

Contando raízes: análise dos numerais cardinais do português na perspectiva da Morfologia Distribuída

Counting roots: an analysis of Portuguese cardinal numbers from the perspective of Distributed Morphology

Juliana Vignado*
Universidade Estadual de Campinas

Resumo: Este artigo propõe uma análise da derivação de numerais cardinais no português a partir da perspectiva da Morfologia Distribuída e argumenta em favor do conteúdo semântico das raízes. A partir da descrição dos dados, a análise defende que as raízes que formam os numerais têm um conteúdo semântico que consiste na cardinalidade que expressa. Assim, o significado das palavras derivadas depende do significado nuclear subespecificado da raiz e da estrutura em que ela está inserida. Em português, numerais podem ser argumentos (“três é um número primo”) ou predicados (“eu vi os dois meninos”). Nesta análise, isso é explicado pela categorização da raiz, que pode ser realizada por um nominalizador ou um adjetivizador. Foram analisados numerais simples e numerais complexos. O primeiro tipo se refere àqueles formados a partir de uma raiz e um categorizador (“nove”, “quarenta”, “cem”); enquanto o segundo se refere àqueles formados por uma composição (“dezesseis”, “duzentos”), coordenação (“vinte e três”) e adjunção (“dois mil”). Proponho que há na Lista 3 instruções de interpretação de numerais complexos que refletem o conhecimento enciclopédico sobre o funcionamento do sistema decimal. Assim, a restrição sobre a seleção de raízes é explicada a partir do conteúdo semântico da raiz e do conhecimento enciclopédico.

Palavras-chave: Numeral. Morfologia Distribuída. Semântica. Raiz.

Abstract: This article proposes an analysis of the derivation of cardinal numerals in Portuguese from the perspective of Distributed Morphology and argues in favor of the semantic content of roots. Based on the description of the data, the analysis maintains that the roots that form numerals have semantic content consisting of the cardinality they express. Thus, the meaning of derived words depends on the underspecified core meaning of the root and on the structure in which it is embedded. In Portuguese, numerals can function as arguments (“três é um número primo”) or as predicates (“Eu vi os dois meninos”). In this analysis, this contrast is explained by the categorization of the root, which may be carried out by a nominalizer or an adjectivizer. Both simple and complex numerals are analyzed. The former refers to those formed from a root and a categorizer (“nove, quarenta, cem”), whereas the latter refers to those formed by compounding (“dezesseis, duzentos”), coordination (“vinte e três”), and adjunction (“dois mil”). I propose that List 3 contains interpretive instructions for complex numerals that reflect encyclopedic knowledge about the functioning of the decimal system. Thus, restrictions on root selection are explained in terms of the semantic content of the root and encyclopedic knowledge.

Keywords: Numeral. Distributed Morphology. Semantics. Root.

1 INTRODUÇÃO

O objetivo deste artigo é propor uma análise da derivação de numerais cardinais (doravante cardinais) no português sob a perspectiva da Morfologia Distribuída. A discussão considera os cardinais simples e complexos e foca em aspectos morfossintáticos

*Doutoranda em Linguística. Instituto de Estudos da Linguagem, Unicamp, Campinas, São Paulo, Brasil; juvign@gmail.com

e semânticos da derivação de cardinais. A partir da análise dos dados, argumenta-se em favor do conteúdo semântico das raízes, um debate importante para esse quadro, como será apresentado em mais detalhes no panorama teórico.

Conforme autores como Comrie (2005) e Hammarström (2010), numerais cardinais são expressões linguísticas que se referem a números inteiros e positivos; isto é, o conjunto dos números naturais em termos matemáticos. Trata-se de uma definição abrangente que engloba a grande variação verificada a respeito da expressão de cardinalidade nas línguas naturais, sem perder de vista o fato comum a essa classe de palavras: a expressão de uma cardinalidade.

As ideias centrais da Morfologia Distribuída consistem em assumir que a formação de palavras é sintática e que existam elementos atômicos não composicionais, chamados ‘raízes’, compreendidos como as peças básicas para a formação de palavras e sentenças, juntamente com os traços formais provindos da Gramática Universal (Marantz, 1996). As raízes são neutras quanto à sua categoria e são categorizadas ao se combinarem com um núcleo categorizador na sintaxe; e o conteúdo das raízes não é decomponível na sintaxe.

No desenvolvimento dessa teoria, emerge um debate importante sobre a natureza das raízes, que consiste na defesa ou não de que raízes têm conteúdo semântico. Nessa direção, argumento que os numerais fornecem evidências a favor da presença de conteúdo semântico nas raízes. Guerreiro *et al.* (2020) chamam de ‘princípio da cardinalidade’ o fato de que expressões linguísticas são associadas a valores cardinais definidos, e tais valores definidos dependem dos valores cardinais associados às raízes envolvidas e da estrutura sintática. Este artigo defende a existência de um conjunto de raízes no português que tem semântica cardinal, como será explorado em mais detalhes na seção 3.

FLP 27(2)

Analisando dados do inglês, do hebraico e do grego, Alexiadou e Lohndal (2017) propõem que é possível aventar uma tipologia linguística de raízes e categorias funcionais, na qual as línguas se organizam em uma escala. Em uma extremidade, estariam línguas como o hebraico, em que os elementos funcionais são fundamentais para determinar a interpretação de uma raiz e da palavra derivada, assim, as raízes teriam um significado mais abrangente; e, na outra extremidade, estariam línguas como o inglês, em que a interpretação de uma raiz e da palavra derivada dependem mais do significado da própria raiz, assim, as raízes teriam um significado estrito. Como será apresentado na seção 3, as raízes que formam numerais do português exibem um significado bastante regular nas palavras derivadas, diante disso, busco explorar o seu significado a fim de observar se esses elementos apresentam um significado abrangente ou estrito.

A investigação sobre os numerais e a natureza desse conjunto de raízes levantam ainda outras questões. É consenso na literatura que muitos cardinais apresentam um significado composicional: numerais como “duzentos” (2×100) e “dezoito” ($10 + 8$), por exemplo, são interpretados a partir de uma relação de multiplicação e adição entre o significado — que consiste numa cardinalidade — dos morfemas que compõem a palavra. Adição e multiplicação não são operações linguísticas, mas são operações amplamente observadas na interpretação de numerais das mais diversas línguas (*cf.* Huford, 1987). Numerais compõem a gramática de uma língua e são derivados assim como todas as palavras, expressões e sentenças. Desse modo, esta análise busca esclarecer

onde tal interpretação — que relaciona conhecimento linguístico e conhecimento matemático — acontece na arquitetura da gramática.

Dito de outra forma, se numerais complexos são formados por estruturas que são interpretadas a partir de operações matemáticas, como adição e multiplicação, como essa interpretação é gerada na arquitetura da gramática? Para responder a essa pergunta, explorarei os dados de numerais a partir da perspectiva da Morfologia Distribuída, a fim de explicar como se dá a sintaxe numérica no português. Sendo assim, dentre outros aspectos, esta análise se concentra em aspectos morfossemânticos do fenômeno.

Evidentemente, o conjunto de dados analisados neste artigo não permite propor generalizações fortes sobre a natureza de todas as raízes da língua portuguesa, o que exigiria um *corpus* maior e mais diversificado. De todo modo, esta pesquisa visa contribuir com o debate sobre a natureza das raízes, mais especificamente, no que concerne à presença de conteúdo semântico inerente.

Em termos semânticos, no português, cardinais podem ser argumentais, quando nominais (1-a); ou predicativos, quando adjetivais (1-b); e um numeral predicativo pode ainda modificar um numeral argumental (1-c).

- (1) a. Sete é ímpar.
- b. Maria tem dois filhos.
- c. Tem quatro setes vermelhos no baralho.

