

A não uniformidade das bases numerais: uma visão dos contrastes das bases cardinais do português brasileiro

DOI: <http://dx.doi.org/10.21165/el.v54i1.3889>

Fernando Valls Yoshida¹

Resumo

As bases cardinais *-enta*, *-centos*, *mil* e *milhão*, *bilhão*, ..., organizam os numerais do português brasileiro em um sistema de base decimal. Apesar de sua função matemática comum, defendo que as bases cardinais não podem ser caracterizadas de maneira linguisticamente uniforme. Ao contrário, este artigo sistematiza contrastes entre elas nos domínios sintático (possibilidade de elipse e distribuição com relação a quantificadores), categorial (comportamento de concordância em gênero e número e licenciamento de construção pseudopartitiva) e morfológico (diferenças em estruturas morfológicas e graus de produtividade). A não uniformidade das bases numerais aqui defendida tem impactos, de um lado, na medida em que a Faculdade da Linguagem é sensível à caracterização matemática das bases e, de outro, na diversidade gramatical inerente a esses objetos no português brasileiro.

Palavras-chave: Numerais cardinais; bases numerais; Morfologia; Sintaxe.

¹ Universidade de São Paulo (USP), São Paulo, São Paulo, Brasil; fevayo@usp.br; <https://orcid.org/0000-0002-4400-4869>

The non-uniformity of numeral bases: insights from contrasts of Brazilian Portuguese cardinal bases

Abstract

The cardinal bases *-enta* ('-ty'), *cem* ('hundred'), *mil* ('thousand') and *milhão* ('million'), *bilhão* ('billion'), ..., organize Brazilian Portuguese numerals into a decimal numeral system. Despite their common mathematical function, I argue that cardinal bases cannot be characterized in a linguistically uniform manner. Rather, this paper systematizes contrasts between them at the levels of syntax (possibility of ellipsis and distribution with respect to quantifiers), category (gender and number agreement behavior and licensing of pseudopartitive constructions) and morphology (differences in morphologic structure and degrees of productivity). The non-uniformity of numeral bases defended here has impacts both on the extent to which the Faculty of Language is sensitive to the mathematical characterization of bases and on the inherent grammatical diversity of these objects in Brazilian Portuguese.

Keywords: Cardinal numerals; numeral bases; Morphology; Syntax.

Introdução

Sistemas numerais linguísticos são frequentemente descritos e analisados em termos de bases. Em (1), os diferentes numerais cardinais² para o número 91 ilustram, a partir do contraste entre o valor numérico das bases *juu* 'dez' (1a), *kūrū* 'doze' (1b) e *vingt* 'vinte' (1c), como a escolha de uma base afeta a organização de um sistema numeral.

1. a. kyuu ju ichi [Japonês (japônica): Kaiser *et al.* (2013)]
 nove dez um
 "nove dezenas e um (≈ noventa e um)"

b. bā-kūru bā-tāmà ná bī-tāmà [Berom (benue-congolesa): Comrie (2022)]
 PL-doze PL-sete mais CL-sete
 "sete dúzias e sete (≈ noventa e um)"

c. quatre-vingt-onze [Francês (românica)]
 quattro-vinte-onze
 "quattro vintenas e onze (≈ noventa e um)"

² São de interesse deste artigo apenas os numerais ditos cardinais (e.g., *três*), ainda que o espírito de análise aqui defendido possa ser extrapolado para os numerais ordinais (e.g., *terceiro*), multiplicativos (e.g., *triplo*), fracionários (e.g., *terço*) etc.

Motivado por esse papel estruturante das bases, Hurford (1975), um dos pioneiros na teoria gramatical de numerais linguísticos, reservou a elas uma projeção especial na sintaxe: M. Em seu modelo, regras de estrutura frasal (2) geram estruturas profundas (3a), a partir das quais são derivadas representações fonológicas (3b) e semânticas (3c). Articulado ao projeto da Teoria Padrão (Chomsky, 1965), Hurford (1975) tem interesse não só em capturar generalizações descritivas de sistemas numerais, mas em investigar os universais cognitivos que fundamentam a capacidade humana para representar números por meio da linguagem.

12. Regras de estrutura frasal dos numerais do inglês

- | | | |
|--------|---|--|
| NUMBER | → | $\{1, \text{PHRASE}\} (\text{NUMBER})$ |
| PHRASE | → | NUMBER M |
| M | → | $\{10, \text{NUMBER M}\}$ |
- a. Estrutura profunda: $[\text{NUMBER} [\text{PHRASE} [\text{NUMBER} 1 [\text{NUMBER} 1] [\text{NUMBER} 10] [\text{NUMBER} 1]]]$
- b. Forma fonológica: /twəntiwan/
- c. Forma semântica: $((1 + 1) \times 10) + 1 = 21$

A motivação inaugural de Hurford (1975) foi atualizada e reinterpretada à luz de diversos desdobramentos teóricos: a Teoria de Regência e Ligação (Ionin; Matushansky, 2006, 2018), a Gramática de Construções (Booij, 2010), o Minimalismo (Di Sciullo, 2015, 2022), a Nanossintaxe (Wągiel; Caha, 2020; Wągiel, 2023), entre outros. No âmbito desses desenvolvimentos, parte importante do debate se concentra em determinar a configuração sintática de construções *[numeral nome]*: se o numeral é um núcleo que toma o NP quantificado como complemento (4a), ou se é especificador de um núcleo na projeção funcional do NP quantificado (4b).

