UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE LINGUÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM SEMIÓTICA E LINGUÍSTICA GERAL

JULIANA VIGNADO NASCIMENTO

A interface sintática e semântica na análise dos sistemas numerais
do karitiana e do kamayurá

Versão corrigida

São Paulo

2019

JULIANA VIGNADO NASCIMENTO

A interface sintática e semântica na análise dos sistemas numerais
do karitiana e do kamayurá

Dissertação apresentada ao Programa de Pós-Graduação
em Semiótica e Linguística Geral do Departamento de
Linguística da Faculdade de Filosofia, Letras e Ciências
humanas da Universidade de São Paulo para obtenção do
título de mestre.

Versão corrigida

São Paulo

2019

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na Publicação
Serviço de Biblioteca e Documentação
Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

V784i Vignado, Juliana N.
A interface sintática e semântica na análise dos
sistemas numerais do karitiana e do kamayurá /
Juliana N. Vignado ; orientadora Ana Müller. - São
Paulo, 2019.
125 f.

Dissertação (Mestrado)- Faculdade de Filosofia,
Letras e Ciências Humanas da Universidade de São
Paulo. Departamento de Linguística. Área de
concentração: Semiótica e Linguística Geral.

1. LINGUÍSTICA. 2. NUMERAL CARDINAL. 3. SEMÂNTICA
FORMAL. 4. SINTAXE E SEMÂNTICA DA LINGUAGEM NATURAL.
5. TUPI-GUARANI. I. Müller, Ana, orient. II. Título.

ENTREGA DO EXEMPLAR CORRIGIDO DA DISSERTAÇÃO/TESE

Termo de Ciência e Concordância do (a) orientador (a)

Nome do (a) aluno (a): Juliana Vignado Nascimento

Data da defesa: 04/06/2019

Nome do Prof. (a) orientador (a): Ana Lúcia de Paula Müller

Nos termos da legislação vigente, declaro **ESTAR CIENTE** do conteúdo deste **EXEMPLAR CORRIGIDO** elaborado em atenção às sugestões dos membros da comissão Julgadora na sessão de defesa do trabalho, manifestando-me **plenamente favorável** ao seu encaminhamento e publicação no **Portal Digital de Teses da USP**.

São Paulo, 05/08/2019

(Assinatura do (a) orientador (a))

JULIANA VIGNADO NASCIMENTO

A interface sintática e semântica na análise dos sistemas numerais do karitiana e do kamayurá

Dissertação apresentada ao Programa de Pós-Graduação
em Semiótica e Linguística Geral do Departamento de
Linguística da Faculdade de Filosofia, Letras e Ciências
humanas da Universidade de São Paulo para obtenção do
título de mestre.

São Paulo, 4 de Junho de 2019

BANCA EXAMINADORA

Prof.^a Dr.^a Ana Lucia de Paula Müller
Orientadora

Prof.^a Dr.^a Maria Filomena Spatti Sandalo
Universidade Estadual de Campinas

Prof.^a Dr.^a Suzi Lima
University of Toronto

Prof. Dr. Thomas Daniel Finbow
Universidade de São Paulo

São Paulo,

2019

“God made integers, all else is the work of man”

Famoso aforismo do matemático Leopold Kronecker (1823- †1891)

“One possibility is that the natural numbers result from a simple constraint on the language faculty, hence are not given by God (...)”

Chomsky (2005)

Agradecimentos

Antes de tudo, preciso frisar que a ordem dos agradecimentos aqui não implica em uma ordem de importância, implica apenas na organização dos agradecimentos que são muitos e necessários. Assim, primeiro serão apresentados os agradecimentos aos envolvidos no processo da pesquisa e em seguida os agradecimentos aos meus familiares.

É impossível contemplar a todas as pessoas que passaram pelo meu caminho durante o desenvolvimento desta dissertação, mas eu gostaria de poder agradecer a todas as moças da cantina que me serviram café, comida e conversa nos meus dias inteiros no prédio da Letras, aos moços do xerox que seguraram aquele fiado de impressão em dias com muita pressa e pouco dinheiro, aos motoristas do fretado que semanalmente me levaram com segurança de Campinas a São Paulo e a tantos outros trabalhadores que com sua força de trabalho permitiram que eu chegasse até aqui. Obrigada!

Gostaria também de fazer uma homenagem especial a linguista brasileira Lucy Seki que faleceu em 2017, o ano do início desta pesquisa. A dedicação incansável com que essa importante linguista brasileira empreendeu durante toda sua vida para a descrição e análise de línguas indígenas brasileiras, entre elas o kamayurá, foi o que possibilitou parte da investigação aqui empreendida. Ademais, o fato de uma mulher ter sido a autora de uma das primeiras gramáticas de língua indígenas escritas por um linguista brasileiro é fonte de grande inspiração e admiração de minha parte. Eu agradeço, portanto, a Lucy Seki que foi uma das pioneiras no campo da descrição e análise de línguas indígenas e a tantas outras linguistas que realizaram seu trabalho com muito empenho, dedicação e resistência e abriram o caminho para que jovens mulheres e linguistas, como eu, pudessem realizar suas pesquisas.

Agradeço ao CNPq que concedeu a mim duas bolsas de iniciação científica e uma bolsa de mestrado que possibilitaram a realização dessa pesquisa. O auxílio financeiro fornecido por essa instituição é um recurso valioso para o desenvolvimento da pesquisa científica do Brasil. Agradeço também a Universidade de São Paulo que concedeu a mim uma bolsa de iniciação científica. Os programas de bolsa da universidade são bastante importantes para o desenvolvimento de pesquisas científicas

por estudantes da graduação. É graças a esses auxílios que muitos jovens, como eu, iniciam sua carreira acadêmica e conseguem se manter na universidade, por isso, sua importância merece ser ressaltada.

Agradeço a minha querida orientadora Ana Müller pela orientação prestada durante os últimos seis anos. Ao longo de três iniciações científicas e do mestrado a professora Ana me ensinou e me formou com muito carinho e dedicação. Serei eternamente grata pelas correções minuciosas, pelas broncas, pelas motivações e por todo apoio e compreensão que recebi quando precisei. Agradeço por encarar comigo o desafio de trabalhar com a formação de sistemas numerais em línguas tupi, um tema desafiador e ainda pouco explorado pela linguística formal. Agradeço a minha orientadora que com cuidado e muita paciência me iniciou e me conduziu pelas veredas da investigação científica. Obrigada, professora, nunca me esquecerei da energia que você dedicou a minha formação como pessoa e como pesquisadora.

A orientação da Ana não me rendeu apenas boas diretrizes de como exercer o ofício de pesquisadora, mas também me inseriu em uma “família acadêmica” incrível e inspiradora. Junto com o GELI – Grupo de Estudos em Línguas Indígenas – aprendi não somente sobre linguística e línguas indígenas, mas também sobre cooperação, organização e ética no fazer científico. Fiz boas amizades que carregarei para a vida toda, fui a congressos intelectualmente inspiradores, organizei eventos científicos e coletas de dados, viajei e compartilhei boas conversas e experiências com os amigos que o grupo me trouxe.

Agradeço aos professores Thomas Finbow e Luciana Storto pelos ensinamentos valiosos sobre línguas indígenas sul-americanas e por estarem sempre dispostos a me ajudar com o desenvolvimento da pesquisa e resolução de problemas descritivos. Agradeço ao professor Marcelo Barra pelas lições em semântica formal e pelas preciosas contribuições a esse trabalho. Agradeço imensamente a oportunidade de contar com professores tão competentes, acessíveis e bacanas na minha formação. Obrigada, professores, vocês sempre serão uma inspiração no exercício da minha carreira profissional.

Agradeço às professoras Luciana Sanchez e Lara Frutos que foram minhas mentoras na semântica formal e minhas amigas ao longo desses anos. Essas semanticistas incríveis foram, e ainda são, fundamentais para minha formação identitária enquanto

mulher e pesquisadora. Agradeço por sempre me lembrarem que lugar de mulher é na ciência e onde ela quiser! Também agradeço todas as risadas, todas as ajudas e todos os conselhos. É uma honra poder contar com amigas tão competentes, divertidas e inspiradoras. Lu e Lara, vocês são mulheres maravilhosas!

Agradeço especialmente aos meus “irmãos” de pesquisa Fernando Ferreira, Lucas Ciola e Thiago Chaves sempre presentes e prontos para me consolarem nos momentos difíceis e me alegrarem em todos os momentos. Meninos, vocês foram um presente na minha vida. Não sei o que seria de mim sem vocês, de verdade, e é difícil expressar todo o meu carinho e amor por vocês nessas linhas. Obrigada pelas piadas, pelas viagens, pelas conversas e pela linguística terapêutica nossa de todos os dias! Vocês sempre estarão no meu coração e nos meus bons pensamentos, meus irmãozinhos!

Agradeço também aos meus estimados “primos” de pesquisa (nossa família é grande e bonita!). Agradeço a Tarcísio Dias, Karin Vivanco e Ivan Rocha que com seus conhecimentos sintáticos sempre me ajudaram no esclarecimento de dúvidas de sintaxe e de língua karitiana. Serei sempre grata pelos cafés, pelos papos e pelas contribuições valiosas. Obrigada por me lembrarem a importância da sintaxe ao longo desses anos!

