CL, assumimos que está em jogo uma restrição de marcação que dá conta da evitação de hiatos, expressando uma tendência bastante comum nas línguas do mundo.

- (5) \*HIATUS (McCarthy 2002:116)  
Hiatos são proibidos.

A seguir, em (6) e (7), propomos a análise de um par de substantivos biformes sexuados, *garoto/garota*. A avaliação se dá através de passos, que, como dissemos, representam cada interação entre GEN e EVAL, e processam uma mudança de cada vez. O ponto de partida são estruturas morfológicas abstratas, isto é, sem substância fônica (representadas por grafias em versalete tanto no *input* quanto nos candidatos). A seta indica o candidato que violou minimamente as restrições mais altas, o *output* ótimo de cada estágio.

- (6) Análise de *garoto*

Step 1: MAX-M(ROOT)>> MAX-M(CL2)>> MAX-M(MASC)

GAROT-CL2-MASC	MAX-M(ROOT)	MAX-M(CL2)	MAX-M(MASC)
→ a. garot-CL2-MASC		*	*
b. GAROT-CL2-o	*	*	
c. GAROT-o-MASC	*		*

Step 2: MAX-M(CL2)>> MAX-M(MASC)

garot-CL2-MASC	MAX-M(CL2)	MAX-M(MASC)
a. garot-CL2-MASC	*	*
→ b. garoto-MASC		*
c. garot-CL2-o	*	

Step 3: \*HIATUS>> MAX-M(MASC) (convergence)

garoto-MASC	*HIATUS	MAX-M(MASC)
a. garotoo	*	
→ b. garoto-MASC		*

No caso de (6), em que se avalia a forma masculina do nome, observe-se que, no primeiro passo, o candidato escolhido é aquele que realiza a raiz, ou seja, que não viola MAX-M(ROOT), apesar de violar as restrições que dão conta da realização dos morfemas de classe e gênero. No passo 2, o *input* da avaliação é, agora, o *output* escolhido no passo 1, isto é, aquele que já realizou a raiz. Não há mais, portanto, geração de candidatos com raiz não escandida. Por isso mesmo, e por economia de espaço, omitimos nesse passo a restrição MAX(ROOT). A mesma lógica se aplica a todos os *tableaux*. A competição agora se dá entre escandir o morfema de classe ou o morfema de gênero. A dominância de MAX-M(CL2) sobre MAX-M(MASC) resolve este conflito em favor da realização de classe. No passo 3, a restrição de marcação que milita contra hiatos domina a realização de gênero masculino. O resultado é a escolha de um *output* masculino

sem expoente fonológico para GEN – em termos estruturalistas, um alomorfe zero. A avaliação *converge* neste passo, já que o *output* de um próximo passo seria incapaz de produzir mudanças na direção do *output* final. (cf.: McCarthy 2000).

(7) Análise de *garota*

Step 2: MAX-M(FEM) >> MAX-M(CL1) (converges at step 3)

garot-CL1-FEM	MAX-M(FEM)	MAX-M(CL1)
a. garot-CL1-FEM	*	*
b. garota-FEM	*	
→ c. garot-CL1-a		*

Por limitação de espaço, no caso da forma feminina, em (7), restringimo-nos a apresentar o passo 2. Aqui, ao contrário do que ocorre em (6), GEN domina CL, fazendo de *c* o candidato ótimo. Os demais passos se assemelham à análise do masculino.

Em (8) e (9), a seguir, apresentamos a análise de substantivos de dois gêneros, sexuados, não marcados fonologicamente, *o atleta* e *a atleta*, respectivamente.