Rothstein (2017) propõe que essa variação possa ser explicada por uma operação de mudança de tipo (*Type Shifting*). Sob a ótica da Morfologia Distribuída, proponho que essa diferença se deva pela estrutura sintática, mais especificamente pelo mecanismo de categorização. Seguindo a autora, entendo que numerais são nomes, por *default*, e defendo que designam uma propriedade cardinal que reflete o conteúdo semântico da raiz.

FLP 27(2)

Os dados analisados foram reunidos por Vignado (2023), que apresenta uma descrição completa dos numerais do português, incluindo cardinais, ordinais, multiplicativos e palavras não numéricas derivadas das raízes com conteúdo cardinal. A relevância da análise reside no fato de que os numerais do português foram pouco analisados formalmente na literatura, e, além disso, a semântica das raízes é um debate atual e relevante para o quadro da Morfologia Distribuída. A pesquisa, também, contribui para uma melhor compreensão da interação entre o conhecimento matemático e o conhecimento linguístico, que são fundamentais para a investigação científica sobre numerais.

2 PANORAMA TEÓRICO

2.1 Numerais

Nesta subseção, são apresentados alguns aspectos dos numerais que serão relevantes para a análise proposta. Chomsky (1998) sugere que sistemas numerais são um subproduto da Faculdade da Linguagem. Como distingue Vignado (2019), um

sistema numeral é a expressão linguística de um sistema numérico; assim, o sistema numeral é a parte de uma língua natural cuja função é expressar cardinalidades, *i.e.*, números inteiros e positivos. Ainda nessa direção, Hauser, Chomsky e Fitch (2002) afirmam que a Faculdade da Linguagem fornece aos numerais a estrutura recursiva que permite formar potencialmente infinitos numerais a partir de um conjunto finito de elementos. Para esses autores, a formação de números exatos só pode existir via estrutura recursiva fornecida pela Gramática. Há ainda outros autores que também defendem que a Faculdade da Linguagem confere aos sistemas numéricos/numerais seu poder composicional, como Hurford, (1975, 1987, 2006), Rothstein, (2013, 2017) e Ionin e Matushanky (2018).

O primeiro ponto importante na pesquisa sobre numerais é a distinção entre numerais simples e numerais complexos, que é consenso na literatura sobre o tema (*cf.* Hurford, 1987; Zabbal, 2005; Rothstein, 2017 e outros). Zabbal (2005) propõe que numerais simples são itens lexicais, e numerais complexos são construídos a partir de numerais simples na sintaxe e interpretados composicionalmente. Para Rothstein (2017), numerais simples são palavras que denotam um número, e numerais complexos podem ser analisados em subpartes cujos significados são construídos composicionalmente, usando multiplicação e adição. Para a autora, a composição de numerais complexos pode ser morfológica — e, portanto, pré-sintática — ou sintática. Assim, numerais como “*sixteen*” (‘dezesesseis’) seriam morfemas livres, mas são composicionais por serem formados a partir do significado de “*six*” (‘seis’) e do sufixo ‘-teen’, que forma numerais de 13 a 19 no inglês. Rothstein sugere que esse tipo de composição ocorre pré-sintaticamente no léxico, o que implica a assunção de um léxico gerativo em alguma medida.

FLP 27(2)

Para Hurford (2006), uma das principais características dos sistemas de regras sintáticas para formação de numerais complexos é o fato de que a categoria NUMERAL é recursiva. Assim, a análise parte da constatação de que numerais são formados na sintaxe e que são recursivos. Seguindo Guerreiro *et al.* (2020), essa característica será chamada de ‘sintaxe numérica’. Por fim, a investigação das regras sintáticas que formam numerais é uma abordagem com potencial para elucidar a representação numérica nas línguas naturais ou, em outras palavras, esclarecer quais mecanismos linguísticos são empregados para a expressão dos números.

É bastante descrito na literatura lexicalista (*cf.* Hurford, 1975, 1987; Zabbal, 2005; Ionin; Matushanky, 2006; Rothstein, 2013, 2017; Guerrero *et al.*, 2020; Di Sciullo, 2022) o fato de que numerais complexos veiculam operações matemáticas em sua interpretação semântica, que refletem sua estrutura sintática — como, por exemplo, no português, “dezenove”, com uma operação de adição, $10+9$, e “trezentos”, com uma operação de multiplicação, 3×100 . Portanto, esta pesquisa considera que numerais são formados na sintaxe; porém, nesta análise, considera-se que todos os numerais são formados na sintaxe, não somente os complexos.

Assim, a partir da constatação de uma morfologia bastante regular e da assunção de que numerais são recursivos e derivados na sintaxe, buscou-se considerar uma teoria linguística cujas ferramentas teóricas permitem descrever e analisar a estrutura interna dessas expressões. Nessa direção, este artigo se propõe a discutir o fenômeno a partir da Morfologia Distribuída, uma teoria sintática não lexicalista. Em suma, a observação

dos numerais complexos evidencia uma grande regularidade na constituição morfológica desse conjunto de dados, sugerindo que as regras sintáticas que operam na construção dessas expressões ocorram abaixo do nível da palavra, o que justifica a implementação de uma teoria sintática desse tipo.

2.2 Morfologia Distribuída

A Morfologia Distribuída é uma abordagem não lexicalista que defende a hipótese do mecanismo único (Halle; Marantz, 1993; Marantz, 1997); ou seja, defende que toda derivação é sintática. Com efeito, as operações que formam as sentenças são as mesmas operações que formam as palavras, como *'merge'* ('concatenar') e *'move'* ('mover'). Essa propriedade do modelo ficou conhecida como *'syntactic structure all the way down'* (Harley; Noyer, 1999), comumente traduzida como 'estrutura sintática por toda a derivação', para empregar os termos de Scher, Bassani e Minussi (2013). A teoria propõe um léxico formado apenas por raízes e traços formais, e não itens lexicais. As raízes são acategoriais, sendo, por isso, categorizadas por elementos funcionais chamados 'categorizadores' (Marantz, 1997; Embick; Marantz, 2008). Tais características serão apresentadas em mais detalhes a seguir.

Como sintetizam Alexiadou e Lohndal (2017), a ideia principal da Morfologia Distribuída é a de que a formação de objetos linguísticos, por menores que sejam, como as palavras, é sintática, e não pré-sintática como assumem as teorias lexicalistas. Além disso, assume-se a existência de elementos atômicos e não composicionais chamados 'raízes'. Halle e Marantz (1993, 1994) e Marantz (1996, 1997), os trabalhos fundadores da teoria, defendem que as propriedades morfológicas, fonológicas e semânticas do então 'item lexical' não se concentram em um único componente da arquitetura da Gramática, o léxico, mas estão distribuídas em diferentes componentes.

Diante disso, na Morfologia Distribuída, em vez de um léxico — gerativo ou não —, há três listas que compõem a arquitetura da Gramática, que são acessadas em diferentes momentos da derivação. A Lista 1, também chamada de 'léxico estrito', contém os elementos primitivos, as raízes e os traços formais. Essa lista é acessada antes da derivação sintática, e todas as derivações são formadas a partir desses elementos primitivos. A Lista 2, ou Vocabulário, contém os Itens de Vocabulário, que são as regras de correspondência entre os nós terminais abstratos derivados na sintaxe e os feixes de traços fonológicos. Note que vários Itens de Vocabulário podem ser candidatos para inserção em um mesmo nó sintático, assim, o sistema precisa decidir qual Item vence essa competição. Essa decisão é definida pelo Princípio de Subconjunto (Halle, 1996). A Lista 3, ou Enciclopédia, contém os significados especiais e o conhecimento de mundo do falante. Essa lista contém os significados não composicionais. Já a semântica composicional é determinada na sintaxe, via estrutura. Isso interessa aos objetivos deste trabalho, pois o significado dos numerais depende da estrutura sintática e da composicionalidade, mas a semântica contida na raiz é não composicional.