4. Debate sobre a estrutura de construções *[numeral nome]*

- a. Hipótese da estrutura em cascata: $[\text{XP} \text{ numeral} [\text{NP} \text{ nome}]]$
- b. Hipótese do especificador: $[\text{YP} \text{ numeral} [\text{Y} [\text{NP} \text{ nome}]]]$

A respeito da maneira com que o embate entre essas propostas tem se desenvolvido, dois são os principais expedientes. Por um lado, evidências sintáticas, semânticas e morfológicas são acumuladas para cada uma das hipóteses, de modo mutuamente exclusivo (ver Corver e Tatsumi, 2023, para uma síntese dos argumentos). Por outro, há a tentativa de reduzir a hipótese adversária como sendo uma mera consequência da hipótese defendida, sendo para isso propostas condições adicionais que derivam propriedades superficiais da estrutura de especificador a partir de uma estrutura fundamental em cascata (cf. Ionin; Matushansky, 2018), ou vice-versa (cf. Witkoś; Dziubała-Szrejbrowska, 2018).

De maneira a conciliar as duas perspectivas em conflito, um terceiro caminho tem sido avançado por parte da literatura recente (Danon, 2012; Klockmann, 2021; Tatsumi, 2021; Corver; Tatsumi, 2023). Essa posição da literatura identifica, apesar dos pontos de disputa, um importante consenso: a hipótese de que o comportamento formal de numerais é universal e de que há uma única configuração estrutural para as construções *[numeral nome]*. Essa assunção, que ecoa Hurford (1975, p. 4), é aqui formulada nos termos de (5). A intuição relevante é a de que é somente sob o pressuposto de que há uma única estrutura, uniforme e universal, que se coloca o debate entre diferentes implementações para uma (suposta) mesma construção.

3. **Hipótese de Uniformidade** A Faculdade da Linguagem opera de maneira uniforme no que diz respeito (i) aos numerais de uma língua e (ii) aos numerais das diferentes línguas naturais humanas.

Nesse sentido, o caminho sugerido pelos autores dessa tradição conciliadora consiste em, de um lado, admitir o conjunto de evidências acumulado de maneira mutuamente exclusiva para cada uma das hipóteses em (4) e, de outro, acomodar a convivência entre as análises adversárias a partir da renúncia da Hipótese de Uniformidade (5). Assim, as análises são diferentes justamente porque os fenômenos são diferentes, admitindo que a Faculdade da Linguagem acomoda tal diversidade. Sob essa perspectiva, perguntas importantes dizem respeito à determinação do repertório translingüístico de estruturas *[numeral nome]* possíveis (Tatsumi, 2021), à maneira com que diferentes configurações podem conviver em uma mesma língua (Danon, 2012; Corver; Tatsumi, 2023) e à possibilidade de incorporação ou perda de certos tipos de construção ao longo da história de uma língua (Klockmann, 2021).

Este artigo avança essa tradição emergente em dois sentidos relevantes. Em primeiro lugar, a partir de um recorte voltado a bases numerais, argumento em favor de uma versão mais forte da negação da Hipótese de Uniformidade (5): não só não é universal a maneira com que numerais em geral são manipulados, mas também a caracterização formal de bases numerais, em particular, é heterogênea (ver Klockmann, 2021, para uma posição semelhante, defendida a partir do inglês e do polonês). Em segundo lugar, assumo neste artigo uma visão centrada no português brasileiro (doravante, PB), com o objetivo de assim contribuir à discussão em torno da Hipótese de Uniformidade (5) com dados de uma língua pouco representada nessa tradição.³

Para tanto, este trabalho explora contrastes entre as bases do PB na interface morfologia-sintaxe, avaliando as maneiras com que elas divergem do ponto de vista sintático, categorial e morfológico. A imagem que emerge dessa comparação tensiona a Hipótese

3 Para uma análise neo-hurfordeana do sistema numeral do PB, ver Vignado (2017, 2018, 2019, 2020).

de Uniformidade (5) e sugere sua abdicação em favor de uma caracterização heterogênea das bases do PB.

Contrastes entre as bases numerais do PB

Contrastes sintáticos

Um ponto importante do debate sobre a (não) uniformidade da representação linguística das bases numerais diz respeito à sua distribuição sintática. Nesse sentido, o sistema cardinal do PB, em sua heterogeneidade, se mostra particularmente informativo. Isso porque, contrariamente ao que prevê a Hipótese de Uniformidade (5), testes de constituinte e de distribuição complementar com quantificadores sugerem que há pelo menos dois padrões de comportamento sintático para bases.

A respeito da distribuição de bases em construções de forma *[unidade base nome]*, uma primeira questão a ser levantada é se a base forma com o nome uma projeção máxima independente da unidade, isto é, *[unidade [_{XP} base nome]]*. Para tanto, o teste da elipse, tradicionalmente assumido como tendo por alvo necessário projeções máximas (cf. Merchant, 1999), é especialmente conveniente. Aplicando-o em construções contrastivas como (6), a diferença entre a agramaticalidade da elipse em (6a–b) e sua gramaticalidade em (6c–d) sugere que apenas *mil* e os milhares altos (i.e., *milhão*, *bilhão* etc.) formam projeções máximas com os nomes que quantificam. Sendo esse o caso, as bases do PB diferem minimamente no constituinte que ocupam.

6. O preço foi...
 - a. *... de oito[~~enta reais~~] para noventa reais.
 - b. *... de oito[~~centos reais~~] para novecentos reais.
 - c. ... de oito [~~mil reais~~] para nove mil reais.
 - d. ... de oito [~~milhões de reais~~] para nove milhões de reais.

A pergunta que naturalmente se segue é se as bases *-enta* e *-centos*, não formando com o nome um constituinte à parte, o fazem com as unidades cardinais, isto é, *[[__{ZP} unidade base] nome]*. Para avaliar essa hipótese, consideremos quantificadores como *muitos*, *vários* e *tantos*, que, como sugere (7), se distribuem de maneira complementar com as unidades cardinais. Nesses termos, se esses quantificadores se distribuem como as unidades cardinais e, por hipótese, se essas unidades ocupam os mesmos constituintes que *-enta* e *-centos* (i.e., ZP), então devemos observar a distribuição complementar desses quantificadores também com relação a essas bases.