(8) Análise de (*o*) *atleta*

Step 2: MAX-M(CL1) >> MAX-M(MASC) (converges at step 3)

atlet-CL1-MASC	MAX-M(CL1)	MAX-M(MASC)
a. atlet-CL1-MASC	*	*
→ b. atleta-MASC		*
c. atlet-CL1-o	*	

(9) Análise de (*a*) *atleta*

Step 2: MAX-M(FEM) >> MAX-M(CL1) (converges at step 3)

atlet-CL1-FEM	MAX-M(FEM)	MAX-M(CL1)
a. atlet-CL1-FEM	*	*
b. atleta-FEM	*	
→ c. atlet-CL1-a		*

A análise que propusemos para (7) e (8) é a mesma que dá conta da formação de (8) e (9). De acordo com a hierarquia proposta em (2), o *a* final de palavras masculinas equivale à realização de CL, ao passo que o *a* final de palavras femininas é sempre expoente de GEN, ao contrário do que ocorre com *o* final, que é preferencialmente marcador de CL em qualquer circunstância. Isso, ao mesmo tempo que explica a excepcionalidade de palavras femininas terminadas em *o* (ex. *tribo*), consegue dar conta da agramaticalidade de formas como *\*atleto*.

Em (10) e (11) estão *tableaux* para os substantivos uniformes *drama* e *porta*, masculino e feminino, respectivamente, ambos terminados em *a*.

(10) Análise de *(o) drama*Step 2:  $\text{MAX-M}(\text{CL1}) \gg \text{MAX-M}(\text{MASC})$  (converges at step 3)

dram-CL1-MASC	Max-M(CL1)	MAX-M(MASC)
a. dram-CL1-MASC	*	*
→ b. drama-MASC		*
c. dram-CL1-o	*	

(11) Análise de *(a) porta*Step 2:  $\text{MAX-M}(\text{FEM}) \gg \text{MAX-M}(\text{CL1})$  (converges at step 3)

port-CL1-FEM	MAX-M(FEM)	MAX-M(CL1)
a. port-CL1-FEM	*	*
b. porta-FEM	*	
→ c. port-CL1-a		*

A análise de (10) e (11), que respeita *ranking* idêntico ao proposto nas análises anteriores, mostra que a mesma generalização apresentada para substantivos biformes se aplica aos uniformes.

Por fim, nomes de CL3, isto é, que terminam sistematicamente em *e* (ex. *o/a presidente*), têm a superficialização de CL sobre GEN garantida pela posição alta de CL3 sobre todas as demais restrições da família MAX-M, como se viu em (2). A distinção de gênero neste caso é mais abstrata, pois é igualmente não realizada na superfície tanto no feminino quanto no masculino. Nos casos em que se licencia uma forma feminina em *a* (ex. *presidente/presidenta*), como afirmamos anteriormente, assumimos que uma gramática alternativa (um *ranking* alternativo, em termos otimalistas), está em jogo:  $\text{MAX-M}(\text{FEM}) \gg \text{MAX-M}(\text{CL3})$ .

Outra vantagem de uma análise paralela em HS-OI está no fato de permitir que se analisem estruturas maiores do que a palavra. Em (12), propomos a análise de um sintagma constituído de um determinante – neste caso, um artigo definido – e um nome de gênero único: *a ponte*. Assumimos, como hipótese de trabalho, que ao artigo corresponde um morfema específico, DEF, que convencionamos estar associado ao fonema *o*. Isso implica que não tenha precedência de realização como é próprio de raízes. Consideramos, ainda, que, por se tratar de palavra funcional, não recebe marcadores de classe, mas que, como os substantivos da língua, se sujeita à flexão de gênero.<sup>18</sup>

18. A análise de frases que apresentamos aqui impõe se admitir que a computação em HS-OI se processa sobre o *output* da sintaxe. Isso quer dizer que o processo de concordância, que toma o nome como nuclear para estender o traço aos demais elementos de um NP já deve ter ocorrido, ou deve ser alcançado por restrições em paralelo capazes de expressar sua precedência sobre a seleção das expoentes de GEN e CL. Esta análise se restringe à avaliação tão somente da seleção desses expoentes fonológicos que substanciam os morfemas expressos no *input*.