Por fim, há ainda um nível morfológico — *'morphological structure'* ou MS — na ramificação para a Forma Fonológica (*'phonological form'* PF), em que são aplicadas as operações particulares, responsáveis por atender aos requerimentos de boa formação morfológica e/ou morfossintática de cada língua. Após a aplicação das regras

morfológicas, ocorre a inserção de Vocabulário, em que a realização fonológica vai ser selecionada a partir da entrada mais subespecificada em relação aos traços fornecidos pela sintaxe. Esse processo é chamado de ‘Inserção Tardia’.

Quanto ao estatuto da raiz, seguindo os pressupostos da teoria, a análise parte da ideia de que as raízes estão no léxico sem categoria, sendo categorizadas na derivação sintática. Os categorizadores primitivos da teoria são *verbalizador*, *nominalizador* e *adjetivizador* e são representados pelas letras minúsculas ‘v’, ‘n’ e ‘a’. Abaixo, o exemplo em (2-a) representa uma raiz categorizada como um verbo; em (2-b), categorizada como um nome e, em (2-c), categorizada como um adjetivo.

- (2) a. [vP v √RAIZ]
 b. [nP n √RAIZ]
 c. [aP a √RAIZ]

Esta seção buscou apresentar brevemente algumas premissas básicas do quadro teórico da Morfologia Distribuída adotado para a análise dos numerais proposta aqui. A seguir, a próxima seção apresenta a descrição dos cardinais do português.

3 OS NUMERAIS DO PORTUGUÊS

Nesta seção serão apresentadas propriedades gerais dos numerais cardinais e a proposta de que a semântica cardinal está contida na raiz. Para isso, para além dos cardinais simples e complexos do português, são consideradas também palavras derivadas das mesmas raízes que formam os numerais, para fins de descrição do fenômeno das raízes.

Esta análise emprega uma teoria não lexicalista em que não há itens lexicais predeterminados e tampouco um léxico gerativo, mas sim um léxico estrito — uma lista — composto por raízes e traços formais. Dessa forma, todas as estruturas são derivadas na sintaxe. Adotarei uma definição para numerais simples e complexos semelhante à definição de Rothstein (2017) quanto à composicionalidade dos numerais complexos, mas que se diferencia dela por considerar que todos os numerais são construídos na sintaxe. Nesse sentido, proponho que numerais possam ser simples, *i.e.*, construídos a partir de uma única raiz, ou complexos, *i.e.*, aqueles que são compostos por duas raízes e que podem ser analisados a partir de suas subpartes, cujos significados são construídos composicionalmente. Por exemplo, “sete” é um numeral simples que denota o número 7, já “dezessete” é um numeral complexo, já que é formado por “dez” e “sete” em uma relação de adição 10+7. Há ainda numerais que não são uma única palavra, mas sintagmas, como “vinte e três”. Em síntese, o que define o numeral complexo em nossa proposta é a presença de duas ou mais raízes — ou numerais — e a composicionalidade semântica de sua estrutura interna, necessária para interpretar o valor cardinal da expressão.

No português, os numerais cardinais — simples ou complexos — ocorrem em dois contextos sintáticos, um em que opera como um argumento (3) e um em que opera como um modificador predicativo (4)–(5). Em (3-b) e (5) pode-se observar que um numeral argumental pode ser modificado por um numeral predicativo.

Comportamento semelhante foi observado por Rothstein (2013, 2017) para dados do inglês. Conforme mencionado, a autora defende um tratamento dos numerais cardinais como propriedades, assim como nomes comuns, que operam como predicadores.

- (3) a. **Sete** é ímpar.
b. Tem quatro **oitos** nesse baralho.
- (4) a. Maria tem **dois** filhos
b. Maria tem **duas** filhas
- (5) Eu tenho **três** dois. (no contexto de um jogo de cartas)

Os exemplos em (3) mostram que os cardinais argumentais podem apresentar concordância de número plural. Já no caso dos numerais predicativos, pode haver concordância de gênero, conforme apresentado em (4), que ocorre em “um”, “dois” e na série das centenas, como mostra (6), a seguir.

- (6) a. **Um** menino/**uma** menina ganhou a corrida.
b. Vinte **dois** homens/vinte e **duas** mulheres passaram no concurso.
c. **Duzentos** idosos/**duzentas** idosas receberam o benefício.

Os demais numerais do português não apresentam concordância de gênero, ou seja, são invariáveis. Esta análise não irá discutir o fenômeno da concordância de gênero em detalhes, mas vai assumir a proposta de Resende e Santana (2019) para o português, segundo a qual gênero é um traço do nominalizador, que vem em três diferentes matizes: n^o [], n^o [FEM] e n^o [MASC]. Nessa notação, o categorizador n pode ser subespecificado para gênero, ou contar com traços de gênero masculino e feminino.

FLP 27(2)

Em relação à propriedade de número, com exceção do numeral “um”, que denota um valor cardinal singular por excelência, os demais numerais cardinais, quando predicativos, são invariáveis e fazem com que o nominal que modificam receba a marca de plural. Como mostram os dados em (7), esses elementos não podem receber marca de plural quando em posição predicativa.

- (7) a. Maria tem **sete** cachorros.
b. *Maria tem **setes** cachorros.

Para Rothstein (2017), numerais são propriedades de pluralidades, portanto nomes, que se tornam predicados por uma operação de mudança de tipo ‘*type-shifting*’. Seria possível imaginar que “um” não é um numeral genuíno, mas sim um quantificador existencial. De fato, em uma sentença como “um menino chegou”, não é possível distinguir uma leitura de sintagma cardinal de uma leitura de sintagma com quantificação existencial com facilidade. Além disso, note que o numeral “um” não ocorre na formação de numerais complexos, como mostra o contraste entre “cem” e “duzentos” ou entre “mil” e “dois mil”¹.

¹Não serão considerados nesta análise os numerais maiores como “milhão”, “bilhão”, “trilhão” etc. Uma

Numerais cardinais em posição argumental, como os demais nomes da língua, podem receber sufixo aumentativo, como em (8-a), e diminutivo, como em (8-b), o que é agramatical em posição predicativa, como em (8-c) e (8-d).

- (8) a. O Jair me deve **vintão**.
 b. Por **duzentinhos** eu aceitaria o trabalho.
 c. ***vintões** meninos chegaram.
 d. *Eu recebi **duzentinhos** reais.

Mesmo que a maioria dos numerais cardinais seja invariável, seu comportamento é predicativo e, em alguns casos, há padrões de concordância, o que é característica dos adjetivos. Numerais predicativos podem coocorrer com determinantes, como em “as/essas duas casas amarelas”, de modo que não são propriamente determinantes em português, uma proposta defendida na semântica formal por autores como Barwise e Cooper (1981), Keenan e Faltz (1984) e Hofweber (2005). Em português, esses numerais são pré-nominais, como em (9).

- (9) a. Eu vi **dois meninos**.
 b. *Eu vi **meninos dois**.

Ainda que a posição canônica dos adjetivos em português seja pós-nominal e existam adjetivos que sejam licenciados tanto em posição pré-nominal quanto em posição pós-nominal, Prim (2015) mostra que existem adjetivos que são licenciados somente em posição pré-nominal, note-se o contraste de gramaticalidade entre “mero aprendiz” e “*aprendiz mero”. Esse também é o caso dos numerais cardinais.