7. a. {duas/muitas/várias/tantas/...} pessoas
b. *{muitas/várias/tantas/...} duas pessoas
c. *duas {muitas/várias/tantas/...} pessoas

Confirmando essa previsão, os dados em (8) sugerem que as bases de fato diferem com relação à coocorrência com esses quantificadores. Por um lado, a coocorrência não é licenciada para *-enta* (8a) ou *-centos* (8b), o que aproxima sua distribuição com a das unidades cardinais (7b). Por outro lado, os quantificadores coocorrem com *mil* (8c) e com os milhares altos (8d), reforçando a hipótese de que essas bases formam um constituinte independente das unidades cardinais – e, por extensão, dos quantificadores sob investigação.⁴

8. a. *{muitos/vários/tantos/...} enta linguistas
b. *{muitos/vários/tantos/...} centos linguistas
c. {muitos/vários/tantos/...} mil linguistas
d. {muitos/vários/tantos/...} milhões de linguistas

Outro quantificador relevante para a discussão é o pronome *qu- quantos*. Uma propriedade de interesse é que, quando substituindo unidades cardinais, o pronome admite a realização do restante do NP quantificado, o que inclui desde o núcleo do NP (9) a APs (10) e CPs (11) adjuntos. Nesse sentido, a substituição por *quantos* é um diagnóstico para a posição das unidades cardinais: apenas elementos externos a essa posição podem ser realizados junto a *quantos*.

9. a. Ela tem [três] [_{NP} livros].
b. Ela tem [quantos] [_{NP} livros]?
10. a. Ela tem [três] livros [_{AP} raros].
b. Ela tem [quantos] livros [_{AP} raros]?
11. a. Ela tem [três] livros [_{CP} que foram importados da Bulgária].
b. Ela tem [quantos] livros [_{CP} que foram importados da Bulgária]?

Dessa forma, a hipótese de que *-enta* e *-centos* formam um constituinte independente do nome com as unidades cardinais prevê que essas bases não podem coocorrer com

4 Por completude, cabe notar que a ordem [base quantificador nome], à semelhança de (7c), é igualmente rejeitada por todas as bases (i–iv).

- (i) *enta {muitos/vários/tantos/...} linguistas
- (ii) *centos {muitos/vários/tantos/...} linguistas
- (iii) *mil {muitos/vários/tantos/...} linguistas
- (iv) *milhões (de) {muitos/vários/tantos/...} linguistas

quantos à maneira de (9–11). Em termos da estrutura genérica $[[_{\text{ZP}} \text{unidade base}] \text{ nome}]$, isso é dizer, da parte do pronome, que *quantos* substitui ZP e, da parte das bases, que *-enta* e *-centos*, mas não *mil* e os milhares altos, ocupam ZP.

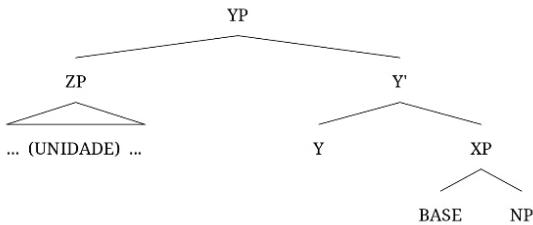
Consideremos, portanto, os diálogos em (12). O padrão de (a)gramaticalidade nas perguntas de B corresponde às predições derivadas estruturalmente: *-enta* (12a) e *-centos* (12b), ocupando ZP, não coocorrem com *quantos*, ao contrário de *mil* (12c) e dos milhares altos (12d), que, externos a ZP, são licenciados em contexto do pronome.

12. a. A: O tonto do meu primo pagou oitenta reais naquele lápis.
B: *Quantos enta reais?
A: *Oito[enta reais].
- b. A: O tonto do meu primo pagou oitocentos reais naquela camiseta.
B: *Quantos centos reais?
A: *Oito[eentos reais].
- c. A: O tonto do meu primo pagou oito mil pesos naquela empanada.
B: Quantos mil pesos?
A: Oito [mil pesos].
- d. A: O tonto do meu primo pagou oito milhões de pesos naquele vinho.
B: Quantos milhões de pesos?
A: Oito [milhões de pesos].

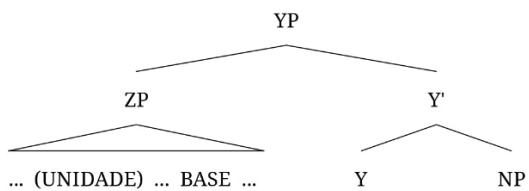
Ainda a partir desses diálogos, notemos que as respostas de A evocam o padrão dos contrastes com relação à elipse em (6), o que sugere uma explicação unificada para os dados levantados nesta subseção. Especificamente, ser alvo da substituição por *quantos* (i.e., integrar ZP) implica não ser alvo de elipse (i.e., integrar XP), e vice-versa. A interpretação sintática dessa oposição é que as bases do PB ocupam dois *loci* sintáticos distintos.

Em suma, os dados aqui levantados sugerem que há dois padrões para o comportamento sintático das bases do PB: bases como *mil* e os milhares altos são alvos de elipse conjunta com o nome e coocorrem com *muitos/vários/tantos/quantos/...*, ao passo que bases como *-enta* e *-centos* não são elididas junto ao nome e se distribuem de modo complementar com *muitos/vários/tantos/quantos/...*. Esses contrastes podem ser acomodados renunciando à Hipótese de Uniformidade (5) e admitindo duas representações sintáticas distintas para as construções $[\text{unidade base nome}]$ do PB (13).