(12) Análise de *a ponte*

Step 1: MAX-M(ROOT)&gt;&gt;\*MAX-M(FEM)&gt;&gt;MAX-M(DEF)

[[DEF-FEM] <sub>DET</sub> [PONT-CL3-FEM] <sub>N</sub> ] <sub>NP</sub>	MAX-M (ROOT)	MAX-M (FEM)	MAX-M (DEF)
→ a. DEF-a pont-CL3-FEM		*	*
b. DEF-a PONT-e-FEM	*	*	*
c. DEF-a PONT-CL3-a	*		*
d. o-FEM pont-CL3-FEM		**	
e. o-FEM PONT-e-FEM	*	**	
f. o-FEM PONT-CL3-a	*	*	

Step 2: MAX-M(CL3)&gt;&gt;\*HIATUS&gt;&gt;MAX-M(FEM)&gt;&gt;MAX-M(DEF)

[[DEF-a] <sub>DET</sub> [pont-CL3-FEM] <sub>N</sub> ] <sub>NP</sub>	MAX-M (CL3)	*HIATUS	MAX-M (FEM)	MAX-M (DEF)
a. DEF-a pont-CL3-FEM	*		*	*
b. oa pont-CL3-FEM	*	*	*	
→ c. DEF-a ponte-FEM			*	*
d. oa ponte-FEM		*	*	
e. DEF-a pont-CL3-a	*			*
f. oa pont-CL3-a	*	*		

Step 3: \*HIATUS&gt;&gt;MAX-M(FEM)&gt;&gt;MAX-M(DEF) (convergence)

[[DEF-a] <sub>DET</sub> [ponte-FEM] <sub>N</sub> ] <sub>NP</sub>	*HIATUS	MAX-M (FEM)	MAX-M (DEF)
→ a. DEF-a ponte-FEM		*	*
b. DEF-a ponte-a	*		*
c. oa ponte-FEM	*	*	
d. oa ponte-a	**		

A análise de (12) mostra, no passo 1, que, enquanto a exponência da raiz é priorizada no nome, no determinante, o *ranking* conduz à priorização da realização do gênero feminino. No passo 2, a posição de MAX-M(CL3) força a escansão da vogal terminal no nome sobre GEN; no determinante, \*HIATUS impede a realização de DEF. No passo 3, \*HIATUS impede que GEN se realize no nome, e nada se altera em relação ao estágio anterior no que concerne ao determinante. A derivação converge neste passo. Pode-se dizer, assim, que o artigo definido feminino em português é, na superfície, produto da realização do morfema de FEM e da não realização de DEF (alomorfe zero).

O *ranking* final que dá conta da exponência de GEN e CL em português apresentada nesta seção é o que segue.

## (13) Ranking final

Max-M(ROOT) >> Max-M(CL3) >> \*HIATUS >> Max-M(FEM) >> Max-M(CL2) >> Max-M(CL1) >> Max-M(MASC) >> Max-M(DEF) >> Max-M<sup>19</sup>

Nesta análise tomamos como pressuposto que os substantivos do português contam subjacentemente com informação morfológica de GEN e CL, sem exceções. A exigência de escansão de dois morfemas gramaticais se resolve operacionalmente com a administração do conflito entre restrições de fidelidade (hierarquizadas entre si em relação à realização desses morfemas) com apenas uma restrição de marcação fonológica, a que concerne à evitação de hiatos. Por outro lado, análises alternativas, que considerem, por exemplo, possível sincretismo entre as categorias gramaticais aqui analisadas, podem necessitar de restrição fonológica adicional, a fim de evitar que, por conta da precedência de escansão da raiz, candidatos fechados por consoantes (de modo geral, mas sobretudo por obstruintes) se superficializem. Uma alternativa é a proposição de restrições de estrutura silábica, capazes de inibir codas indesejadas, em especial em posição final. Outra alternativa, proposta em Schwindt (2011), é uma restrição que imponha que palavras prosódicas sejam fechadas por vogal (possivelmente ativa em diferentes línguas românicas).<sup>20</sup> Uma restrição desse tipo pode operar como argumento coadjuvante para sustentar a categoricidade de vogais terminais no sistema apesar de sua muitas vezes imprecisa associação às informações morfológicas que veiculam.

## 6. Considerações finais

Neste texto discutimos a exponenciação de GEN e CL em português, particularizando-a para a variedade brasileira a partir do exame de sua produtividade. Tomando por base uma hierarquia realizacional para essas marcas (cf.: Schwindt 2011), propusemos uma análise no modelo HS-OI (cf.: McCarthy 2011). Nesta seção, recuperamos o que entendemos como alcances de nosso estudo, além de destacar limitações e questões residuais.