FLP 27(2)

Adicionalmente, Landman (2003) reúne argumentos em favor de tratar numerais cardinais como adjetivos. Para o autor, em sintagmas numéricos, em seus termos, como “*three girls*” ‘três garotas’, o numeral se comporta semanticamente exatamente como os adjetivos intersectivos, ou seja, o significado do sintagma é dado pela intersecção entre a denotação do numeral e a denotação plural do nome “*girls*” ‘garotas’, *i.e.*, na intersecção do conjunto de conjuntos de duas unidades e do conjunto de garotas.

observação preliminar desses numerais nos mostra que parece não existir uma semântica cardinal como a proposta aqui. Note que “**bilhão**” (10⁹) não são “dois milhões”, assim como “**trilhão**” (10¹²) não são “três milhões” e assim por diante. Sugiro que essas palavras refletem uma ordenação das bases, por exemplo, “**milhão**” (10⁶), “**bilhão**” (10⁹), “**trilhão**” (10¹²), “**quatrilhão**” (10¹⁵), que precisa ser mais bem explorada em investigações futuras. Veja que *trilhão*, por exemplo, consiste na terceira base dessa série. Além disso, bases livres de maior valor se comportam de forma diferente de uma base livre como “mil”, como mostram os dados a seguir.

- a. **Vinte mil** pessoas votaram no vereador.
 b. **Vinte milhões** de pessoas não votaram na eleição.

Em (a), “vinte mil” é um numeral complexo formado por adjunção — onde “vinte” modifica “mil” — e o numeral se liga diretamente ao nome que modifica, já em (b) observamos a presença de uma preposição, nesse caso, “milhões” parece funcionar como uma palavra de medida, note que se trata da mesma estrutura de sintagmas com palavras de medida em português, como “dois quilos de farinha” ou “duas dúzias de bananas”.

Os dados mostraram que o numeral argumental se comporta exatamente como um nome com gênero interpretável. Quando argumental, seus adjuntos concordam com o numeral, como mostra o contraste de gramaticalidade entre “tem três oitos vermelhos no baralho” e “*tem três oitos vermelhas no baralho”². Em contrapartida, quando são predicativos ‘um’, ‘dois’ e a série das centenas concordam em gênero com o nome que modificam, reforçando a hipótese de que são adjetivos. Em síntese, o numeral argumental e numeral predicativo têm comportamentos sintáticos diferentes, e, portanto, pertencem a categorias sintáticas diferentes.

Esta análise defende que o numeral argumental é um nome que designa uma propriedade cardinal de tipo $\langle n \rangle$ (cf. Rothstein, 2017), portanto, ele é categorizado por um nominalizador. Já os numerais predicativos são numerais cardinais, e sua natureza predicativa é definida pela recategorização por um adjetivizador, como será esclarecido mais adiante.

Outra análise possível seria considerar a existência de um categorizador numeralizador, *i.e.*, que categorizaria raízes como numerais. Contudo, como mostram os dados, a natureza morfossintática e semântica dos numerais sugere que eles funcionam como os demais nomes e adjetivos da língua e a categoria ‘numeral’ não inclui apenas os cardinais, mas numerais ordinais, multiplicativos e fracionários, cuja semelhança com os cardinais consiste no valor expresso pela raiz, mas cujos significados são diferentes. Ademais, propor um novo primitivo nesse quadro teórico exigiria evidências translinguísticas sobre o fenômeno para sustentar tal implementação, o que foge ao escopo desta análise.

A presente proposta segue Rothstein (2017), porém, diferentemente da proposição da autora, não requer uma operação de mudança de tipo, pois explica a variação a partir da categorização da raiz ou estrutura. Note que, apesar da categorização e da natureza semântica da expressão, o valor cardinal da raiz sempre se mantém no numeral, como mostra ??, em que é apresentada a semântica dos numerais argumentais e predicativos.³

FLP 27(2)

- (10) a. $\llbracket \text{quatro}_{arg} \rrbracket_{\langle n \rangle} = 4$
 b. $\llbracket \text{quatro}_{pred} \rrbracket_{\langle e, t \rangle} = \lambda x. |x| = 4$

A cardinalidade da raiz se mantém mesmo em numerais que não são cardinais, como mostram os dados em (11), de numerais ordinais, numerais multiplicativos, ou, ainda, em palavras que não são numerais, de forma que não é o caso que isso ocorra porque esses

²À primeira vista, pode-se argumentar que na sentença “tem três oitos vermelhas no baralho” a palavra ‘cartas’ está elipsada na estrutura justificando sua gramaticalidade. Contudo, nesse caso, o numeral não receberia a marca de plural. Como mostram as sentenças a seguir.

- a. *Tem três [cartas] **oitos vermelhas** no baralho.
 b. Tem três [cartas] **oito vermelhas** no baralho.

³Valls Yoshida (2024) segue Ionin e Matushansky (2018) na hipótese de que cardinais são de tipo semântico $\langle \langle e, t \rangle \rangle$. Bylinina e Nouwen (2020) mostram que a ocorrência de numerais com determinantes em inglês — como também é o caso do português — é um contra-argumento a esse tipo de tratamento semântico.

elementos pertencem a uma classe gramatical própria, sendo, portanto, mais adequado postular que essa semântica está contida na raiz.

- (11) a. $\sqrt{\text{TRI}}$ — três, trinta, triplo, tríade, triângulo, trisal.
 b. $\sqrt{\text{NOV}}$ — nove, nono, nônio, novena.
 c. $\sqrt{\text{CEM}}$ — cem, centésimo, centena, centenário.
 d. $\sqrt{\text{MIL}}$ — mil, milésimo, milhar, milenar, milênio.

Os dados mostram que, mesmo as palavras que não são numerais, como “triângulo”, “novena”, “centenário”, “milenar” *etc.*, por exemplo, mantêm em sua semântica a cardinalidade expressa pela raiz. Esse fato é ilustrado em mais detalhes em (12), com palavras que não são cardinais derivadas da raiz $\sqrt{\text{TRI}}$ ⁴.

- (12) a. triplo: que contém três vezes a mesma quantidade.
 b. tricúspide: que tem três pontas.
 c. trio: conjunto de três elementos.
 d. trivalente: que tem três valências.
 e. triplicar: multiplicar por três.

Essas palavras foram coletadas no dicionário (Novo Dicionário Aurélio, 1986), e os significados apresentados são os primeiros listados para cada verbete. Evidentemente, é possível elencar muitas outras palavras que corroboram a constatação de que as palavras derivadas desse tipo de raiz contêm um significado cardinal como parte de seu significado. De todo modo, esse conjunto de dados é suficiente para os fins desta exposição.

Assim, propõe-se que a consistência verificada no significado das palavras derivadas dessas raízes, seguindo autores como Alexiadou, (2001), Arad, (2003) Alexiadou e Lohndal (2017), vem do fato de que as raízes carregam um significado nuclear subespecificado. Note que uma cardinalidade é um significado bastante simples e, dessa forma, pode ser considerado bastante subespecificado. Nos dados, foi possível observar que, independentemente da classe da palavra e do significado atribuído a ela, o que se mantém é o valor cardinal.

⁴Um parecerista anônimo observou que dados como “tricúspide” ou “trivalente” sugerem que as raízes em questão podem funcionar como prefixos, podendo operar também como itens funcionais, o que revelaria uma certa plasticidade de domínios e configurações em que tais elementos podem ocorrer. Sobre isso, Nóbrega (2014, p. 84) trata a palavra “trigêmeo”, por exemplo, como um composto palavra-palavra, para esse autor, se trata de um composto Numeral-Nome. Esta análise concorda com a primeira afirmação, mas questiona a visão de “numeral” como uma categoria sintática. “Numeral” é uma categoria descritiva importante que abrange palavras cujo significado se relaciona com cardinalidades, mas engloba palavras com semânticas e comportamentos morfossintáticos distintos, como numerais multiplicativos (ex. “triplo”) e cardinais, por exemplo. Os dados analisados aqui sugerem que cardinais se comportam como os demais nomes e adjetivos do português. De todo modo, essa é uma questão relevante para investigações futuras.