13. Dois *loci* sintáticos para as bases do PB
 - a. Em constituição com o nome: *mil* e os milhares altos



b. Em constituição com as unidades: *-enta* e *-centos*



Essas representações capturam a assimetria das bases com relação à sua distribuição na sintaxe, mas são indeterminadas no que diz respeito aos rótulos das projeções X, Y e Z. Para a projeção Y, há um debate paralelo aos interesses deste artigo concentrado em decidir Y como sendo Num, Card ou Q (ver Ionin e Matushansky, 2018, para uma revisão das hipóteses). O mesmo problema se estende à caracterização de X e Z. Para estes, todavia, há um maior consenso de sua identificação com categorias lexicais, o que é evidenciado pelo grande aporte de evidências que indicam que numerais têm o comportamento de partes do discurso típicas, como nomes e adjetivos (cf. Corver; Tatsumi, 2023).

Assumindo, portanto, que X e Z projetam categorias morfossintáticas como N e A, a próxima seção reflete sobre a uniformidade das bases no que se refere ao seu comportamento categorial.

Contrastes categoriais

Comecemos comparando o comportamento categorial de *mil* e dos milhares altos. Por um lado, os numerais *milhão/bilhão/...* são caracteristicamente nominais (Neves, 2011; Vignado, 2019; Valls Yoshida, 2024). Como nomes, requerem complemento introduzido por *de* (14a), em construções pseudopartitivas (cf. Tatsumi, 2021) e controlam concordância em gênero (15a) e em número gramatical (16a) independentemente dos traços do núcleo do NP quantificado. Por outro lado, *mil* não apresenta nenhum dos indicativos de comportamento categorial que identificam os milhares altos como nomes (14–16b).

14. a. milhões *(de) habitantes
- b. mil (*de) habitantes
15. a. um_[USG] milhão_[SG] de pessoas_[PL]
- b. *um_[USG] mil_[SG] pessoas_[PL]
16. a. dois_[UMASC] milhões_[MASC] de pessoas_[FEM]
- b. *dois_[UMASC] mil_[MASC] pessoas_[FEM]

Não se comportando como nome, é possível questionar se *mil* pode ser caracterizado como adjetivo. Do ponto de vista morfológico, *mil* não realiza de maneira transparente traços de gênero ou número gramatical, o que torna opaco seu perfil morfossintático. Do ponto de vista distribucional, todavia, *mil* admite, mantendo a leitura cardinal,⁵ ambas as posições pré-nominal e pós-nominal (17), sendo esta última canonicamente reservada a adjetivos predicativos.

17. a. mil [_{NP} tradições e glórias]
[_{NP} tradições e glórias] mil
- b. mil [_{NP} problemas]
[_{NP} problemas] mil

A comparação entre *mil* e os milhares altos indica que, ainda que ambos estejam sob uma mesma projeção máxima com o NP quantificado, essas bases não projetam, elas próprias, uma mesma categoria morfossintática: os milhares altos são nomes e *mil* é um adjetivo (i.e., [_{NP} base_{N/A} NP]). A essa altura, é interessante observar que se a hipótese proposta defende que as diferenças de (14–16) são essencialmente categoriais, é de se esperar que uma contraparte nominal de *mil* licencie sua ocorrência nos mesmos contextos dos milhares altos. Com efeito, o comportamento sugestivamente nominal de *milhar* (18–20), derivado de *mil*, corrobora essa hipótese.⁶

5 Faz-se necessário distinguir as leituras cardinais em (17), de interesse para o argumento, daquela que Wiese (2003) identifica como a leitura nominal de cardinais (e.g., *camisa nove*, que se lê de maneira distinta de *nove camisas*).

6 Caso análogo pode ser elaborado para *dez*, *doze* e *cem*, que se distribuem como *mil* (i–iiib), mas cujas contrapartes *dezena*, *dúzia* e *centena*, como os milhares altos, se distribuem como nomes (i–iiia). Ao todo, esses dados reforçam que é uma diferença entre nomes e adjetivos aquela que está por trás desses contrastes.

- (i) a. {dezenas/dúzias/centenas} *(de) habitantes
- b. {dez/doze/cem} (*de) habitantes
- (ii) a. uma {dezena/dúzia/centena} de habitantes
- b. *uma {dez/doze/cem} habitantes
- (iii) a. duas {dezenas/dúzias/centenas} de habitantes
- b. *duas {dez/doze/cem} habitantes

18. milhares *(de) habitantes
19. um_[USG] milhar_[SG] de pessoas_[PL]
20. dois_[UMASC] milhares_[MASC] de pessoas_[FEM]

Em oposição aos milhares altos e à semelhança de *mil*, os numerais derivados a partir de *-centos* não manifestam qualquer propriedade sugestivamente nominal, a começar pela rejeição da construção pseudopartitiva (21). No que se refere à concordância em gênero, porém, esses numerais realizam abertamente marcas de concordância com o nome (22), em contraste com *mil*. Também em contraste com o comportamento de *mil* em (17), as centenas rejeitam a posição pós-nominal (23a), assim como os outros quantificadores com os quais se distribuem de modo complementar (23b). À luz dessas observações, os numerais formados por *-centos* e *mil* são sugestivamente ambos adjetivos, mas não à mesma maneira, algo que parece refletir o caso geral da diversidade distribucional dos adjetivos do PB (Müller; Negrão; Nunes-Pemberton, 2002; Quadros Gomes; Gonçalves Sudré, 2021; *i.a.*).⁷

21. oitocentas (*de) pessoas
22. a. oitocentas_[UFEM] pessoas_[FEM]
 - b. oitocentos_[UMASC] habitantes_[MASC]
23. a. *pessoas oitocentas
 - b. *pessoas {muitas/várias/tantas/...}

Os contrastes levantados até então nos permitem avaliar, finalmente, o comportamento categorial de *-enta*. Especificamente, todos os indícios categoriais aqui considerados testam negativo para *-enta*: os numerais derivados dessa base não tomam complementos nominais introduzidos por *de* (24), não admitem modificação cardinal para que se teste determinação de gênero e número gramatical inerentes (25), não realizam marcas de concordância com o nome (26) e não ocorrem em posição pós-nominal (27). A resistência de *-enta* a revelar pistas de sua natureza categorial é um desafio em aberto (cf. Valls Yoshida, 2024) e, para os efeitos deste artigo, contrasta essa com todas as demais bases, para as quais ao menos alguma evidência positiva está disponível.