Agradeço também ao “grupinho do mestrado” que me divertiu e confortou durante a pesquisa. Agradeço especialmente à Raissa Santana pelas dicas de ouro em organização e planejamento e a Jessica Vianna pelas ajudas em semântica formal. Adicionalmente, também agradeço a elas, a Claudia Souza, a Camila Silvestre, a Marcela Guedes, a Iolanda Dias e ao Tarcísio por todos os cafés, desabafos e risadas nessa nossa experiência compartilhada de mestrado. Foi uma honra crescer com vocês, companheiros de caminhada!

Preciso agradecer, é claro, à Érica, a melhor secretária de pós-graduação que um departamento de linguística poderia ter, que, com sua firmeza e organização, mantém sólidas as bases do nosso departamento. Obrigada, Érica, por me ajudar em tantas e tantas dúvidas que eu tive ao longo dos anos e pela dedicação incansável ao departamento, você arrasa!

Agradeço também minha família que ofereceu suporte e amor nos últimos anos para que eu lograsse cumprir essa jornada. Agradeço especialmente a minha mãe, minha companheira incansável durante a vida. Sua determinação, coragem e força me inspiram a sempre tentar ser melhor e a seguir meus sonhos. Agradeço a minha mãe que me criou

no ambiente universitário, o que certamente me motivou na busca pelo conhecimento com humanidade e respeito pelo ofício. Obrigada, mamãe, por me acalantar nos momentos mais difíceis, por fornecer condições práticas (e financeiras) para que eu desenvolvesse minha pesquisa no estado de São Paulo e participasse dos congressos úteis à minha formação e, por fim, por me amar incondicionalmente apesar dos desafios que enfrentamos durante esse período. Obrigada por ser meu porto seguro. Sempre vou te amar ‘infinito mais um’!

Agradeço ao meu pai, neto de bororos e melhor pedagogo que eu conheço, que desde de pequena me estimulou a ler revistinhas e livrinhos. Ao longo da vida nossos momentos sempre foram nossas leituras, nossos debates e nossas viagens juntos. Se eu cheguei até as letras e a linguística foi graças ao hábito de leitura que meu pai estimulou desde tão cedo. Agradeço ao meu pai por me ensinar valores tão necessários como humildade e paciência para que eu pudesse seguir na caminhada sem desistir. Sem suas palavras sábias nos momentos certos, as coisas seriam bem mais difíceis. Me sinto honrada por ter ancestrais indígenas, e isso me inspira grande respeito e uma profunda admiração por essas populações. Obrigada, papai, por ser meu pai-professor-pantaneiro e me formar com tanto amor e carinho, me ensinando sempre a observar com amor e respeitar todo aquele que é diferente de mim.

Agradeço a mamãe e papai, que mesmo estando tão longe, lá no Mato Grosso, estiveram presentes na minha jornada através da pesquisa na Universidade de São Paulo. Agradeço aos meus pais-professores por me ensinarem o apreço pela universidade reforçando sempre a importância de uma universidade pública, universal e acessível. Obrigada por me ensinarem o valor do conhecimento e da educação. Estar longe da família em uma universidade de renome foi um desafio. Sem o apoio e amor de vocês isso tudo não seria possível. Obrigada por compreenderem quem eu sou e por me motivarem a nunca desistir.

Não poderia deixar de agradecer meu núcleo familiar no estado de São Paulo que foi meu suporte e porto seguro sempre que precisei. Agradeço especialmente a minha madrinha Joceli e às minhas primas Monique e Mariana. Esse trio é um time incrível que me ensina sobre companheirismo, amor e família. Agradeço todas as vezes que me levaram ao médico quando eu estava doente, que me levaram para passear quando eu estava cansada, que me deram tantas e tantas caronas facilitando a minha rotina morando sozinha em Campinas. Agradeço por sempre me lembrarem com palavras

carinhosas e motivadoras que eu sou boa no que eu faço e que tenho sido forte. Eu amo vocês!

Por fim, agradeço à vovó Maria, a mulher mais forte que eu conheço. Sua vida foi repleta de desafios, mas a dona Maria nunca desistiu e sempre seguiu em frente. Criou quatro filhos sendo viúva e com pouca instrução, mas com muito trabalho e muita inteligência deu suporte para que minha mãe seguisse a carreira universitária, o que me estimulou a buscar o mesmo caminho. Obrigada, vovó, por me lembrar que não se pode desistir e que a vida é dura, mas que é preciso ter coragem e seguir em frente.

Agradeço a todos que acreditaram em mim. Agradeço a todos os amigos que me deram suporte afetivo ao longo desses anos, a todos que me mandaram energias boas e palavras de conforto quando precisei. Agradeço todas as rezas, orações, fezinhas, boas energias e boas palavras que foram dedicadas ao meu sucesso na realização do mestrado.

Por último, e não menos importante, gostaria de agradecer a mim. Gostaria de agradecer a mim por acreditar em mim e por todo trabalho duro. Gostaria de agradecer a mim por nunca desistir e por tentar fazer o certo. Gostaria de agradecer a mim por ser quem eu sou, apesar dos pesares das imperfeições. Por fim, gostaria de agradecer a mim por ser uma sonhadora incurável e tornar possível esta dissertação!

RESUMO

VIGNADO, J. N. **A interface sintática e semântica na análise dos sistemas numerais do karitiana e do kamayurá.** 133f. 2019. Dissertação (Mestrado) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2019.

Esta pesquisa investiga interface sintática e semântica das construções que expressam números em duas línguas da família linguística Tupi. Os objetivos são i) propor uma análise da interface sintática e semântica da estrutura composicional dos numerais nas línguas Tupi e ii) avaliar o impacto dessa análise para a compreensão da morfossintaxe dos sistemas numéricos em geral. Foram selecionados os sistemas numéricos das línguas do karitiana (Arikém) e do kaïamurá (Tupi-Guarani). No caso do karitiana, os dados foram coletados pela autora deste projeto em uma Iniciação científica financiada pelo CNPq em 2015. No caso do kamayurá a fonte dos dados foi a Gramática do Kamayurá (SEKI, 2000). A pesquisa toma como ponto de partida as seguintes teorias: i) Estratégia de empacotamento proposta por Hurford (1987, 2006, 2010) para explicar a sintaxe da formação dos numerais nas línguas naturais. Nessa abordagem operações sintáticas expressam operações matemáticas, como multiplicação e adição, e numerais complexos são formados a partir de numerais e bases morfológicas multiplicativas; ii) teoria Fregeana para palavras numéricas desenvolvida por Rothstein (2013, 2017) que propõe que numerais complexos são formados a partir de numerais básicos <n> e multiplicadores lexicais <n <e, t>. As hipóteses iniciais foram: i) a estratégia de empacotamento opera na formação recursiva de numerais complexos em línguas Tupi e isso dá apoio empírico à tese de que esse mecanismo é um universal sintático que orienta a formação dos sistemas numéricos nas línguas naturais; ii) o conceito de base morfológica multiplicativa (HURFORD, 1987,2006,2010) é compatível com o conceito de léxico multiplicativo (ROTHSTEIN, 2013). As hipóteses foram confirmadas e a análise dos dados mostrou que: i) ambas línguas expressam os numerais compostos via numerais básicos, bases auxiliares e coordenação, ii) ambas línguas apresentam estruturas aditivas nos numerais compostos e iii) a abordagem foi produtiva para elucidar a composição desses sistemas numerais.

Palavras-chaves: Numeral, Semântica, Sintaxe, Interface, Tupi.

ABSTRACT

VIGNADO, J. N. **The syntactic and semantic interface in the analysis of Karitiana and Kamayurá numeral systems.** 2019. 133f. Thesis (Master) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2019.

This research aims to investigate the syntactic and semantic interface of the numerical expressions in two indigenous languages of different subfamilies of the Tupi linguistic family. Our goals are i) to propose an analysis of the syntactic and semantic interface of the numerical compositional structure in the Tupi languages and ii) to evaluate the impact of this analysis for the comprehension of the morphosyntax of numerical systems in general. The numeral systems analyzed of the languages Karitiana (Arikém), Kaiamurá (Tupi-Guarani). In the case of Karitiana, data were collected by the author of this project in 2015 in a Scientific Initiation funded by CNPq. For kamayurá, the source of data was the Kamyurá Grammar (SEKI, 2010). The research takes as its starting point the following theories: i) Packing strategy proposed by Hurford (1987, 2006, 2010) to explain the syntax of numerals formation in natural languages. In this approach, syntactic operations express mathematical operations, such as multiplication and addition and complex numerals are formed from a basic numeral and a multiplicative morphemic base; ii) Fregean theory for numerical words developed by Rothstein (2013, 2017) which proposes that complex numerals are formed from basic numerals <n> and lexical multipliers <n <e, t>. The initial hypothesis was: i) the packaging strategy operates on the recursive formation of complex numerals in Tupi languages and this gives empirical support to the thesis that this mechanism is a syntactic universal that guides the formation of numerical systems in natural languages; ii) the concept of multiplicative morphemic base (Hurford, 1987, 2006, 2010) is compatible with the concept of lexical multiplier (Rothstein 2013). The hypotheses were confirmed and the data analysis showed that: i) both languages express the compound numerals via basic numerals, auxiliary bases and coordination, ii) both languages have additive structures in the compound numerals and iii) the approach was productive to elucidate the composition of these numerical systems.