4 ANÁLISE

A descrição dos dados mostrou que a cardinalidade das raízes se mantém nas palavras derivadas, mesmo quando não se trata de um numeral. Nesta análise, esse conteúdo será denominado “semântica cardinal”. A semântica cardinal presente em algumas raízes da língua portuguesa é representada pela notação de cardinalidade utilizada pela matemática, em que um valor cardinal ‘*n*’ é representado como $|n|$, como em (13).

$$(13) \quad \llbracket \sqrt{\text{RAIZ.N}} \rrbracket = |n|$$

Essa notação para representar cardinalidades também é utilizada na literatura linguística por Ferreira (2008, 2012) e Rothstein (2013, 2017). Assim, a natureza cardinal do numeral provém da cardinalidade da raiz e não é preciso propor um traço no categorizador ou categorizador especial para derivar esse tipo de numerais, basta que a raiz seja categorizada por um nominalizador.

Veja que, nessa perspectiva, a série dos numerais de 11 a 15, as dezenas e as bases maiores como *mil* constituem numerais simples⁵. Nesse caso, o que diferencia “cinco” de “quinze” e de “cinquenta”? A partir dos dados, pode-se observar três categorizadores ‘*n*’ que formam os cardinais em português: $\emptyset$ (‘quatro’), *-z-* (‘onze’) e *-nt-* (‘quarenta’). Propor que o cardinal é derivado por um nominalizador torna necessário explicar como ele se torna um adjetivo, quando se realiza como um numeral predicativo.

Como mostrado por Rothstein (2017), e mencionado anteriormente, o numeral é um nome por *default* que, assim como todos os nomes, designa uma propriedade que, no caso dessas expressões, é uma propriedade cardinal. Nessa direção, esta análise propõe que numerais são derivados por um nominalizador em um primeiro ciclo, e posteriormente são recategorizados por um adjetivizador para se tornarem numerais predicativos. Esse mecanismo, previsto pelo quadro da Morfologia Distribuída, captura o fenômeno do comportamento semântico dessas expressões, explicando a alternância entre o comportamento argumental e predicativo, que, como apresentado na seção 3, têm consequências morfossintáticas bastante claras. Além disso, não implica propor nominalizadores e adjetivizadores com a mesma forma fonológica, como seria necessário para derivar numerais simples sem o mecanismo de recategorização por um ‘*a*’ nulo.

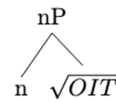
Diante do que foi exposto, a análise proposta considera um numeral simples todo aquele que é formado apenas por uma raiz, como os numerais “um”, “dois”, “nove”, “quinze”... “trinta”, “quarenta”, “dez”, “cem”, “mil”⁶. Para fins de exposição, tomemos

⁵ Isso implica raízes que expressam cardinalidades de 11 a 15, veja que algumas dessas raízes também formam outras palavras, como “dúzia” e “quinzena”.

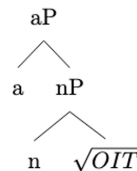
⁶ É importante distinguir entre o que é um numeral simples do que é um numeral básico. Nesta proposta, o numeral simples tem uma estrutura sintática simples composta por apenas uma raiz categorizada. Já um numeral básico pode ser entendido como uma unidade de um sistema numérico — como 1 a 9 no sistema decimal — que se constrói em cima de bases e numerais básicos de forma recursiva. Nas pesquisas sobre o tema, é comum que ambos sejam tratados da mesma forma, isso se deve à abordagem lexicalista que inicia derivação a partir de itens lexicais pré-existentes, assim como um sistema numérico se desenvolve sobre unidades e bases. Do ponto de vista desenvolvido aqui, ‘simples’ e ‘complexo’ dizem respeito a

o numeral “oito” como exemplo, formado pela raiz $\sqrt{\text{OIT}}$ e a vogal temática ‘-o’, de modo que a estrutura morfológica de numerais simples é RAIZ-VT. Assim, “oito” é nome que designa uma propriedade cardinal em (14-a) e um adjetivo em (14-b).

(14) a.



b.



Do ponto de vista semântico, (14-a) é de tipo $\langle n \rangle$ enquanto (14-b) é de tipo $\langle e, t \rangle$. Como observa Rothstein (2017), reforçando a análise de que numerais do tipo (14-a) denotam uma propriedade cardinal n , existem ainda propriedades de tipo $\langle n, t \rangle$, como é um número par em sentenças como “Dois é um número par” — veja que aplicar este tipo de predicado sobre entidades de tipo $\langle e \rangle$ gera agramaticalidade como “*João é um número par”.

Em síntese, uma análise pautada na categorização da raiz/estrutura permite explicar por que numerais cardinais argumentais e predicativos têm comportamentos morfossintáticos diferentes sem precisar recorrer para uma operação semântica de mudança de tipo. Conforme apresentado na descrição dos dados, o numeral argumental pode receber morfologia de número plural e de grau — o que não é possível para sua forma predicativa — e, no caso de alguns numerais, a categorização por um adjetivizador provoca concordância de gênero, reforçando a proposta.

FLP 27(2)

A partir das considerações apresentadas, podemos observar em (15) um exemplo de derivação que ilustra a análise proposta por este trabalho a partir da perspectiva da Morfologia Distribuída. Para todos os cardinais, quando predicativos, considere o 2º ciclo da derivação com uma recategorização por a que tem consequências para a interpretação semântica do item derivado.

(15) Derivação de numerais

$\sqrt{|6|} \rightarrow \text{‘seis’}$

1º ciclo:

i. Sintaxe $\rightarrow \sqrt{|6|} + n(\emptyset)$

ii. PF $\rightarrow$ inserção de vocabulário $\rightarrow \sqrt{|6|} \leftrightarrow /ses/$
 $\rightarrow$ inserção de vogal epentética $/ses/ \rightarrow /seis/$

iii. LF $\rightarrow \llbracket \text{seis} \rrbracket_{\langle n \rangle} = \text{propriedade cardinal } |6|$

estrutura sintática e à composicionalidade dos numerais. Portanto, esses conceitos dizem respeito a aspectos estritamente linguísticos e não a aspectos matemáticos.

2º ciclo:

- i. Sintaxe $\rightarrow nP + a(\emptyset)$
- ii. PF $\rightarrow /seis/$
- iii. LF $\rightarrow \llbracket seis \rrbracket_{\langle e,t \rangle} = \lambda x. |x| = 6$

Como vimos nos numerais simples, a mesma raiz pode ser categorizada por todos os categorizadores, expressando valores diferentes (ex. “quatro”, “quatorze”, “quarenta”). Isso ocorre porque um núcleo categorizador concatenado a uma raiz pode influenciar seu significado, especificando um domínio de interpretação idiossincrática. Note que o significado de numerais de outros tipos, como os ordinais e os multiplicativos, reforçam essa perspectiva. Nos numerais “quatro” (cardinal), “quarto” (ordinal) e “quádruplo” (multiplicativo), pode-se observar com clareza a mesma raiz, com valor $|4|$. A semântica dessas palavras, contudo, é diferente, e é definida pelo contexto do sufixo derivacional empregado. Para Marantz (1997; 2001), morfemas derivacionais são itens funcionais. Assim, trata-se de categorizadores diferentes que levam a interpretações idiossincráticas diferentes.

Portanto, nessa perspectiva, as estruturas derivadas por categorizadores diferentes fornecem instruções diferentes na Lista 3, acessada na Forma Lógica LF. Tais instruções, apresentadas em (16), refletem o conhecimento extralinguístico e, portanto, enciclopédico sobre o sistema numérico decimal, subjacente ao sistema numeral do português.

(16) Instrução para Lista 3

- a. $n[\sqrt{|n|} + \emptyset] \leftrightarrow$ interprete como n
- b. $n[\sqrt{|n|} + -z-] \leftrightarrow$ interprete como $n + 10$
- c. $n[\sqrt{|n|} + -nt-] \leftrightarrow$ interprete como $n \times 10^7$

FLP 27(2)

Esses diferentes categorizadores podem gerar contextos de alomorfia da raiz. Como apresentado na seção 2, na Morfologia Distribuída, muitos autores assumem um mecanismo chamado de Inserção Tardia, em que os itens de vocabulário da Lista 2 são inseridos pós-sintaticamente. Um exemplo de alomorfia desse tipo é apresentado em (17).