24. oitenta *(de) habitantes
25. *dois/duas oitenta habitantes

7 Como bem nos aponta um(a) parecerista anônimo(a), a identificação unificada de *-centos* e *mil* como bases adjetivas mascara, por um lado, as diferenças dessas bases com relação a adjetivos protótipicos e, por outro, os contrastes entre elas enquanto numerais adjetivais (cf. admissão da posição pós-nominal para *mil*, mas não para formas derivadas de *-centos*). Um desdobramento previsto da proposta aqui apresentada é o refinamento da sintaxe das bases adjetivais, tendo em vista ainda seus efeitos semânticos para a interpretação de sintagmas nominais quantificados por numerais (cf. nota de rodapé 9).

26. oitenta/*oitento/*oitentas/*oitentos {pessoas/habitantes}
27. *pessoas oitenta

No todo, os dados mobilizados nesta subseção atribuem às bases uma dimensão de flexibilidade a mais, adicionando à diversidade sintática uma camada de heterogeneidade categorial: os milhares altos são nomes; *mil* e os derivados de *-centos* são adjetivos, ainda que não à mesma maneira; e os derivados de *-enta* são categorialmente evasivos e nesse sentido contrastam com todas as demais bases.

Tomando o caminho que parte da estrutura frasal e avança ao domínio da palavra, a próxima subseção se dedica a explorar a composição morfológica das bases do PB.

Contrastes morfológicos

Parte das assimetrias exploradas na subseção *Contrastes sintáticos* naturalmente se relaciona à diferença das bases em sua autonomia morfológica com relação às unidades. Bases que são morfemas livres, como *mil* e os milhares altos, são licenciados em contextos em que *-enta* e *-centos*, sendo morfemas presos, não o são. É importante notar, mesmo assim, que as bases presas não são igualmente presas e contrastes entre elas revelam estruturas morfológicas particularmente distintas.

Uma primeira aproximação a essa diferença pode ser feita em termos de constituintes prosódicos (cf. Nespor; Vogel, 1986): as dezenas são formadas por uma única palavra fonológica (28a), ao passo que as centenas, por duas (28b). Essa distinção é um primeiro indicativo de que *-enta* e *-centos* são objetos morfológicos de tamanhos distintos, e que não se combinam às unidades cardinais segundo uma mesma estrutura.

28. [sete]_ω
 - a. [setenta]_ω
 - b. [sete]_ω[centos]_ω

Outro indicativo desse contraste diz respeito ao tipo de estrutura morfolonológica com que se combinam. A esse respeito, as unidades cardinais do PB são radicais complexos formados a partir da combinação do expoente de uma raiz a uma vogal temática (cf. Valls Yoshida, 2024). Crucialmente, as bases sob investigação são sensíveis a essa diferença: *-enta* se combina diretamente com os expoentes de raízes cardinais (29a), ao passo que *-centos* exige radicais marcados tematicamente (29b). O argumento para o cardinal *oit-o* em (29), de terminação *-o* (ou de Classe I, nos termos de Alcântara, 2010), pode ser extrapolado aos cardinais *quatr-o* (também de Classe I) e a *set-e* e *nov-e* (de Classe III).

29. [radical [raiz oit]-o]
a. [raiz oit]-enta
b. [radical oito]-centos

Também são contrastivos processos fonológicos sensíveis à localidade com a raiz, como a neutralização vocálica,⁸ condicionada dialetalmente no PB. O exemplo (30) ilustra como o expoente *n/O/ve* admite variação na aplicação da neutralização vocálica em contexto de *-enta* (30a), mas não em contexto de *-centos* (30b). A impossibilidade categórica de *n[o]vecentos* se alinha às demais assimetrias morfológicas: *-centos* se combina a uma porção maior de estrutura – *i.e.*, um radical (29) que forma uma palavra fonológica completa (28) – e por isso não guarda com o expoente de *nove* relação local o suficiente para engatilhar a neutralização vocálica. O mesmo parece estar por trás do padrão de variação em *s[ε/*e]te* ~ *s[ε/e]tenta* ~ *s[ε/*e]tecentos*.

30. *n[ɔ]ve, *n[o]ve*
a. *n[ɔ]venta, n[o]venta*
b. *n[ɔ]vecentos, *n[o]vecentos*

Em suma, as evidências podem ser capturadas assumindo que essas bases engajam em dois processos de formação de palavras distintos: *-enta* é um sufixo derivacional e *-centos* forma com as unidades cardinais compostos morfológicos. Para os propósitos deste artigo, isso significa que tampouco sob a lente da morfologia as bases podem ser identificadas de maneira homogênea, advogando contra a Hipótese de Uniformidade (5).

É importante frisar que a oposição das bases em morfemas livres e morfemas presos não necessariamente se reduz a uma diferença em complexidade morfológica: afinal, os milhares altos, morfemas livres, são morfologicamente complexos. Sob uma perspectiva diacrônica, o numeral *milhão* é uma inovação românica a partir da maior das bases latinas *mille* 'mil'. A esse respeito, o cognato histórico para *milhão* no PB foi primeiro atestado no francês antigo, em 1270 (Bloch; von Wartburg, 1962 [1932]), e é por sua vez um possível empréstimo do italiano antigo, analisado composicionalmente como *mili-one*, lit. 'mil-AUG', 'milhão' (Dressler; Merlini Barbaresi, 1994). A derivação de bases altas a partir de bases menores é um expediente regular translinguisticamente (Comrie, 2022).