Key-words: Numeral, Semantics, Syntax, Interface, Tupi.

LISTA DE TABELAS

Tabela 1 – Regras sintáticas na formação de numerais

Tabela 2 – Exemplos das regras sintáticas na formação de numerais

Tabela 3 – Numerais básicos do karitiana

Tabela 4 – Numerais de 6 a 10 em karitiana

Tabela 5 – Numerais de 11 a 15 em karitiana

Tabela 6 – Numerais de 16 a 20 em karitiana

Tabela 7 – Paradigma da formação de numerais em karitiana

Tabela 8 – Bases morfológicas do karitiana (VIGNADO, 2017)

Tabela 9 – Palavras e morfemas que compõe numerais kamayurá

Tabela 10 – Numerais básicos em kamayurá

Tabela 11– Numerais de 6 a 10 em kamayurá

Tabela 12 – Numerais de 11 a 15 em kamayurá

Tabela 13 – Numerais de 16 a 20 em Kamayurá

Tabela 14 – Paradigma da formação dos numerais em kamayurá

Tabela 15 – Bases morfológicas do kamayurá

Tabela 16 – Acabar, karitiana vs kamayurá

LISTA DE ABREVIACÕES

1p	Primeira pessoa do plural
ADVZ	Morfema Adverbializador
CAUS	Morfema causativo
COP	Cópula
DECL	Morfema de modo declarativo
ENF	Morfema enfático
FUT	Futuro
NFUT	Morfema de não futuro
VT	Vogal temática
PART	Particípio
POSP	Posposição

FORMA DE APRESENTAÇÃO DOS DADOS

Os dados com suas respectivas glosas serão apresentados conforme 1:

1) *Descrição do dado (quando houver)*

Segmentação morfêmica

Glosa

‘Tradução em português’

(Referência, Ano, página, exemplo)

No corpo do texto, os dados serão apresentados em sua forma ortográfica e em itálico. Todo dado de outros pesquisadores será acompanhado de sua referência, enquanto os dados coletados por mim não virá acompanhado por referência. Quando conveniente, vai-se uniformizar as glosas ou assumir outra glosa conforme as discussões apresentadas.

O asterisco (*) indica que o dado em questão é agramatical. Já o símbolo do jogo-da-velha (#) corresponde a um dado gramatical que é semanticamente impossível em um dado contexto.

SUMÁRIO

1. Introdução.....	1
2. A questão dos numerais indígenas.....	8
3. Metodologia.....	19
4. Panorama Teórico.....	26
4.1. Pressupostos teóricos e desenvolvimento da teoria da gramática gerativa.....	27
4.2. A sintaxe interna dos numerais e a estratégia de empacotamento...	35
4.3. A semântica composicional de numerais.....	52
5. Apresentação dos dados e análise.....	73
5.1. O karitiana	73
5.1.1. Os numerais karitiana.....	73
5.1.2. Análise da formação dos numerais.....	83
5.1.3. Resultados.....	92
5.2. O kamayurá.....	94
5.2.1. Os numerais kamayurá.....	94
5.2.2. Análise da formação dos numerais.....	100
5.2.3. Resultados.....	105
6. Considerações finais.....	107
7. Referências bibliográficas.....	112

1. Introdução

Neste capítulo são apresentados brevemente o objeto de estudo desta pesquisa, os objetivos e a estrutura desta dissertação. O **tema geral** de investigação desta pesquisa é **a morfossintaxe e a semântica composicional dos numerais em duas línguas tupi pertencentes a subfamílias distintas**. O **objetivo central** é realizar uma análise da interface sintática e semântica da **estrutura composicional dos numerais nessas línguas** a fim de avaliar o impacto dessa abordagem para **a compreensão sobre sistemas numéricos em geral**.

Um sistema numérico é a parte de uma língua natural cuja função é expressar números inteiros e positivos, isto é, números naturais. Portanto, numerais são a expressão linguística de números, e um número é uma abstração matemática, ou seja, uma cardinalidade. É importante ressaltar que esse é o conceito de número que interessa a esta pesquisa, a abstração matemática, e não deve ser confundido com a notação científica proporcionada pelos algarismos arábicos, que tal qual o sistema numérico de uma língua natural, é apenas uma expressão dessa abstração. Numerais diferem de várias maneiras da notação por algarismos arábicos ainda que ambos sistemas sejam meios de expressar números que se formam de maneira recursiva. Algarismos, claramente, proporcionam um conhecimento universal sobre cardinalidades, o que confere uma vantagem na compreensão da semântica dos numerais em relação a outros conjuntos lexicais cuja representação do significado semântico se torna mais problemática, tal como adjetivos escalares como *alto*, cujo valor depende do contexto. É possível observar nas frases *João é um anão alto* e *João é um jogador de basquete alto* que o valor semântico de *alto* é bastante diferente em contextos diferentes. O conjunto lexical dos numerais e os que derivam dele, como horas e ordinais, possuem uma sequência convencionada e fixa para os seus itens, fato não observado nos demais conjuntos lexicais de uma língua, com exceção dos dias e meses que também tem uma ordem convencionada. Logicamente, a sequência intrínseca aos numerais se deve ao domínio semântico dessa categoria de palavras, a sequência dos números naturais.

Assim, **numeral e número são conceitos distintos**, o número é um conceito da matemática enquanto numeral é uma categoria de palavras, que é o tópico de interesse desta pesquisa. É comum encontrar na literatura o termo numerosidade para referir-se à quantidade que um numeral expressa, i.e., sua cardinalidade. Nesta análise optou-se por utilizar o termo cardinalidade, apesar de como autores citados os utilizam, mas vale

ressaltar que os termos são sinônimos. Assim, cardinalidade, ou numerosidade, se refere a quantia de entidades designada por um conjunto qualquer, ou seja, o tamanho do conjunto. Por exemplo, a cardinalidade de $|dois\ meninos|$ é 2^1 .

Nesse sentido, esta análise também irá considerar a **distinção entre sistema numérico e sistema numeral**. O termo sistema numérico se refere a um sistema numérico recursivo no sentido matemático, como por exemplo o sistema decimal e o sistema binário. O termo sistema numeral se refere ao sistema de numerais de uma língua natural, ou seja, à expressão linguística de um sistema numérico subjacente. Essa distinção é importante pois sistemas numerais de línguas naturais apresentam irregularidades, como em (1) francês *quatre-vingt*, ‘oitenta’, em contraste à *soixante*, ‘sessenta’, em (2), e *dix-huit*, ‘dezoito’, em (3).

1)	quatre-vingt	quatre-vingt
		quatro.vinte
	oitenta	4x20
2)	soixante	six-a-nte
	sessenta	seis.VT.dez
		6x10
3)	dix-huit	dix-huit
	dezoito	dez.oito
		10+8

É preciso observar que o sistema numeral difere do sistema numérico. Nos sistemas numerais das línguas naturais podem existir alterações na base numérica, o que em um sistema numérico no sentido matemático é algo fixo. Os exemplos em (1), (2) e (3)

¹ As barras $||$ são a notação matemática para cardinalidade. Se A é o conjunto denotado pelo sintagma *quatro meninos*, a cardinalidade de A pode ser notada como $|A|=4$.

mostram que o sistema numeral pode apresentar irregularidades. Esse tipo de fato linguístico leva a duas considerações importantes sobre a formação dos sistemas numerais e a investigação da expressão linguística dos sistemas numéricos. A primeira é assumir que essas irregularidades são idiossincrasias das línguas naturais e devem ser avaliadas dentro do sistema da língua em questão a fim de encontrar as motivações linguísticas que as justifiquem, essas podem ser diacrônicas ou sistêmicas. A segunda, diz respeito aos significados de base numérica, conceito que pode ser entendido como a base numérica de um sistema matemático, por exemplo 10 em um sistema decimal, mas também pode referir-se a base morfológica numeral que forma um paradigma de numerais compostos dentro de um sistema, como {-ty} em inglês, que forma dezenas como em *fifty* e *sixty*, e o morfema {-cento} que forma centenas em português como em *duzentos* e *trezentos*. Por isso, para o primeiro caso, mantém-se a nomenclatura base numérica, mas para o segundo é preciso assumir o conceito de base morfológica numeral, isto é, um morfema que representa uma cardinalidade e é utilizado de maneira recursiva na formação de numerais. Assim, o português, por exemplo, apresenta um sistema numeral cuja base numérica é decimal e as bases morfológicas numerais são {dez-}, {-cento}, {milhão} e outras. Hurford (1987) chama esse tipo de morfemas de base morfológica multiplicativa e Rothstein (2013) de multiplicador lexical (ver seções 4.2 e 4.3 respectivamente).

As línguas naturais comumente apresentam um conjunto de numerais básicos, isto é, um conjunto limitado de palavras cardinais a partir das quais os outros numerais são construídos, como no karitiana *myhint* ‘um’, *sypomp* ‘dois’, *m̃j̃jymp* ‘três’, *otadnamynt* ‘quatro’ e *yjpyt* ‘cinco’ e no português *um*, *dois*, *três*, *quatro*, *cinco*, *seis*, *sete*, *oito*, *nove*, *dez* ²³. Para além desses numerais, muitas línguas têm um sistema numérico expandido e também apresentam numerais compostos, como *sypomp yjpy ota ot* ‘sete’ em karitiana e *vinte e sete* em português⁴. Numerais compostos são formados a partir dos numerais básicos e de bases morfológicas numerais regulares via mecanismos linguísticos variados, como regras morfossintáticas, coordenação, sufixação, justaposição ou reduplicação. Esses mecanismos linguísticos expressam operações matemáticas, por

² Há registros de línguas que aparentam não conter numerais como o Pirahã (ver EVERETT, 2005).