⁷Para as dezenas, Valls Yoshida (2024) oferece uma análise alternativa. Nessa perspectiva, as dezenas são formadas pelo sufixo derivacional ‘-enta’ que, em sua análise, é tratado como uma raiz. Ainda que Valls Yoshida levante bons argumentos teóricos em sua análise, esta pesquisa traz uma proposta que abarca a derivação dos numerais cardinais de forma geral, não somente das dezenas, buscando uma análise capaz de explicar todo o sistema de forma unificada. Considero que ‘-nt-’ seja o item de vocabulário associado ao categorizador, seguindo a perspectiva de que morfemas derivacionais sejam itens funcionais (Marantz, 1997; 2001). A partir desta abordagem, temos raiz, categorizador e vogal temática na estrutura. Veja que o numeral vinte configura uma evidência em favor desta análise, já que é uma dezena, mas não apresenta o sufixo ‘-enta’, fortalecendo, assim, a hipótese de [RAIZ-nt-VT]. Outra diferença importante é que, para o autor, as raízes não contêm conteúdo semântico, perspectiva contrária àquela proposta aqui, que defende que um sistema numeral se constrói não somente a partir de conhecimentos linguísticos, mas também de conhecimentos extralinguísticos sobre o sistema decimal, que se relacionam com o conteúdo cardinal das raízes.

a. $\sqrt{|5|} \leftrightarrow /sĩk/$ (cinco, cinquenta)
 b. $\sqrt{|5|} \leftrightarrow /kĩ/$ ______n $[-z]$ (quinze)
 ______n $[\sqrt{|n|} + \sqrt{|100|}]$ (quinhentos)

A partir da natureza morfossintática dos numerais complexos do português, pode-se observar que eles são formados a partir de três processos: composição, coordenação e adjunção. Esses numerais são apresentados em (18), a seguir.

- Composição [N+N]: dezesseis, [...] dezenove, duzentos, trezentos, [...] novecentos.
- Coordenação: vinte e um, duzentos e dezesseis.
- Adjunção: dois mil, vinte mil, trezentos mil.

Primeiramente, vamos considerar os numerais formados por composição, o ponto central da análise dos numerais complexos. Os numerais formados por composição constituem duas séries distintas: de 16 a 19, e as centenas. Na primeira série, temos duas raízes transparentes na estrutura que são interpretadas em uma relação de adição, como em (19-a). No caso das centenas, também pode-se observar suas raízes, mas essas são interpretadas em uma relação de multiplicação, como em (19-b).

- (19) a. dezesseis (10+6), dezessete (10+7), dezoito (10+8) e dezenove (10+9)
b. duzentos (2x100), trezentos (3x100), [...] oitocentos (8x100), novecentos (9x100)

Vignado, J. Contando raízes: análise dos numerais cardinais do português na perspectiva da Morfologia Distribuída

compostos multiplicativos, uma raiz $\sqrt{|n|}$ e uma $\sqrt{|100|}$. A relevância das raízes que expressam cardinalidades relacionadas a dez não é surpreendente, posto que esses valores representam as bases do sistema decimal, que orienta o sistema numeral do português. Note que nem toda raiz é licenciada para formar numerais compostos. Veja que compostos como “*oitoeito” e “*dozENTOS” não são numerais em português, e se configuram como estruturas agramaticais. Além disso, a ordem é definida, o que impede estruturas como “*setedez” ou “*centoseis”⁸.

Do ponto de vista fonológico, temos “cem” por um lado, e “novecentos” por outro. Note que o português também conta com palavras como “centésimo”, “centena”, “centenário”. Trata-se de uma alomorfa da raiz, cuja forma fonológica *default* é /sét/, já que a forma /sẽ/ só se manifesta no numeral simples derivado dessa raiz. Nesse caso, temos os itens de vocabulário apresentados abaixo, em (20).

- (20) a. $\sqrt{|100|} \leftrightarrow /sét/$ (centena, duzentos)
 b. $\sqrt{|100|} \leftrightarrow /sẽ/ \text{ — } n[\sqrt{|100|} + \emptyset]$ (cem)

Esta análise defende que a restrição sobre as raízes licenciadas em numerais compostos é esperada diante do fato de que o português emprega o sistema decimal. Hurford (1987) propõe que numerais que contêm a operação de multiplicação sejam formados a partir de bases morfêmicas multiplicativas que refletem a base numérica de um sistema. Ou seja, no caso de um sistema decimal, as bases morfêmicas seriam os múltiplos de 10, como dez (10^1), cem (10^2) e mil (10^3), fato que se reflete nos numerais do português.⁹ Argumento que tais bases também constituem raízes com conteúdo semântico cardinal.

FLP 27(2)

Para propor uma derivação para os cardinais complexos, a análise adota a definição de composto de Nóbrega (2014), em que um composto é formado quando duas ou mais raízes categorizadas, concatenadas em uma determinada relação gramatical — subordinação, atribuição ou coordenação —, são recategorizados por um núcleo categorial.

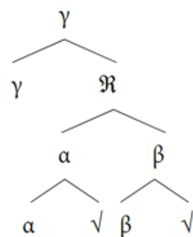
Desse modo, assume-se que os numerais compostos têm a mesma estrutura, em (21), dos demais compostos do português brasileiro (*cf.* Nóbrega, 2014), em que α , β e γ são categorizadores; R determina a relação gramatical entre α , β e γ , recategoriza a

⁸Veja que a ordem é fixa e característica desses numerais; a expressão “sete e dez”, por exemplo, é interpretada como um horário, e não como um cardinal.

⁹Evidentemente, esse raciocínio permite imaginar bases muito maiores como 10^{231} ; contudo, é preciso destacar que numeral e número não dizem respeito aos mesmos objetos científicos, como aponta Vignado (2019). A língua nomeia números conforme a necessidade de seu uso. Apesar disso, ainda que pela lógica recursiva os numerais sejam potencialmente infinitos, como os números, eles não o são pelas limitações impostas por (i) questões cognitivas, como memória e processamento, e (ii) extralinguísticas, como cultura e sociedade. Um exemplo são os sistemas numerais restritos de línguas que nomeiam seus numerais até valores bem menores do que o português, como as línguas nativas sul-americanas karitiana e avá guarani. Essas línguas têm numerais até 20, sendo que, a partir de 5, utilizam uma estratégia baseada em associar partes do corpo a valores cardinais (Vignado; Frutos, 2024). Outro exemplo dessa distinção são as irregularidades de um sistema numeral, visto que sistemas numéricos são regulares por natureza. O francês adota o sistema decimal, mas exhibe numerais irregulares como “*quatre vingt*” ‘oitenta’ — lit.: ‘quatro vinte’ —, em que emprega a base numérica 20 em uma estrutura de multiplicação, e não 10, como esperado.

estrutura e garante propriedades ligadas à integridade lexical, conforme o autor.

(21)



(exemplo adaptado de Nóbrega, 2014, p. 110)

Posteriormente, a estrutura é enviada para o processo de SPELL OUT, onde é enviada para PF e LF. Em PF, vogais temáticas são adjungidas aos categorizadores e ocorre a inserção de itens de vocabulário.

Diante da proposta de que numerais sejam, em um primeiro ciclo, categorizados por um nominalizador, numerais complexos são tratados aqui como compostos [N+N]. De acordo com Nóbrega (2014, p. 82), a relação interna entre os elementos de um composto [N+N] pode ser de três tipos: (i) subordinação, quando um nome seleciona outro a partir de uma “relação de” (ex. “aracnofobia”); (ii) atribuição, quando um nome não núcleo modifica o nome núcleo (ex. “peixe-espada”); ou de coordenação, quando os nomes estão numa relação de conjunção (ex. “ator-diretor”).