No que concerne ao exame de produtividade, assumindo o pressuposto de que análises formais devem dar conta do que é disponível e rentável na gramática de uma língua, examinamos quantitativamente o comportamento dos

19. Omitimos, por simplificação, ao longo da análise, a restrição geral MAX-M, a última de nosso *ranking*, que penaliza a não realização de quaisquer morfemas, assinalando uma marca de violação a cada item não escandido. As restrições de realização de morfemas específicos estão em relação de estringência com essa restrição mais geral.

20. É fato que, em diversas línguas românicas, a ocorrência de codas silábicas é bastante restrita, e que se limita ainda mais no final de palavras. A maior incidência de epêntese vocálica no PB em posição final de vocábulo do que em outras posições (cf.: Collischonn 2002:210-211) e sua produtividade no âmbito dos estrangeirismos, como em [ˈpukɪ], para PUC, ou [ˈtɔpɪ] para *top*, é uma das principais evidências dessa exigência.

substantivos do PB quanto à realização de vogais terminais e sua relação com informação morfológica de GEN e CL em uma amostra de léxico dicionarizado e em uma amostra de uso do sul do Brasil – esta última contrastada quanto à frequência lexical a um corpus de referência. De modo geral, os resultados das duas primeiras amostras são muito coerentes, evidenciando o papel gramatical dessas marcas. Há, no mapeamento geral da língua, isomorfismo entre a terminação *o* e gênero masculino, de um lado, e entre a terminação *a* e o gênero feminino, de outro, com alguma vantagem de frequência de feminino. Essa correlação diz respeito, majoritariamente, a gênero gramatical, não sexo biológico ou gênero social. Se observados, contudo, subconjuntos de dados, como o dos poucos nomes que estabelecem pareamento com sexo biológico ou gênero social, e mesmo o dos biformes e comuns de dois na língua em uso, apesar de sua frequência também limitada, constata-se que o isomorfismo mencionado se mantém, e as formas masculinas passam a prevalecer – o que consideramos evidência adicional à tese mattosiana de masculino como forma não marcada em português. A análise de frequência lexical mostrou predomínio de formas de alta frequência em todas as categorias analisadas, sem sugerir qualquer privilégio, mas revelou alguma relevância no domínio do emprego das palavras fechadas pela vogal *e*.

Considerando a continuidade desta pesquisa, vislumbramos demandas para refinamento, como o tratamento de palavras fechadas por consoante, por ditongo nasal, das atemáticas em geral e das formas truncadas. Também aprimoramento da análise de frequência, considerando subconjuntos de dados, bem como experimentos envolvendo contextos específicos de concordância, podem trazer mais elementos à discussão aqui empreendida.

Em relação à escala de realização adotada e ao modelo analítico, sustentamos que hierarquizar CL e GEN enquanto categorias independentes permite dar conta de gramáticas mais e menos produtivas na língua. A análise de restrições ranqueadas (paralela e não modular), além disso, lida a um só tempo com restrições de naturezas e grandezas supostamente distintas, o que escapa a análises estritamente localistas. No caso do mapeamento de GEN e CL capta, por exemplo, intuições como a exigência de boa formação fonológica como reguladora da exponenciação; também consegue captar interações entre diferentes elementos de uma frase, como determinante e nome num NP, por exemplo, no que diz respeito à ordem de exponenciação. No domínio mais particular do modelo que adotamos, o da HS-OI, em que se admitem interações locais entre GEN e EVAL (paralelo-serial), pode-se lidar, entre outros aspectos, com o privilégio de determinados mapeamentos sobre outros, quando o que está em jogo são *outputs* intermediários mais do que efeitos de superfície. Além disso, a hipótese realizacional da morfologia garante uma interpretação coerente dos zeros alomórficos preconizados pelo estruturalismo (e não propriamente de *morfemas zero*). Uma demanda que se impõe a partir do uso desse modelo é a

de se estabelecer um diálogo com arquiteturas que incluam o processamento morfosintático, como a da Morfologia Distribuída, já que nessa abordagem teórica se interpreta o fenômeno da exponenciação fonológica de morfemas de modo similar.