³ O zero não foi considerado um numeral básico aqui pois há uma discussão em aberto sobre seu status como numeral na qual este trabalho não pretende se aprofundar (sobre o zero ver WELLMAN & MILLER, 1986; HANLON, 1988).

⁴ Alguns autores se referem a essa categoria como numerais complexos. Contudo, nesta análise optou-se por usar o termo numerais compostos para não haver um ruído com o conceito de número complexo da matemática, seguindo a sugestão recebida de um parecerista anônimo.

exemplo, o número *duzentos e sete* do português contém uma multiplicação via sufixação, em *duzentos* ‘dois x cento’, e uma adição expressada pela conjunção ‘e’, em *e sete*.

Ionin & Matushansky (2006) observaram que as operações semânticas básicas para a interpretação de numerais compostos são a adição e a multiplicação, embora existam línguas que também utilizam outras operações. Abaixo observamos exemplos do português *quarenta e sete* em (4) e do inglês *two hundred* em (5):

4)	quarenta e sete	quatro-VT-nta e sete
		$(4 \times 10) + 7$
5)	two hundred	two hundred
	dois cem	2×100

Esta análise propõe que para o melhor entendimento da categoria dos numerais é preciso considerar que existem línguas que utilizam outras operações, como o caso do yourubá que utiliza a subtração. Nessa língua existem palavras para as dezenas e no caso das unidades, de 1-4 são somadas à dezena e de 5-9 são subtraídas da dezena seguinte (ESIRI, 2011), como mostrado em (6):

6)	ẹẹrìndílógún	ẹrin-díl-ógún
	dezesseis	quatro.de.vinte ⁵
		$20-4$

(ESIRI, 2011, p.81)

⁵ Do original *four from twenty* (ERISI, 2011)

Ainda que essas operações variem nas línguas naturais, é consenso que adição e multiplicação são as operações mais comuns. Também existe variação na ordem dos constituintes em uma expressão numérica. Por exemplo, nas línguas românicas, como o português, os constituintes se organizam do maior valor cardinal para o menor, como em *duzentos e cinquenta e três*, em contrapartida, o alemão organiza seus constituintes de forma oposta, como em *vierundsechzig*, literalmente ‘quatro e sessenta’. As variações nas operações semânticas matemáticas que são codificadas na língua e a ordem dos constituintes são fenômenos linguísticos interessantes que precisam ser melhor descritos nas línguas naturais e, especialmente, nas línguas indígenas (ver capítulo 2).

Mesmo com as irregularidades e idiosincrasias encontradas em sistemas numéricos de línguas naturais, como observado no francês nos exemplos (1), (2) e (3) acima, uma análise mais detalhada permite observar vestígios de um sistema regular que evidencia em sua formação mecanismos linguísticos regulares. Ainda que no francês existam numerais como *quatre-vingt*, oitenta, seu sistema numeral evidencia um sistema numérico de base 10, com observamos em numerais como *deux cents*, ‘duzentos’, e *soixante*, ‘sessenta’. Considerar constatações idiosincráticas para analisar as irregularidades é mais adequado da perspectiva de uma linguística que universaliza a capacidade humana da linguagem, como o faz a linguística gerativa adotada por esta análise (ver seção 4.1), do que postular um esquema subjacente não universal para cada língua e desconsiderar os padrões regulares observados nas línguas naturais (Chomsky, 2006).

Nesta dissertação foram observados os numerais compostos das línguas Tupi karitiana (Arikém) e kaïamurá (Tupi-Guarani) ⁶. Para isso, esta análise apresenta a descrição e a análise da formação dos sistemas numerais dessas línguas a partir da abordagem da linguística gerativa e da interface entre sintaxe e semântica. No desenvolvimento da argumentação são apresentados tópicos relevantes para tal abordagem como pressupostos teóricos da linguística gerativa, a sintaxe na formação de numerais e a sua semântica composicional. A partir de uma análise prévia sobre o karitiana desenvolvida em uma iniciação científica pela autora desta dissertação, foram levantadas as hipóteses que motivaram esta pesquisa: i) sistemas numerais constituem um tópico passível de ser investigado pelo panorama da linguística gerativa já que existem padrões de formação desses sistemas amplamente observados nas línguas naturais; ii) a estratégia de

⁶ As subfamílias correspondentes são apresentadas entre parênteses.

empacotamento (HURFORD, 1987) parece conduzir a formação recursiva de numerais compostos em línguas Tupi, o que dá apoio empírico à tese de que esse mecanismo é um universal que conduz a formação dos sistemas numerais nas línguas naturais; iii) o conceito de base morfológica multiplicativa (HURFORD, 1987, 2006, 2010) é compatível com o conceito de léxico multiplicador (ROTHSTEIN, 2013, 2017) o que permite uma abordagem baseada na interface entre a sintaxe e a semântica.⁷

Diante das considerações iniciais apresentadas, o objetivo geral desta pesquisa foi investigar as construções sintáticas que expressam cardinalidades em duas línguas da família Tupi partindo da consideração de que sistemas numéricos são um subproduto da faculdade da linguagem (CHOMSKY, 1998). Assim, esta análise buscou: i) realizar uma análise da interface sintática e semântica da estrutura dos numerais nas línguas karitiana e kamayurá; ii) mostrar que a interface sintaxe-semântica é uma abordagem produtiva na elucidação do problema da análise linguística de expressões numéricas e iii) avaliar o impacto dessa análise para a compreensão dos sistemas numéricos em geral.

Rothstein (2013) defende que a principal questão linguística sobre numerais é “como se pode modelar o significado dos numerais para poder explicar seu comportamento gramatical e linguístico?”. A análise proposta por esta pesquisa vai seguir as questões de investigação propostas pela autora e pretende responder i) quais são os numerais básicos e compostos?; ii) como numerais compostos são construídos e quais são os processos aritméticos envolvidos na sua composição? iii) os numerais refletem esses processos aritméticos em sua estrutura?, iv) essas expressões são construídas no léxico ou na sintaxe?

Esta dissertação está organizada como será apresentado a seguir. No capítulo 2 é introduzida a questão da análise de numerais em línguas indígenas e informações básicas sobre a família Tupi, visando destacar a importância dessa investigação em face da escassez de trabalhos sobre o tema numerais nessas línguas, apresenta também o que há de literatura sobre numerais de línguas nativas brasileiras e também alguns problemas verificados nessas análises e em tipologias sobre o tema. O capítulo 3 apresenta a metodologia adotada por este trabalho, a origem dos dados e a justificativa da escolha das línguas. O capítulo 4 apresenta o panorama teórico adotado, esse capítulo é composto por 3

⁷ Iniciação científica desenvolvida com o auxílio financeiro do CNPq durante os anos de 2015 e 2016 sob a orientação da professora Dr^a. Ana Lúcia de Paula Müller, que também orientou esta dissertação. Essa iniciação científica foi contemplada com uma menção honrosa no 24º Simpósio Internacional de Iniciação da USP.

seções. Na seção 4.1 é apresentado o panorama teórico geral da linguística gerativa, incluindo uma breve introdução de seu desenvolvimento, seus pressupostos teóricos e a justificativa da sua escolha. Na seção 4.2 é apresentada a teoria da Estratégia de Empacotamento (HURFORD, 1987; 2006; 2010), sobre a qual se baseia a análise sintática dos numerais nesta dissertação, a investigação da proposta de um mecanismo universal na formação dessa classe de palavras e por fim reflexões acerca de uma roupagem minimalista, no sentido de Chomsky (1995), sobre essa teoria. Já a seção 4.3 apresenta e justifica, em face a outras propostas teóricas, a semântica composicional para numerais adotada aqui (ROTHSTEIN, 2013, 2017). Então, o capítulo 5 traz a apresentação dos dados e a análise, este capítulo está organizado em 2 seções, cada uma relacionada a uma língua na seguinte ordem karitiana e kamayurá. Cada seção é composta por uma subseção de apresentação da língua e dos dados, uma subseção de análise e por fim uma subseção de resultados. Finalmente, o capítulo 6 apresenta os resultados e a conclusão da investigação desenvolvida, buscando contemplar os impactos dessa análise para o tema numerais na linguística teórica e descritiva.

2. A questão dos sistemas numerais indígenas

Com o objetivo de ilustrar o panorama geral sobre sistemas numerais de línguas indígenas, neste capítulo são apresentados tópicos relevantes da investigação linguística sobre sistemas numerais, ressaltando sua importância, a escassez de abordagens formais sobre o tema e alguns problemas na descrição e análise dessa classe de palavras.

A investigação linguística sobre sistemas numerais de línguas não indo-europeias de uma perspectiva formal é uma importante contribuição para o desenvolvimento de teorias que buscam universais linguísticos, já que possibilita uma análise minuciosa de dados de línguas tipologicamente distantes e subdescritas contribuindo para um conjunto de dados disponíveis na literatura mais completo e diversificado. Isso vale, especialmente, sobre análises de línguas indígenas brasileiras, que apesar de esforços recentes, ainda constituem um conjunto de línguas subdescritas na literatura da linguística formal.