Uma vez fixado, o numeral *milhão* foi reanalisado (Hurford, 1975; Bauer, 2021) e o morfema *-ilhão* é ativo na derivação da sequência completa dos milhares altos do PB (31). A esse respeito, é especialmente interessante notar que *-ilhão* se combina a expoentes de uma

8 Para uma perspectiva morfológica deste fenômeno, ver Lee (1995), Schwindt (2013), Ulrich e Schwindt (2018), entre outros.

variedade de tipos de numerais (32), o que complexifica a composição morfológica das bases altas do PB.⁹

31. milhão, bilhão, trilhão, quatrilhão, quintilhão, sextilhão, septilhão, octilhão, nonilhão, decilhão
32. a. Formas greco-latinas
bi- (e.g., *bissexual*), *tri-* (e.g., *tricampeão*), *oct-* (e.g., *octógono*), *dec-* (e.g., *década*)
b. Formas cardinais
quatr- (e.g., *quatro*)
c. Formas ordinais
quint- (e.g., *quinto*), *sext-* (e.g., *sextº*), *sept-* (e.g., *sétimo*), *non-* (e.g., *nono*)

Retomando o contraste entre as bases do PB, as evidências morfológicas aqui consideradas sugerem que elas se realizam em termos de expedientes de formação de numerais distintos, listados explicitamente em (33).

33. Expedientes derivacionais a partir de bases no PB
 - a. Sufixação por *-enta*
trinta, quarenta, cinquenta, sessenta, setenta, oitenta, noventa
 - b. Composição com *-centos*
duzentos, duzentas, trezentos, trezentas, quatrocentos, quatrocentas, seiscientos, seiscentas, setecientos, setecentas, oitocentos, oitocentas, novecentos, novecentas
 - c. Derivação de *-ilhão*
milhão, milhões, bilhão, bilhões, trilhão, trilhões, quatrilhão, quatrilhões, quintilhão, quintilhões, sextilhão, sextilhões, septilhão, septilhões, octilhão, octilhões, nonilhão, nonilhões, decilhão, decilhões

Uma vez identificados, esses diferentes processos de formação de palavras podem ser, ainda, contrastados no que diz respeito à sua produtividade morfológica (Quadro 1), um fenômeno complexo, que pode ser abordado sob diferentes vieses, formais e quantitativos (cf. Bauer, 2001). No que se refere à regularidade desses processos (Aronoff, 1976; Basílio, 1977), as bases *-enta*, seguida de *-centos*, são as que mais acumulam idiossincrasias na forma da unidade cardinal com que se combinam. Quanto à formação de neologismos, apenas *-centos* e *-ilhão* derivam numerais inovadores, cuja interpretação é vaga quanto ao seu valor numérico. Ainda, sob uma perspectiva quantitativa para a produtividade (Corbin, 1987; Yang, 2022), as bases *-enta* e *-centos* empatam no número de cardinais canônicos (i.e., excetuando-se os inovadores) com que se combinam, atrás do número

9 A diversidade do paradigma (31) no que se refere ao tipo de morfe com que *-ilhão* se combina não é uma exclusividade dos milhares altos, e se atesta no PB também na composição morfológica dos numerais ordinais, multiplicativos e fracionários. Ver Greenberg (2000) para uma discussão completa.

maior de *types* a que se aplica *-ilhão*. Quanto ao número de *tokens*, a média da frequência lexical no *corpus* LexPorBR (Estivalet; Meunier, 2017) sugere ainda uma nova graduação de produtividade das bases, em que *-ilhão* é o mais produtivo, uma ordem de grandeza à frente de *-enta*, seguido finalmente por *-centos*.

Quadro 1. Contrastos de produtividade entre as bases do PB

Base	Regularidade	Numerais vagos	Frequência (<i>types</i>)	Frequência média (<i>tokens</i>)
<i>-enta</i>	Parcial: exceções em <i>trinta, quarenta</i> e <i>sessenta</i> .	Não.	> 7	111,53
<i>-centos</i>	Parcial: exceções em <i>duzentos</i> e <i>trezentos</i> .	Sim: <i>trocentos, bucentos, ...</i>	> 7	15,466
<i>-ilhão</i>	Parcial: exceção em <i>milhão</i> .	Sim: <i>gazilhão, zilhão, ...</i>	> 10	2532,30

Fonte: Elaboração própria

Ao todo, as evidências mobilizadas nesta subseção evidenciam a riqueza morfológica do PB na representação de suas bases cardinais. As bases *-enta* e *-centos* são morfemas presos e se distinguem na medida em que, na relação com as unidades cardinais, um é um sufixo derivacional e o outro forma com elas compostos morfológicos. Bases que são morfemas livres podem ser morfologicamente simples, como *mil*, mas não necessariamente: os milhares altos são derivados a partir do morfema preso *-ilhão*. Tomados ao todo, os processos de formação de cardinais do PB são diversos morfologicamente e contrastam ainda em produtividade, o que se observa na (ir)regularidade morfofonológica, na (im) possibilidade de derivar numerais inovadores e nas diferenças em frequência de *types* e *tokens*.

Uma vez mais, as bases cardinais do PB resistem a uma análise unificada. A diversidade formal com que se revelam põe sérias dificuldades para uma análise comprometida com sua substância linguística, mais que sua identidade no papel que exercem na estruturação do sistema cardinal do PB.