Recebido em: 07/12/2016

Aprovado em: 07/07/2017

Email:

Luiz Carlos Schwindt [schwindt@ufrgs.br](mailto:schwindt@ufrgs.br)

## Referências

- ALCÂNTARA, Cíntia da C. 2010. As classes formais do PB. *Letras de Hoje*, v. 45, n. 1: 5-15.
- ARMELIN, Paula. 2014. Classifying Nominals in Brazilian Portuguese: a unified account for gender and inflectional class. In: Ludmila Veselovská; Markéta Janebová. (Eds.). *Complex Visible Out There: Language Use and Linguistic Structure*. 1 ed. Olomouc: Palacky University, p. 67-82.
- CAMARA JR., Joaquim M. 1970. *Estrutura da língua portuguesa*. Petrópolis: Vozes.
- COLLISCHONN, Gisela. 2002. A epêntese vocálica no português do sul do Brasil. In: BISOL, Leda; BRESCANCINI, Cláudia (Org.). *Fonologia e variação: recortes do português brasileiro*. Porto Alegre: EDIPUCRS, p. 205-230.
- \_\_\_\_\_.; SCHWINDT, Luiz Carlos. 2015. Uma coisa é uma coisa, outra coisa é outra coisa: por que a distinção entre gênero social e gramatical em língua portuguesa é necessária ao idioma. *Jornal Zero Hora*. 12/12/2015.
- CORBETT, Greville. 1991. *Gender*. Cambridge University Press.
- CORBIN, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tubinga: Max Niemeyer Verlag.
- KATAMBA, Francis; STONHAM, John. 2006. *Morphology*. New York: St. Martin's Press.
- HARRIS, James W. 1991. The exponence of gender in Spanish. *Linguistic Inquiry*, v. 22, n. 1: 27-62.
- \_\_\_\_\_. 1996. The syntax and morphology of class marker suppression in Spanish. In: ZAGONA, Karen. (Ed.). *Grammatical theory and romance languages*. Amsterdam: John Benjamins, p. 99-122.
- MATHEWS, Peter H. 1974 *Morphology: an introduction to the theory of word-structure*. Cambridge University Press.
- MCCARTHY, John J. 2000. Harmonic serialism and parallelism. In: HIROTANI, Masako. (Ed.). *Proceedings of the North East Linguistics Society 30*. Amherst, MA: GLSA Publications, p. 501-524.
- \_\_\_\_\_. 2002. *A thematic guide to Optimality Theory*. Cambridge University Press.

- \_\_\_\_\_. 2007. *Hidden generalizations: phonological opacity in Optimality Theory*. London: Equinox.
- \_\_\_\_\_. 2011. Pausal phonology and morpheme realization. In: BOROWSKY, Tony; KAWAHARA, Shigeto; SHINYA, Takahito; SUGAHARA, Mariko. *Prosody Matters: Essays in Honor of Lisa Selkirk*. London: Equinox.
- PRINCE, Alan.; SMOLENSKY, Paul. 2004[1993]. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA, and Oxford, UK: Blackwell. ms, Rutgers University e University of Colorado-Boulder.
- SCHWINDT, Luiz Carlos. 2011. Zeros na morfologia nominal portuguesa à luz da Optimal Interleaving Theory. *ReVEL*, **edição especial n. 5**. [www.revel.inf.br](http://www.revel.inf.br).
- \_\_\_\_\_; WETZELS, Willem Leo. 2016. **The morphology and phonology of inflection**. In: COSTA, João; MENUZZI, Sergio M.; WETZELS, Willem Leo. *Handbbook of Portuguese Linguistics*. Wiley-Blackwell, p. 188-209.
- VILLALVA, Alina. 2007. *Morfologia do português*. Lisboa: Universidade Aberta.
- WOLF, Matthew A. 2008. Optimal Interleaving: serial phonology-morphology interaction in a constraint-based model. *Doctoral Dissertation*. University of Massachusetts Amherst. ROA-996.