Os sistemas numerais das línguas naturais não foram alvos de grandes debates intelectuais e linguísticos no século XX, como apontado por Hurford (1987). Para esse autor, a morfossintaxe e a semântica intrincada dos numerais são em si suficientes para justificar o interesse nesses sistemas. Ele levanta duas principais razões para o não interesse da linguística moderna sobre sistemas numerais durante o período dos 60 ao 80, reproduzidos a seguir⁸:

- a) Numerais são uma área especial do estudo do significado, essas expressões têm um valor cardinal e semântico convencionado nas línguas naturais, e, além disso, a estrutura de sistemas numerais não representa suficientemente a estrutura da língua de forma geral. Assim, é um tópico marginal no que concerne a temas centrais em linguística.
- b) A língua é usada apenas para prover nomes para um pequeno conjunto de números naturais que uma comunidade deseje nomear. Para outros tipos de números como números negativos, fracionais, irracionais entre outros é preciso ir além da língua e

⁸ Traduções livres da autora desta pesquisa.

criar notações técnicas. Assim, sistemas numerais não revelam a natureza dos números de forma geral.

Ambos argumentos colocam os sistemas numerais como um objeto científico periférico. Hurford ainda destaca que por vezes linguistas dizem que numerais podem interessar a matemáticos, mas são apenas um interesse marginal à Linguística, enquanto os matemáticos dizem que numerais podem interessar a linguistas, mas que são apenas um interesse marginal à Matemática. Sobre isso, o autor desenvolve uma analogia interessante e afirma que, se fosse esse o caso, seria como argumentar que o ornitorrinco não é um animal interessante porque não é prototipicamente mamífero, nem pássaro e nem réptil, mas aparentemente algo entre essas categorias.

Ora, entidades como os vírus são importantes no estudo da biologia justamente por estarem no limiar dos seres vivos. Contrariamente aos dois argumentos apresentados acima, objetos alocados na fronteira entre domínios intelectuais diferentes são de grande importância, já que configuram pontes em potencial entre os domínios envolvidos e ajudam a olhar fatos de um domínio a partir da perspectiva de outro domínio.

De fato, os argumentos apresentados em (a) e (b) não são compatíveis um com o outro como aponta Hurford. O primeiro diz, grosso modo, que numerais são objetos linguísticos atípicos, já que as entidades que denotam, i.e., números, são bastante diferentes das entidades envolvidas no resto da língua, como pessoas, objetos, eventos, etc. Evidentemente, numerais envolvem operações aritméticas em sua interpretação, o que não é encontrado em outros subsistemas da linguagem humana. Isso significa dizer que em certa medida existe uma relação entre a forma desse subsistema linguístico e o tipo de entidade que ele denota, relação que o segundo argumento nega ao afirmar que numerais não tem relação com a natureza dos números.

Ainda que não exista de fato uma relação direta entre a natureza ontológica dos números e sua expressão linguística, há um pareamento claro entre numerais e números naturais, já que aqueles são a expressão linguística destes. Os sistemas numerais são de forma bastante óbvia bem integrados às línguas que pertencem, estão sujeitos a operações sintáticas comuns a outras classes de palavra, como concordância, marcação de caso, topicalização e tantas outras.

Estas expressões linguísticas, portanto, estão sujeitas a restrições como estão todos os itens que compõe um sistema linguístico, i.e., uma língua. Assim, são apresentadas a seguir algumas propriedades dos numerais que evidenciam sua integração com as restrições da língua. Por exemplo, normalmente, palavras não numéricas não ocorrem na estrutura interna dos numerais como em (7):

7) *Dois fortes mil homens lutaram essa batalha

Nas línguas naturais, numerais costumam ter um lugar bem definido em construções sintáticas maiores, por exemplo, em inglês eles ocorrem entre um determinante e um adjetivo atributivo como mostrado pelas sentenças em (8):

- 8) a. The two English boys came to visit us.
b. *The English two boys came to visit us.

Como apontado por Cobbert (1978), em muitas línguas numerais tem características distribucionais e morfológicas compartilhadas com a classe dos adjetivos e dos nominais. Para Hurford (1987), numerais são primeiramente adjetivos e também ocorrem como nominais. A diferença de ordem entre numerais e outros adjetivos e nomes seria derivada de suas características semânticas.

Esta análise, contudo, defende que é preciso considerar uma gama mais diversificada de línguas para confirmar tais afirmações. Muitas línguas da família Tupi são conhecidas por não apresentarem uma classe de adjetivos, ou ainda apresentarem uma quantidade muito restrita de adjetivos. Por isso, é preciso verificar se nessas línguas os numerais são de fato adjetivos, como sugere boa parte da literatura sobre numerais, ou se pertencem a outra classe de palavras, apresentando, assim, um comportamento linguístico diferente do que é observado nas línguas mais bem investigadas pela linguística.

Muitas vezes, línguas nativas apresentam fatos linguísticos bastante diferentes de línguas bem descritas como o português e o inglês. O karitiana, por exemplo, é uma língua em que o sistema verbal é muito mais rico que o sistema nominal, já que nessa língua não há material funcional, como determinantes e quantificadores, no sintagma nominal (MULLER et al, 2006). Considerando essas diferenças, se faz preciso analisar línguas nativas para se obter teorias e análises mais completas sobre sistemas numerais de línguas

naturais. Nesse sentido, esta pesquisa optou por investigar os sistemas numerais de duas línguas da família Tupi, o kartiana e o kamayurá.

A família Tupi é um dos maiores grupos linguísticos da América do Sul e contém a subfamília Tupi-Guarani, uma das subfamílias indígenas mais amplamente descritas na literatura sobre línguas indígenas sul-americanas (RODRIGUES, 1986). No entanto, ainda faltam descrições linguísticas mais completas dos sistemas numerais das línguas dessa subfamília e, ainda mais, das demais subfamílias da família Tupi. O tema é normalmente apresentado como uma pequena parte da descrição ou da gramática de uma língua ou então como parte de tipologias. E apesar do esforço de alguns autores em esclarecer a morfologia e a origem dos sistemas numerais dessas línguas, não há um aprofundamento da discussão sobre a morfossintaxe e a semântica dessas expressões linguísticas.

Estudar numerais de línguas indígenas do ponto de vista formal pode avançar a compreensão do desenvolvimento dos sistemas numéricos e contribuir para a teoria linguística com dados dessas línguas. Além disso, a necessidade de descrições de línguas indígenas brasileiras é reforçada pelo fato de que das cerca de 160 línguas indígenas atualmente faladas no Brasil, existe algum tipo de descrição de apenas 83 delas (LEITE e FRANCHETTO, 2001). A pesquisa linguística sobre línguas indígenas que vai além de aspectos tipológicos e funcionais e adota uma perspectiva teórica e que aborda níveis mais ‘altos’ de análise (como a sintaxe e semântica) ou, ainda, que adote uma perspectiva teórica gerativa, só muito recentemente vêm sendo desenvolvidas (GABAS e MOORE, 2006). Ainda mais raras são as abordagens via interface entre os níveis de análise, o que reforça a necessidade desse tipo de abordagem. Entretanto, é preciso destacar os esforços recentes para realizar esse tipo de análise com as línguas indígenas.

Muitas línguas indígenas sul-americanas têm substituído ou suplementado os seus sistemas de contagem nativos por empréstimos do português ou espanhol, o que torna esses sistemas um de seus traços mais ameaçados. Essa ameaça ocorre não somente no caso das línguas indígenas sul-americanas, como também de línguas nativas do mundo inteiro, constituindo-se, portanto, prioridade na documentação dessas línguas (EPPS & SALANOVA, 2012). Comrie (2005) também destaca que os sistemas numéricos de línguas minoritárias são elementos mais ameaçados que as próprias línguas, fato que se justifica mais por razões sócio-políticas do que linguísticas.

Quando uma língua é falada em contato com outra língua que goza de mais prestígio ou que abrange um maior número de falantes algumas partes dessa língua são mais ameaçadas que outras. Numerais podem ser emprestados entre línguas vizinhas. Esses empréstimos podem ocorrer naturalmente em alguns contextos, como, por exemplo, no contato via comércio onde preços e valores são debatidos e, portanto, numerais são bastante utilizados (MAZAUDON, 2009).

Existem alguns estudos sobre os numerais em línguas indígenas. Green (1997) classifica os numerais das línguas indígenas brasileiras sob parâmetros tipológicos; Medeiros e Miranda (2009) que apresenta um estudo preliminar dos quantificadores numéricos do krenak; Fargetti (2007) sobre os numerais do yudjá (Juruna); Fargetti & Sumaio (2015) que apresentam uma reanálise do numerais em yudjá e Lima (2012) sobre a estratégia de empacotamento nos numerais do yudjá.

Ainda assim, Fargetti & Sumaio (2015) destacam que a análise linguística de numerais ainda é uma lacuna nos estudos de línguas indígenas, ainda mais levando em consideração a base numérica matemática dos sistemas e a correlação com sua expressão linguística. Nesse tipo de análise é preciso destacar que numerais têm dois níveis, como “número” pela via da lógica numérica matemática subjacente, e como “numeral” pela via da expressão linguística (ver seção 4).