Considerações finais

A hipótese de que a Faculdade da Linguagem é diretamente sensível à estrutura matemática dos sistemas numerais abriu uma extensa e produtiva tradição de estudos linguísticos. Para a questão das bases numerais, em particular, isso permitiu à tipologia, por um lado, abstrair as diferenças entre as línguas e generalizar a regularidade da expressão de sistemas numerais em torno de bases (Greenberg, 1978, 2000; Comrie, 2013, 2022) e aos estudos gramaticais, por outro, refletir sobre o lugar privilegiado

que bases ocupam na derivação de numerais complexos (Hurford, 1975; Booij, 2010; Rothstein, 2013).

Se a noção de base, assim concebida, é produtiva a um nível apropriado de abstração da substância linguística, uma investigação voltada às suas propriedades genuinamente gramaticais revela uma caracterização heterogênea das bases numerais. Acrescentando a uma tradição não-uniformista dos numerais linguísticos (Danon, 2012; Klockmann, 2021; Tatsumi, 2021; Corver; Tatsumi, 2023), este artigo traz à discussão a empiria do PB e com isso revela a riqueza gramatical dos expedientes que representam as bases do sistema cardinal dessa língua. Especificamente, os argumentos delineados neste artigo apontam para dois *loci* que bases podem ocupar na sintaxe: ou em constituição com o nome quantificado, ou com as unidades cardinais que as modificam. Em cada um desses *loci*, as bases do PB variam em categoria morfossintática (viz., nomes e adjetivos) e nas propriedades típicas de cada categoria que manifestam. Não só quanto à morfologia flexional as bases divergem: no que se refere à sua diversidade derivacional, as bases não parecem ser codificadas por um expediente morfológico unificado, e se revelam em sufixos derivacionais e em compostos, com graus diferentes de produtividade.

Ao longo da discussão aqui desenvolvida, o aprofundamento da caracterização gramatical das bases numerais do PB é reconhecidamente limitado à medida em que é suficiente para contrastar as bases nos níveis sintático, categorial e morfológico. Tal posição é fundamentada na decisão de concentrar aqui os esforços na empreitada de desafiar a Hipótese de Uniformidade (5). Com sorte, tendo sido esse objetivo satisfatoriamente atingido, são esperados na continuidade deste trabalho desenvolvimentos de uma caracterização pormenorizada do comportamento gramatical das bases do PB para que, com isso, se possa apreciar seu lugar na Gramática com maior precisão.

Agradecimentos

Agradeço à prof. Dr^a Ana Paula Scher pelos debates precisos que acompanharam o desenvolvimento deste artigo, à prof. Dr^a Indaiá de Santana Bassani pelas ricas sugestões que inspiraram a estrutura e a redação do trabalho e à Marcela Nunes Costa pelo olhar sempre atento e as críticas sempre tão construtivas que em muito contribuíram à realização deste artigo. Todos os equívocos restantes são de minha responsabilidade. Pelo financiamento de pesquisa, agradeço à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP – processo #2023/17997-0).

Referências

ALCÂNTARA, C. As classes formais do português brasileiro. *Letras de Hoje*, v. 45, n. 1, p. 5-15, 2010.

- ARONOFF, M. *Word Formation in generative grammar*. Cambridge: MIT Press, 1976.
- BASÍLIO, M. *Aspects of the structure of lexicon: evidence from Portuguese*. 1977. Tese (Doutorado em Linguística) – University of Texas at Austin, Austin, 1977.
- BAUER, B. Counting systems. In: LEDGEWAY, A.; MAIDEN, M. (ed.). *The Cambridge Handbook of Romance Linguistics*. Cambridge: Cambridge University Press, 2022.
- BAUER, B. Formation of Numerals in the Romance Languages. In: *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press, 2021.
- BAUER, L. *Morphological productivity*. Cambridge: Cambridge University Press, 2001.
- BLOCH, O.; VON WARTBURG, W. *Dictionnaire étymologique de la langue française*. 11. ed. Paris: Presses Universitaires de France, 1962. (Originalmente publicado em 1932.)
- BOOIJ, G. Constructions and lexical units: An analysis of Dutch numerals. *New impulses in word formation*, p. 81-100, 2010.
- BRESNAN, J.; MCHOMBO, S. The lexical integrity principle: Evidence from Bantu. *Natural Language & Linguistic Theory*, v. 13, n. 2, p. 181-254, 1995.
- CHOMSKY, N. *Aspects of the Theory of Syntax*. Cambridge: MIT Press, 1965.
- COMRIE, B. Numeral bases. In: DRYER, M.; HASPELMATH, M. (ed.). *The world atlas of language structures online*. 2013.
- COMRIE, B. The Arithmetic of Natural Language: Toward a typology of numeral systems. *Macrolinguistics*, v. 10, n. 1, p. 1-35, 2022.
- CORBIN, D. *Morphologie dérivationnelle et structuration du lexique*. Tübingen: Niemeyer, 1987.
- CORVER, N.; TATSUMI, Y. Number names: Internal structure and morphological marking. In: ACKEMA, P.; BENDJABALLAH, S.; BONET, E.; FÁBREGAS, A. (ed.). *The Wiley Blackwell Companion to Morphology*. Wiley-Blackwell, 2023.
- DANON, G. Two structures for numeral-noun constructions. *Lingua*, v. 122, n. 12, p. 1282-1307, 2012.

DI SCIULLO, A. Complex Cardinal Numerals and the Strong Minimalist Thesis. *Philosophies*, v. 7, n. 4, p. 81, 2022.

DI SCIULLO, A. On the domain specificity of the human language faculty and the effects of principles of computational efficiency: contrasting language and mathematics. *Revista Lingüística*, v. 11, n. 1, 2015.