Tomemos os tipos propostos por Nóbrega para pensar os numerais complexos formados por composição. Os numerais com interpretação aditiva apresentam — de forma bastante evidente — uma relação de coordenação, dado que a natureza lógica da conjunção também leva a uma interpretação aditiva. Quanto aos numerais complexos com interpretação multiplicativa, não se trata de uma relação de subordinação. Estabelecendo um paralelo com “aracnofobia”, por exemplo, que consiste em uma ‘fobia de aranha’, quando pensamos em um numeral como “duzentos”, seu significado não é ‘centos de dois’ ou ‘dois de centos’, assim, não há nesse tipo de numeral uma relação de subordinação onde um nome seleciona outro. Dessa forma, considero que esses numerais exibem uma relação de atribuição, em que o nome núcleo é aquele que representa a base do sistema, i.e., “cento”, que é modificado por outro nome que representa uma unidade do sistema. Assim, a modificação imposta pela relação de atribuição é interpretada como multiplicação em compostos cujas raízes têm conteúdo semântico cardinal. A literatura sobre compostos do mostra que a modificação nesse caso se dá pela relação de atribuição (*cf.* Bisetto; Scalise, 2005; Nóbrega, 2014).

Dito de outra forma, nos dados em (19), a mesma estrutura gera numerais com interpretação aditiva, “dezessete”, e numerais com interpretação multiplicativa, “quatrocentos”. Isso levanta a questão: o que diferencia uma interpretação da outra? Nesta análise a interpretação aditiva desses compostos em LF advém de uma relação gramatical de coordenação, em que a projeção funcional R de uma conjunção abstrata (&) toma dois constituintes categorialmente idênticos, α e β , em uma estrutura simétrica. Da mesma forma, a interpretação multiplicativa em LF advém da relação gramatical de atribuição, que consiste em uma relação de modificação em que um elemento não núcleo β é adjungido ao núcleo α , gerando uma estrutura assimétrica de adjunção.

Tal análise se alinha à proposta clássica — e lexicalista — de Hurford (1987) em que numerais complexos são formados por coordenação (aditivos) ou por modificação (multiplicativos). Assim, a interpretação é gerada pela relação gramatical interna ao composto, e não pela postulação de traços [ADD] e [MULT], como proposto por Vignado (2023), o que torna a derivação mais econômica, e alinhada à literatura sobre compostos no português.

Na exposição sobre os numerais simples, vimos que há categorizadores n que derivam os numerais em português. O categorizador n ($\emptyset$) também é parte da estrutura dos numerais formados por composição. No contexto do composto, proponho que as instruções de interpretação contidas na Lista 3 sejam responsáveis pela interpretação de multiplicação ou de adição.

(22) Instrução para Lista 3

- a. $R_{coord}[\sqrt{|n|} + \sqrt{|n|}] \leftrightarrow$ interprete como $n + 10$, $|n| \in 1,2,3,4,5,6,7,8,9$
- b. $R_{atrib}[\sqrt{|n|} + \sqrt{|100|}] \leftrightarrow$ interprete como $n \times 100$, $|n| \in 1,2,3,4,5,6,7,8,9$

O mecanismo proposto, que atribui ao conhecimento enciclopédico em L3 à interpretação adequada para os numerais compostos, é uma forma de capturar o fenômeno sem precisar incluir na sintaxe noções que são matemáticas, seja através de traços que veiculem tais operações, seja por questões meramente estruturais.

Veja que uma análise como a de Hurford (1987), em que há uma estrutura sintática aditiva e uma estrutura multiplicativa, carece de uma explicação sobre onde, na arquitetura da gramática, ocorre a interpretação adequada dessas estruturas, já que também ocorrem em outros contextos — como a coordenação, por exemplo, que não é exclusiva aos numerais. Já uma análise como a de Zabbal (2005) implica a assunção de traços — [ADD] e [MULT] — que se justificam somente pela interpretação da relação de adição ou multiplicação entre as partes do numeral complexo. Tais operações não são estritamente linguísticas, mas refletem um conhecimento que pertence a outro domínio: o matemático. Nesse sentido, é natural que esse conhecimento esteja alocado no conhecimento enciclopédico do falante, e não em feixes de traços específicos contidos na Lista 1 a priori.

Por fim, há ainda os numerais complexos que não são compostos, mas formados por formas livres em um sintagma, *i.e.*, aqueles formados por coordenação e adjunção de numerais. Nesse caso, cada numeral do sintagma tem acento primário próprio e pode ser realizado sozinho. Numerais complexos desse tipo com interpretação de adição, como “trinta e dois”, têm uma conjunção ‘e’ pronunciada. Note que, diferentemente dos compostos, não se trata de uma conjunção abstrata marcada por uma relação R, mas de uma conjunção de fato, confirmando sua natureza sintagmática. Assim, proponho que se trata de estruturas de coordenação entre nPs, [nP e nP]. A interpretação aditiva desses numerais é produto da natureza semântica das estruturas de coordenação. Numerais como “dois mil”, por sua vez, consistem em uma estrutura de adjunção de nPs, [nP[nP]] e a interpretação multiplicativa é resultado da relação de modificação imposta pela estrutura.

FLP 27(2)

5 CONSIDERAÇÕES FINAIS

Este artigo buscou descrever e analisar, sob a perspectiva da Morfologia Distribuída, os numerais do português, que constituem um conjunto de dados pouco explorado em abordagens formais. Para isso, foram apresentados dados de numerais cardinais, simples e complexos, a fim de apresentar o panorama completo dessas expressões em português. Argumentou-se que a constância do significado cardinal presente nas raízes em expressões derivadas e as restrições extralinguísticas que advêm do conhecimento enciclopédico sobre o sistema decimal servem como evidência para análise de que essas raízes têm conteúdo semântico que consiste em uma cardinalidade definida.

O modelo proposto considera que o português tem raízes com conteúdo cardinal para os valores 1-15, 100 e 1000. A partir da classificação de Alexiadou e Lohndal (2017), retomando as questões iniciais, nota-se que o significado das raízes com conteúdo semântico cardinal é estrito, e não abrangente, já que é regular e consistente nos outros tipos de numerais e nas palavras derivadas (ex. “três”, “triplo”, “trigêmeo”, “tridente”, “tripartite”).

Ainda no sentido da defesa do conteúdo semântico das raízes, Minussi e Bassani (2017) mostram que não considerar que há algum tipo de semântica que individualize as raízes já na Lista 1 tem consequências teóricas importantes. Para os autores, isso implica:

- i) a perda da possibilidade de explorar teórica e empiricamente as restrições que a semântica pode impor sobre as raízes, uma vez que há algum grau de restrição semântica sobre o contexto morfossintático em que uma raiz pode ocorrer — como observamos no caso dos numerais; e
- ii) a retirada do conteúdo conceitual da Lista 1 significa que esse conteúdo deve ser alocado em outro lugar na gramática, a Lista 3, que, dessa forma, é sobrecarregada com regras de correspondência.

Os dados apresentados mostraram que há restrições relacionadas às raízes na formação de numerais complexos que refletem o conhecimento enciclopédico sobre o sistema numérico decimal e como os números se constroem sobre essa lógica. Veja que a língua emprega mecanismos linguísticos que, aliados ao conhecimento sobre o sistema decimal, norteiam a organização do sistema numeral e a derivação e a interpretação dos numerais. Assim, buscou-se capturar na análise esse fenômeno.

Conforme foi exposto, a raiz concatenada a um categorizador específico influencia o significado da palavra formada, especificando um domínio de interpretação idiossincrática na Lista 3, que diferencia valores cardinais de numerais simples derivados por categorizadores diferentes — que se realizam como sufixos derivacionais diferentes. Assim, esta análise defende que atribuir à Lista 3 interpretações idiossincráticas e restrições extralinguísticas que se aplicam sobre os cardinais constitui uma abordagem produtiva, no sentido de esclarecer como se dá a influência do conhecimento sobre o funcionamento do sistema numérico decimal sobre a formação do sistema numeral de uma língua natural.