Muitas vezes, há uma confusão na literatura entre os elementos matemáticos de um sistema numeral e seus elementos linguísticos, o que gera equívocos na interpretação e descrição desses sistemas. Comrie (2005) propõe uma classificação orientada pela base numérica, i.e., o valor n sobre o qual as expressões numéricas se constroem seguindo um padrão. Assim, a lógica da recursividade numérica da perspectiva da matemática é expressa pela fórmula $x.n + y$, i.e., algum número x qualquer multiplicado pela base numérica n e somado a um número y qualquer, como o exemplo em (9) do sistema decimal do português cuja base n é 10.

9) *quarenta e dois*: $42 = (4 \times 10) + 2$

Comrie lista os seguintes tipos de sistema numéricos: i) decimal, como o português; ii) vigesimal-decimal híbrido, como o francês; iii) puramente vigesimal, como o diola-fogny (da família Nigero-congolesa, falada na Gambia e no Senegal) ; iv) outras bases, como o ekari (da família Trans–New Guinea, falada na indonésia) de base 60; v) de extensão de

partes do corpo, como o kobon (da família Trans–New Guinea, falada na Papua Nova Guiné) ; vi) restrito, que forma expressões para valores até no máximo cinco, como o pirahã (da família Mura, falada na região Amazônica do Brasil).⁹

Entretanto as nomenclaturas escolhidas por esse autor não são claras em relação à classificação proposta. Tais nomenclaturas hibridizam o conceito de base matemática com a sua expressão linguística, ou seja, com sua morfossintaxe. Por exemplo, a consideração sobre “expressões de partes do corpo” se refere a morfologia da expressão numérica, e não ao conceito de base numérica que o autor utiliza. E também, da perspectiva da Matemática, não é possível haver uma base híbrida, pois um sistema numérico matemático se constrói sobre uma base fixa, mesmo que sua expressão linguística seja opaca em relação a isso apresentando irregularidades, como o sistema numeral do francês apresentado no capítulo anterior, por exemplo.

Diante da questão da hibridização do conceito de base matemática e expressão linguística, esta análise argumenta que um bom caminho analítico e descritivo reside no fato de que sistemas numerais são um objeto científico no limiar entre a linguística e a matemática, como apontado por Hurford (1987). Por isso, esta pesquisa defende que analisar sistemas numerais a partir de suas características linguísticas e matemáticas propicia uma melhor compreensão de sua formação composicional. Ou seja, compreender a lógica matemática e a expressão linguística como elementos relacionados, mas independentes, ajuda no esclarecimento da sua composição, já que a análise matemática e a análise linguística fornecem informações distintas, mas complementares sobre o sistema numeral.

A base de um sistema numérico e/ou numeral pode ser tratada como um ‘bloco de montar’ fundamental para a construção do sistema. Tanto em sistemas numéricos quanto em sistemas numerais, a base é definida pelo maior número ou numeral que é utilizado para construir números/numerais maiores. Assim, por exemplo, em um sistema decimal, em que a base é 10, deve haver numerais básicos até 10 e numerais compostos acima desse valor. No caso de sistemas numerais, deve apresentar nomes simples para suas bases morfêmicas multiplicativas como, por exemplo, *hundred* ou *thousand*.

⁹ Exemplos do autor (COMRIE, 2005).

Em um trabalho tipológico, Green (1997) descreve 47 sistemas numéricos de línguas indígenas. A autora defende classificar esses sistemas numerais a partir de suas bases em: i) de base 1: kampa, kulina, tenharim, nadëb, sanuma; ii) de base 2: xavante, arara, bororo, kayapó, etc; iii) de base 3 como o atroari; iv) de base 5 como o mundukuru; v) de base 10 como pa'ikwené; vi) e sistemas em que aparentemente domina a base 20: karajá, kadiwéu, makuxi, tikuna, paresi, juruna, karitiana entre outras.¹⁰ Sua descrição, contudo, não apresenta análises exaustivas de cada língua que justifiquem a classificação proposta e não define com precisão o conceito de 'sistema numérico', nos termos da autora, considerado. Green não deixa claro se 'sistema numérico' se refere a um sistema no sentido matemático, como o sistema binário ou o sistema decimal por exemplo, ou a um sistema numeral, i.e., a expressão linguística de um sistema numérico.

Essa autora diz que as línguas cujo sistema numérico é de base vigesimal têm palavras distintas para os numerais de 1 a 5; o termo para 5 significa “nossa mão”, “todos os dedos da mão” ou “o fim da mão”; e em diante são usados a outra mão e os pés para expressar os numerais maiores. Para ela alguns exemplos de línguas que utilizam esse sistema são: Karajá (Macro-Jê; Alford, 1987; Karajá, Fortune e Fortune, 1989), Rikbaktsá (Macro-Jê; Boswood, 1978), Urubú-Kaapor (Tupi-Guaraní; Kakumasu e Kakumasu, 1988), Kadiwéu (Guaikurú; Griffiths, 1975), Karitiana (Arikém; Landin, 1983), Tikúna (isolada; Anderson, 1958), Makuxí (Karib; Williams, 1932), entre outras.

O caráter tipológico e quantitativo desse estudo de Green, entretanto, não contemplou a análise sintática e semântica dos termos numerais de forma mais profunda, e inclui algumas notas de rodapé sobre exceções à descrição geral que ela faz desses sistemas. diz:

“Os sistemas vigesimais dessas línguas são quinários, pois os numerais maiores do que cinco compõem-se dos mesmos cinco termos, e o numeral 10 é considerado uma combinação de ‘cinco mais cinco’ (mão mais mão).”

A própria autora observa o caráter quinário desses sistemas numéricos, de modo que uma análise mais cuidadosa da formação desses numerais revela que talvez não seja o

¹⁰ Exemplos da autora (GREEN, 1997).

caso de um sistema numérico vigesimal, mas sim de um sistema numérico quinário, como se pretende argumentar no capítulo 5.

Na tipologia de Comrie (2013) os sistemas numéricos de base 5 estão no grupo de sistemas numéricos de outras bases, entretanto não julgamos essa classificação adequada pois sistemas quinários são bastante comuns nas línguas naturais, especialmente nas línguas indígenas brasileiras. Entretanto, como já foi dito, essa classificação apresenta problemas quanto a distinção do que é base matemática e do que é base morfológica. Essa análise reforça a necessidade de elaborar melhor tipologias linguísticas de numerais, de modo a considerar tanto as questões linguísticas quanto as questões matemáticas contidas na formação dessas palavras e assim descrever melhor esses fenômenos linguísticos.

Não distinguir entre número e numeral pode provocar equívocos em análises de sistemas numéricos. Provavelmente, Green (1997) avaliou o sistema numérico karitiana como de base vigesimal porque não analisou a morfologia dessas palavras e tampouco considerou o processo recursivo de composição das expressões numéricas em relação a sua base numérica, o que a levou a considerar o corpus numérico da língua como um conjunto lexical de números básicos. Contudo, uma observação acurada mostra que a recursividade desse sistema é sobre cinco.

Diferente da descrição de Green (1997), a reanálise do karitiana feita por Vignado (2016) mostrou que esse é um sistema de base 5, e não de base 20. Pois há uma regularidade no uso de bases morfológicas que se referem a partes do corpo, por exemplo *yjpy* ‘nossa mão’ e *yjpi* ‘nosso pé’ que são utilizadas em uma lógica de 5 em 5, o que é evidência de que o valor da base é 5 (para mais detalhes ver seção 5.1).

Uma reanálise do sistema numeral yudjá (Juruna), feita por Fargetti e Sumaio (2015), argumenta que originalmente esse é um sistema de base 2, e não de base vinte como Green (1997) e Fargetti (2007) haviam proposto. À primeira vista, foram considerados 20 numerais básicos que se expressam via as palavras que se referem aos dedos. Contudo, a partir de considerações diacrônicas do sistema e da observação da morfologia das expressões numéricas, as autoras levantaram duas hipóteses bastante distintas: 1) houve um processo de redução do sistema numérico seguido por um resgate do conhecimento que havia sido esquecido; 2) trata-se apenas de um processo de expansão do sistema numérico. A partir dessas hipóteses, as autoras concluíram que no caso do Yudjá, trata-se de um

processo de expansão do sistema numeral, e não de redução. Dessa forma, os numerais compostos são inovações e não configuram um conhecimento anterior que tenha sido resgatado.

De acordo com Winter (1999) a tendência observada nas línguas do mundo é seguir na direção da expansão dos sistemas numéricos de acordo com a necessidade dos falantes e do contato com outros sistemas e culturas, e nunca no sentido de sua redução. Em um estudo inicial sobre os numerais yudja (Juruna), Fargetti (2007) propõe que o sistema numérico juruna era composto por 20 numerais básicos e era vigesimal. Contudo, em um Segundo momento, Fargetti e Sumayo (2015) fazem uma reanálise dos numerais Juruna diante do conjunto de dados linguísticos e de informações extralinguísticas, como, por exemplo, a resistência dos mais velhos a números maiores e também por eles contarem com os dedos em duplas. Sua reanálise desse sistema propõe que originalmente o sistema numérico dessa língua era de base dois, no sentido matemático, e continha apenas 4 numerais básicos em seu sistema de contagem. Elas assumem que o sistema numérico yudjá era de base dois em um estágio anterior na sua diacronia e continha quatro numerais básicos e, posteriormente, o sistema se expandiu e passou a expressar os numerais usando as palavras que expressam os dedos das mãos e dos pés e tornou-se um sistema numeral quinário, i.e., de base 5.