DRESSLER, W.; MERLINI BARBARESI, L. *Morphopragmatics: Diminutives and Intensifiers in Italian, German, and Other Languages*. Berlim: Mouton de Gruyter, 1994.

ESTIVALET, G.; MEUNIER, F. Corpus psicolinguístico Léxico do Português Brasileiro. *Soletrias*, n. 33, p. 212-229, 2017.

GREENBERG, J. Generalizations about numeral systems. In: GREENBERG, J.; FERGUSON, C.; MORAVCSIK, E. (ed.). *Universals of human language*, v. 3, p. 249-295, 1978.

GREENBERG, J. Numeral. In: BOOIJ, G.; LEHMANN, C.; MUGDAN, J. (ed.). *Morphologie: Ein internationales Handbuch zur Flexion und Wortbildung*, v. 1, p. 770-783, 2000.

HAMMARSTRÖM, H.; FORKEL, R.; HASPELMATH, M.; BANK, S. *Glottolog 5.0*. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2024.

HURFORD, J. *The Linguistic Theory of Numerals*. Cambridge: Cambridge University Press, 1975.

IONIN, T.; MATUSHANSKY, O. *Cardinals*: The syntax and semantics of cardinal containing expressions. Cambridge: MIT Press, 2018.

IONIN, T.; MATUSHANSKY, O. The composition of complex cardinals. *Journal of semantics*, v. 23, n. 4, p. 315-360, 2006.

KAISER, S.; ICHIKAWA, Y.; KOBAYASHI, N.; YAMAMOTO, H. *Japanese: A Comprehensive Grammar*. 2. ed. Nova Iorque: Routledge, 2013.

KLOCKMANN, H. Deconstructing base numerals: English and Polish 10, 100, and 1000. In: DOČEKAL, M.; WĄGIEL, M. (ed.). *Formal approaches to number in Slavic and beyond*. Berlim: Language Science, 2021. p. 297-322.

LEE, S. H. *Morfologia e fonologia lexical do português do Brasil*. 1995. Tese (Doutorado em Linguística) – Universidade Estadual de Campinas, 1995.

MERCHANT, J. The syntax of silence: Sluicing, islands, and identity in ellipsis. 1999. Tese (Doutorado em Linguística) – University of California, Santa Cruz, 1999.

MÜLLER, A. L.; NEGRÃO, E.; NUNES-PEMBERTON, G. Adjetivos no português do Brasil: predicados, argumentos ou quantificadores? In: ABAURRE, M. B.; RODRIGUES, A. C. S. (org.). *Gramática do português falado: novos estudos descritivos*. Campinas: Editora Unicamp, 2002.

NESPOR, M.; VOGEL, I. *Prosodic Phonology*. Dordrecht: Foris, 1986.

NEVES, M. H. *Gramática de usos do português*. São Paulo: Editora Unesp, 2011.

QUADROS GOMES, A.; GONÇALVES SUDRÉ, T. Entre a modificação e a quantificação: adjetivos como determinantes plurais em português brasileiro (PB). *Domínios de Lingu@gem*, v. 15, n. 1, 2021.

ROTHSTEIN, S. A Fregean semantics for number words. In: *Proceedings of the 19th Amsterdam Colloquium*. 2013.

SCHWINDT, L. C. Neutralização da vogal pretônica e formação de palavras em português brasileiro. *Organon*, v. 28, n. 54, p. 137-154, 2013.

TATSUMI, Y. *Linguistic Realization of Measuring and Counting in the Nominal Domain: A Cross linguistic Study of Syntactic and Semantic Variations*. 2021. Tese (Doutorado em Linguística) – University of Connecticut, 2021.

ULRICH, C.; SCHWINDT, L. C. O status morfoprosódico dos sufixos inho/zinho, mente e íssimo no Português Brasileiro. *D.E.L.T.A.*, v. 34, p. 769-788, 2018.

VALLS YOSHIDA, F. Uma abordagem morfológica às dezenas cardinais do português brasileiro: -enta como um sufixo derivacional. *Veredas – Revista de Estudos Linguísticos*, v. 28, n. 2, 2024.

VIGNADO, J. *A interface sintática e semântica na análise dos sistemas numerais do karitiana e do kamayurá*. 2019. Dissertação (Mestrado) – Universidade de São Paulo, São Paulo, 2019.

VIGNADO, J. A sintaxe e a semântica de expressões numéricas do karitiana. *Revista Letras*, n. 96, p. 546-565, 2017.

VIGNADO, J. Análise formal de numerais: perspectivas para uma análise minimalista. *Caderno de squibs: temas em estudos formais da linguagem*, v. 4, p. 88-96, 2018.

VIGNADO, J. Evidências morfossintáticas do português, do inglês e do karitiana (tupi) para uma semântica composicional de numerais adequada à interface sintática. *Cadernos de Linguística*, v. 1, p. 1-13, 2020.

WĄGIEL, M. The structure and meaning of label numerals. *Journal of Slavic Linguistics*, v. 31, n. FASL 30 issue, p. 1-20, 2023.

WĄGIEL, M.; CAHA, P. Universal semantic features and the typology of cardinal numerals. *Catalan journal of linguistics*, v. 19, p. 199-229, 2020.

WIESE, H. *Numbers, language, and the human mind*. Cambridge University Press, 2003.

WITKOŚ, J.; DZIUBAŁA-SZREJBROWSKA, D. Cardinal numerals and complex numerals as specifiers. *Poznan Studies in Contemporary Linguistics*, v. 54, n. 4, p. 605-635, 2018.

YANG, C. Systematicity and Arbitrariness In Language: Saussurean rhapsody. In: PAPAFRAGOU, A.; TRUESWELL, J.; GLEITMAN, L. (ed.). *The Oxford Handbook of the Mental Lexicon*. Oxford: Oxford University Press, 2022.