A expressão de uma cardinalidade maior, seja no domínio dos números naturais ou no domínio dos numerais, é potencializada pelo estabelecimento de bases, como a

FLP 27(2)

base decimal. Retomando Guerreiro *et al.* (2020), numerais refletem o ‘princípio da cardinalidade’, o fato de que expressões linguísticas são associadas a valores cardinais definidos. Na proposta apresentada aqui, tais valores cardinais definidos dependem de semântica cardinal em raízes que representam unidades e bases relevantes para o sistema; da estrutura sintática, do tipo de relação R (em compostos); e dos conhecimentos extralinguísticos que regem as restrições que se aplicam sobre essas expressões.

Por fim, a análise segue Rothstein (2017) na assunção de que numerais denotam propriedades, como os demais nomes, mais especificamente propriedades cardinais, e propõe que as raízes que contêm semântica cardinal possam ser categorizadas por um nominalizador, quando numerais argumentais; ou recategorizadas por um adjetivizador, quando são numerais predicativos. A recategorização é um mecanismo previsto pela Morfologia Distribuída que, no caso dos numerais, captura o mesmo fenômeno que Rothstein (2017) explica como uma operação semântica de mudança de tipo. Assim, a análise permite explicar por que numerais cardinais argumentais e predicativos têm comportamentos morfossintáticos e semânticos diferentes sem recorrer a uma operação semântica abstrata.

Para pesquisas futuras sobre os cardinais do português, é necessário explorar mais a questão da marcação de gênero, a fim de elucidar como se dá a marcação de gêneros em numerais predicativos, especialmente em numerais complexos, que exibem variações como “duzentos” e “duas meninas” e “duzentas e duas meninas”, por exemplo. É preciso também propor uma descrição mais detalhada dos aspectos fonológicos. De todo modo, para os fins propostos neste artigo, buscou-se apresentar os pontos mais relevantes sobre tais temas.

FLP 27(2)

Recebido em Novembro 2025

Publicado em Janeiro 2026

REFERÊNCIAS

ALEXIADOU, A. **Functional structure in nominals**: nominalization and ergativity. Amsterdam; Philadelphia: John Benjamins, 2001.

ALEXIADOU, A.; LOHNDAL, T. The structural configurations of root categorization. *In*: BAUKE, L.; BLÜMEL, A. (eds.). **Labels and roots**. Walter de Gruyter GmbH & Co KG: Boston/Berlin, 2017. p. 203-232.

BARWISE, J.; COOPER, R. Generalized quantifiers and natural language. *In*: KULAS, J.; FETZER, J. H.; RANKIN, T. L. (eds.). **Philosophy, language, and artificial intelligence**: resources for processing natural language. Dordrecht, the Netherlands: Springer, 1981. p. 241–301.

BISETTO, A.; SCALISE, S. The classification of compounds. **Lingue e Linguaggio**, v. 4, n. 2, p. 319-332, 2005.

BOOIJ, G. Constructions and lexical units: An analysis of Dutch numerals. *In*: OLSEN, S. **New impulses in word formation**. Hamburg: Helmut Buske Verlag, 2010. p. 81–100.

BYLININA, L.; NOUWEN, R. Numeral semantics. **Language and Linguistics Compass**, v. 14, n. 8, 2020.

COMRIE, B. Numeral Bases. *In*: DRYER, M. S.; HASPELMATH, M. (eds.) **The World Atlas of Language Structures Online**. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2005. n.p.

EMBICK, D.; MARANTZ, A. Architecture and blocking. **Linguistic Inquiry**, Cambridge, MA, v. 39, n. 1, p. 1–53, 2008.

FERREIRA, M. Sintagmas Cardinais Complexos e Implicaturas Escalares. **Revista Letras**, v. 75/76, p. 197–212, 2008.

FERREIRA, M. Tópicos contrastivos com sintagmas cardinais complexos. **Revista de Estudos da Linguagem**, v. 20, n. 1, p. 49-64, 2012.

GUERRERO, D. *et al.* Is thirty-two three tens and two ones? The embedded structure of cardinal numbers. **Cognition**, v. 203, p. 104331, 2020.

HALLE, M. Distributed Morphology: impoverishment and fission. *In*: LECARME, J.; LOWENSTEIN, J.; SHLONSKY, U. (eds.) **Research in Afroasiatic grammar: Papers from the third Conference on Afroasiatic Languages**. Sophia-Antipolis: John Benjamins, 1996.

HALLE, M.; MARANTZ, A. Some key features of Distributed Morphology. *In*: CARNIE, A.; HARLEY, H.; BURES, T. (eds.). **MIT working papers in Linguistics: Papers on Phonology and Morphology**, Cambridge, n. 21, p. 275-288, 1994.

HAMMARSTRÖM, H. Rarities in numeral systems. *In*: WOHLGEMUTH, J.; CYSOUW M. (eds.) **Rethinking universals**. How rarities affect linguistic theory. Berlin: De Gruyter, 2010. p. 11-60.

HARLEY, H.; NOYER, R. State-of-the-article: Distributed Morphology. **Glott International**, v. 4, n. 4, p. 3-9, 1999.

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

HURFORD, J. R. **Language and number: the emergence of a cognitive system**. Oxford: Blackwell, 1987.

HURFORD, J. R. A. performed practice explains a linguistic universal: Counting gives the Packing Strategy. **Lingua**, v. 117, n. 5, p. 773-783, 2006.

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

LANDMAN, F. Predicate-argument mismatches and the adjectival theory of indefinites. *In*: COENE, M.; D'HULST, Y. (eds.). **From NP to DP: The syntax and semantics of noun phrases**. Amsterdam, the Netherlands: John Benjamins, 2003, p. 211-238.

MARANTZ, A. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. **University of Pennsylvania Working Papers in**

Linguistics, v. 4, n. 2, p. 221-225, 1997.

MARANTZ, A. **Words and things**. (Manuscrito). 2001.

NÓBREGA, V. A. **Tópicos em composição: estrutura, formação e acento**. 2014. Dissertação (Mestrado em Letras) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2014.

PRIM, C. de S. Os adjetivos exclusivamente pré-nominais do português. **Signum: Estudos da Linguagem**, v. 18, n. 2, p. 377-403, 2015.

RESENDE, M. S.; SANTANA, B. P. A relação entre raízes, gênero, classe e significado. **Revista da ABRALIN**, s.l., v. 18, n. 1, 2019.

ROTHSTEIN, S. A Fregean semantics for number words. *In*: ALONI, M.; FRANKE, M.; ROELOFSEN, F. (eds.) **Proceedings of the 19th Amsterdam Colloquium**. Amsterdam: Institute for Logic, Language, and Computation, 2013. p. 179–86.

ROTHSTEIN, S. **Semantics for counting and measuring**. Cambridge: Cambridge University Press, 2017.

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

VIGNADO, J. N. Roots, cardinality, and syntax: an analysis of numerals in Portuguese. *In*: DM 30 – 30 years distributing morphology from north to south, 2023, São Paulo: Universidade de São Paulo. **Apresentação de trabalho**, 2023.

VIGNADO, J. N.; FRUTOS, L. Um estudo comparativo entre os numerais do Karitiana e do Avá Guaraní. *In*: FERREIRA, L. F.; FRUTOS, L.; COELHO, O. (orgs.). **Jornadas pelos significados: contribuições de Ana Müller para a semântica**. Campinas: Pontes Editores, 2024. p. 104-144.

ZABBAL, Y. **The syntax of numeral expressions**. Manuscrito, University of Massachusetts, Amherst, 2005.

FLP 27(2)