Reanálises de sistemas numerais como as de Fargetti e Sumayo (2015) e Vignado (2016) mostram que é comum na literatura linguística o ruído na distinção dos elementos matemáticos de um sistema numeral e dos seus elementos linguísticos, o que gera equívocos na interpretação e descrição desses sistemas. Os fatos apresentados sobre o yudjá e o karitiana mostram a necessidade de considerar as fronteiras entre os níveis de análise linguístico e matemático de um sistema numeral para obtenção de análises mais precisas e adequadas.

Mais recentemente os numerais das línguas indígenas brasileiras têm figurado pesquisas em semântica formal que buscam investigar quantificadores e a distinção massivo-contável nessas línguas. Sanchez-mendes (2009) investigou os quantificadores da língua karitiana, essa autora mostrou que no karitiana há apenas quantificação adverbial, mas não apresenta quantificadores nominais. Ainda que este trabalho não trate especificamente de numerais, mostra que não há determinantes no karitiana. Desse modo, os numerais do karitiana não são determinantes. Esta análise considera que os numerais

nessa língua são elementos adverbiais, como vai-se argumentar em mais detalhes na subseção 5.1.1 que apresenta a língua karitiana e seu sistema numeral.

Recentemente, Lima (no prelo) tem investigado a numerosidade nas línguas indígenas brasileiras sob a perspectiva da semântica formal e da aquisição de numerais. Numerosidade é o conceito que autora utiliza para se referir a expressões linguísticas que envolvem cardinalidade, seja ela exata ou aproximada. Nesse trabalho, a autora parte de quatro línguas não aparentadas kawaiwete, munduruku, pirahã, yudja e mostra que os numerais variam em estrutura e extensão através das línguas. Por exemplo, no pirahã não há palavras que se referem a quantidades numéricas exatas e em munduruku o sistema de numerais é restrito a poucos números, já no kawaiwete e no yudja os numerais são reconhecidos por falantes adultos e vão até vinte.

Determinar o tamanho de um sistema numeral não é simples já que o maior numeral observado em interações entre falantes da língua ou narrativas não necessariamente representa os limites dos numerais de uma língua (EPPS et al., 2012). A não trivialidade disso se evidencia quando se pergunta, por exemplo, qual o tamanho do sistema numeral do português. Para responder essa pergunta, pode-se verificar qual é a última base desse sistema numeral, no caso do português parece ser *trilhão* a maior base utilizada no uso corrente. Existe também a expressão *zilhão*, mas essa expressão não é associada a um valor cardinal específico, como são as bases multiplicadoras de um sistema, mas se refere a quantias enormes. Assim, refere-se a uma numerosidade que se define por ser enorme, mas não a uma cardinalidade, que deve ser exata. Esse exemplo ilustra a dificuldade em se definir o tamanho de um sistema numeral.

Alguns problemas na análise de numerais foram levantados neste capítulo, e reforçam a necessidade análises formais mais detalhadas para essa categoria de palavras. Corver et al (2007) observam que duas importantes questões na investigação de numerais de uma perspectiva formal são: (i) quais são as regras estruturais para construir expressões numéricas complexas? e (ii) como a variação que vemos entre as línguas naturais é restrita? Esta pesquisa pretende responder a essas questões propondo uma análise baseada na interface entre uma teoria sintática na formação de numerais compostos e uma teoria da semântica composicional dessas palavras.

Este capítulo apresentou considerações importantes sobre sistemas numerais enquanto objeto científico da linguística e sua interface óbvia com noções da matemática a fim de aprofundar as análises linguísticas. Apresentou brevemente a família Tupi, considerando o estado de arte da linguística sobre os sistemas numerais dessa família citando alguns trabalhos que contemplaram o tema. Por fim, apresentou alguns problemas em análises e tipologias de numerais atentando a necessidade de uma melhor delimitação do que concerne à matemática e o que concerne a linguística na compreensão desses sistemas.

O próximo capítulo apresenta a metodologia que foi adotada nesta pesquisa. Assim apresenta as distinções entre número e numeral e sistema numérico e sistema numeral, essa categorização é fundamental para abordagem adotada nas análises. Também justifica a escolha das línguas analisadas e apresenta a origem dos dados que constituem o ponto de partida para a análise.

3. Metodologia

Este capítulo apresenta a metodologia de coleta e análise de dados adotada no desenvolvimento desta pesquisa. Dessa forma, o capítulo conta com a origem dos dados analisados, apresentando a metodologia de coleta de dados aplicada no caso dos dados do karitiana. A seguir, apresenta uma distinção conceitual importante proposta por essa pesquisa para a descrição e análise dos dados que consiste na diferenciação de número e sistema numérico *versus* numeral e sistema numeral. Por fim, apresenta a escolha das teorias que baseiam a análise proposta por esta pesquisa.

A primeira fonte dos dados do karitiana e do kamayura foi a literatura linguística sobre essas línguas, bucou-se consultar gramáticas, artigos e listas de itens lexicais. No caso do karitiana, considerou-se o dicionário karitiana-português de Landim (1984) e o Léxico karitiana desenvolvido por Storto & Rocha (não publicado). A autora desta pesquisa desenvolveu três iniciações científicas sobre a língua karitiana e seus numerais. Ao longo dessas iniciações científicas e do desenvolvimento do mestrado, foram realizadas quatro coletas de dados.

Para a realização das coletas de dados, os informantes, todos bilingues, falantes de karitiana e português, foram trazidos até a Universidade de São Paulo onde foram consultados por pesquisadores da língua. Em 2013 foram consultados dois homens adultos, E e O. Em 2014 foram consultados dois informantes homens adultos, J e L. Em 2015 foram consultados outros dois homens adultos, I e C. E, já no mestrado, em 2018 foram consultadas duas mulheres adultas EL e M. Assim, no total, foram consultados 8 falantes adultos diferentes. Buscou-se contemplar ambos os sexos nas coletas de dados mediante a possibilidade de participação dos falantes nas coletas de dados que foram realizadas no estado de São Paulo.

Em todas as coletas foi aplicado um formulário de consentimento para participação das coletas de dados. O formulário consiste em um documento que apresenta as condições de participação nas coletas de dados e é apresentado nos anexos desta pesquisa. A realização das coletas de dados consistiu na aplicação de formulários de coleta dados baseados no método de elicitación controlada (MATTHEWSON, 2004). Há uma dificuldade na obtenção de dados espontâneos de numerais, como apontado por Epps et al.

(2012) a interação entre falantes nem sempre evidencia os maiores numerais de um sistema. Por isso, foi necessário elicitare a lista de numerais.

No desenvolvimento das iniciações científicas, foi investigada a semântica sentencial dos numerais em karitiana, e para a obtenção dos dados necessários foi preciso criar contextos para verificar as condições de verdade das sentenças. Matthewson (2004) afirma que esse método fornece pistas das condições de verdade e de felicidade das sentenças. As condições de verdade representam o significado de uma sentença mostrando em que situação ela é verdadeira. Já as condições de felicidade são as condições de uso de uma sentença. Segundo a autora há dois tipos de elicitação: i) pedir traduções e ii) pedir julgamentos. Em ambos os tipos, é desejável o uso de uma meta-língua que seja conhecida pelo pesquisador e pelo consultor. No caso dessa pesquisa, a meta-língua foi o português.

As coletas buscaram contemplar os dois tipos. Desse modo, em um primeiro momento foram pedidas traduções de sentenças, Matthewson recomenda pedir tradução de sentenças e não de palavras isoladas ou sintagmas já que pode ser difícil para o consultor isolar a sequência relevante e as vezes pode não haver uma tradução palavra a palavra de uma língua para a outra. Foram pedidas traduções de sentenças como por exemplos a tradução apresentada em (10) que foi dada pelo falante O em 2013.

10) ‘Na escola tem quarenta e cinco alunos’

O: *Quarentat e cincot alunot naka’agnagit escola pip*

O passo seguinte é atribuir uma tradução morfema a morfema com base em conhecimentos prévios em morfologia, fonologia e sintaxe sobre a língua. Em (11) é apresentada a sentença já glosada, que assim pode ser analisada e utilizada em outras coletas.

11) *Quarentat e cincot aluno naka’agnagit escola pip*
Quarenta-t e cinco-t aluno naka-’agnagi-t escola pip
quarenta- ADVZ e cinco-ADVZ aluno DECL-cop.pl-NFUT escola em

Com sentenças traduzidas por falantes do karitiana, o segundo passo foi testar os julgamentos das sentenças. Então foram apresentados contextos como o exemplo em (10) aplicado ao informante I em 2014.

12) *No ano passado tinham 40 alunos na escola, esse ano entraram mais 5 crianças. Agora a escola tem 45 alunos. Eu posso dizer 'quarentat e cincot alunot naka'agngit escola pip'*

I: Sim, 'tá' bom.

Dessa forma, as duas maneiras de elicitar dados, tradução e julgamento, se complementam e o pesquisador pode ter uma certeza maior quanto a semântica de seus dados. Entretanto, durante o desenvolvimento da investigação sobre a semântica sentencial dos numerais, a lógica da formação composicional dos numerais em karitiana se mostrou interessante e tornou-se o foco da pesquisa diante do fato da necessidade de se descrever e analisar sistemas numerais de línguas brasileiras.

Então, paralela à coleta de dados via elicitación controlada, buscou-se coletar uma lista completa de numerais na língua. A investigação partiu de listas de léxicos anteriores como Landim (1984) e Storto & Rocha (não publicado) e selecionou a lista de numerais já existente. Essa lista foi apresentada aos consultores com a finalidade de verificar a aceitabilidade desses numerais e se de fato os falantes o conheciam.

A tradução dada em (10) mostra que para números maiores os falantes preferem os numerais do português, e isso se deve a razões extralinguísticas de contato. Ainda assim, os falantes não tiveram problemas em traduzir os numerais em karitiana inseridos em sentenças na língua. Eles também reconheceram as formas numerais sem problemas. Muitas vezes, durante a elicitación, os falantes forneceram sentenças com numerais em português, e quando foi pedido que dessem o mesmo numeral em karitiana, eles reformulavam a sentença e recitavam o numeral karitiana sem problemas. As coletas de dados forneceram evidências de que os falantes dessa língua têm conhecimento do sistema numeral da língua, ainda que no uso corrente acabem por utilizar mais os numerais básicos na língua.

A observação do sistema numeral do karitiana levou a pesquisa sobre outros sistemas numerais de línguas indígenas brasileiras. Em um primeiro momento, foram consultados os sistemas numerais do tupi antigo e do nheengatu durante disciplinas de língua tupi oferecidas pelo curso de Letras da Universidade de São Paulo. Entretanto, optou-se por analisar os numerais do kamayurá pela disponibilidade dos dados do seu sistema numeral. A fonte dos dados dessa língua consistiu nos dados da gramática do

kamayurá desenvolvida por Seki (2000). Essa autora apresenta uma lista completa dos numerais do kamayurá de 1 a 20 e também apresenta o significado dos morfemas envolvidos na composição desses numerais. A análise morfológica dessas expressões apresentada por Seki facilitou a análise dos dados.

Portanto, os dados dos sistemas numéricos indígenas foram coletados via coleta de dados pautada em elicitación controlada, no caso do karitiana, e via pesquisa bibliográfica, no caso do kamayurá. Depois de uma seleção sistemática dos dados de numerais das línguas, a sua morfossintaxe foi descrita e analisada. Por fim, a análise da formação de sistemas numerais dessas línguas foi feita a partir dos pressupostos teóricos apresentados no capítulo 4. Sobre a apresentação dos dados, buscou-se glosar os morfemas envolvidos na formação dos numerais descritos a partir da descrição desses morfemas na literatura. Os dados de outros pesquisadores apresentam referências explícitas, já os dados que não apresentam referência são da autora desta pesquisa.

Como discutido anteriormente, esta pesquisa assume que **numeral** e **número** são conceitos distintos, número é um conceito da matemática enquanto numeral é um conceito da linguística. Número é um conceito abstrato da matemática, abrange várias concepções de número como números naturais, números reais, números irraciais, números fracionados, números negativos entre outros. Numeral é um conceito linguístico que se refere às palavras e expressões que expressam números naturais em uma língua como *dois* ou *vinte e sete* em português ou *two* e *twenty-seven* em inglês.

Os numerais de uma língua contemplam apenas algumas concepções de número da matemática, como por exemplo, números naturais como 2 com o numeral *dois*, números fracionados como *um terço* e números ordinais como *segundo*. Esta pesquisa contempla apenas os numerais cardinais das línguas analisadas, mas ressalta a importância de investigações linguísticas sobre os numerais de línguas naturais de maneira geral.

Concepções de números mais complexos são expressos apenas na matemática, via notação técnica através do uso de algarismos, sinais matemáticos e letras, como, por exemplo, -2, π ou e (o número de Euler). Esses exemplos mostram que ainda que exista um pareamento entre números e numerais nas línguas naturais. Esses conceitos não são sinônimos e se referem a objetos bastante diferentes e por isso a distinção entre número e numeral se faz necessária na investigação de numerais.

De forma análoga, se estabelece a distinção entre **sistema numérico** e **sistema numeral**. O termo ‘sistema numérico’ se refere a um sistema recursivo e posicional de unidades numéricas no sentido matemático e frequentemente é representado por algarismos, como por exemplo o sistema decimal e o sistema binário. Nesses sistemas há o estabelecimento de uma base numérica sobre a qual se constrói o sistema via uma disposição posicional, por exemplo, o número 243 apresenta algarismos alocados em uma lógica posicional de um sistema decimal que representa 2 centenas, 4 dezenas e 3 unidades. Já o termo ‘sistema numeral’ se refere ao conjunto de numerais de uma língua natural, esse tipo de sistema é composto por numerais simples e numerais compostos. Sistemas numerais são a expressão linguística de um sistema numérico, e ainda que tais sistemas apresentem padrões morfossintáticos bastante claros, é preciso atentar que eles têm características próprias de um sistema linguístico e por vezes não constituem uma correspondência exata de um sistema numérico.

Essa distinção é importante pois sistemas numerais de línguas naturais são passíveis a irregularidades e idiossincrasias como todos os outros subsistemas de uma língua natural. Por exemplo, no sistema numeral do português os numerais de 11 a 19 são construídos via o morfema prefixal {dez-}, exceto no caso dos números *onze*, *doze* e *treze* em que se utiliza o alomorfe sufixal {-ze}. No inglês ocorre algo similar, para essa série de números usa-se o morfema sufixal {-teen} como em *thirteen* e *fourteen*, por exemplo, exceto nos numerais *eleven* e *twelve*, o que se justifica com uma análise diacrônica, pois esses numerais derivam do proto-germânico **ain-lif* e **twa-lif* (DANTZIG, 1954). Assim, também é preciso considerar razões diacrônicas para explicar a aparente irregularidade desse paradigma em português e também nas demais línguas naturais.

Ainda que a distinção número/numeral seja fundamental para a análise desenvolvida por esta pesquisa, não se pode ignorar que a matemática é uma poderosa ferramenta na classificação de sistemas numerais, já que a observação da recursividade numérica de um sistema ajuda no esclarecimento da recursividade sintática dessas expressões, ou seja, é fundamental observar ‘de quanto em quanto’ os padrões morfossintáticos de um sistema são alterados para compreendê-lo melhor. Como apontou Hurford (1987), é na intersecção entre a matemática e a linguística que reside o ‘mistério’ dos numerais e tal característica intrínseca e fundamental não deve ser ignorada (ver capítulo 2).

Em síntese, o numeral é a representação linguística de um número, mas não se resume a isso já que está inserido em um sistema maior que é a língua. A face linguística de um numeral está sujeita a modificações diacrônicas, restrições linguísticas e outras idiossincrasias, como estão sujeitos todos os elementos de uma língua. Portanto, análise linguística dos numerais deve abranger ambos níveis de análise de um sistema numeral, i.e., o matemático e o linguístico, de maneira complementar a fim de evitar interpretações equivocadas da recursividade desses sistemas.

A constatação de padrões recorrentes na formação de sistemas numerais nas línguas naturais sugere que, de certo modo, há algo de universal nesse fenômeno no que concerne a natureza da cognição humana e da linguagem. Tais fatos foram amplamente observados por Hurford (1975, 1987, 2007) que dentro do panorama da linguística gerativa propôs um conjunto de regras sintáticas para numerais que culminou na sua teoria da Estratégia de Empacotamento.

Essa teoria tem sido endossadas por dados de sistemas numerais de línguas bastante distantes (ver HURFORD, 1987; LIMA, 2012; AMAECHI, 2014; KOŞANER, 2016; VIGNADO, 2017). Além disso, na literatura, Hurford é o autor que mais se dedicou à análise linguística de numerais a partir de uma perspectiva que universaliza a faculdade da linguagem. Por isso, este autor será bastante utilizado na investigação da sintaxe dos numerais que serão analisados por esta pesquisa. Adicionalmente, esta pesquisa busca confrontar as regras sintáticas propostas por Hurford com alguns pressupostos do Programa Minimalista com o intuito de dialogar com a versão mais recente da Gramática Gerativa e fomentar o debate sobre sistemas numerais na linguística brasileira ainda que esta análise não logre implementar detalhadamente esse tipo de formalização (ver seção 4.1 e 4.2).

As operações aritméticas envolvidas na interpretação de numerais conduziram à necessidade de considerar teorias semânticas que contemplem tais operações para melhor compreender o significado das estruturas que expressam numerais. Para isso, optou-se por aplicar a teoria semântica para numerais de Rothstein (2013, 2017) que propõe uma análise formal para a denotação dessas expressões que diferencia a semântica de numerais cardinais da semântica de multiplicadores como *hundred* (ver seção 4.3).

A partir das teorias elencadas, verificou-se se a estratégia de empacotamento opera na formação dos numerais compostos no karitiana e no kamayurá fornecendo evidências para a universalidade desse mecanismo na formação desse tipo de expressões nas línguas naturais. Partindo da perspectiva de que a estrutura sintática tem uma interface com a semântica, esta pesquisa parte da interface entre o componente sintático e o componente semântico a partir da proposta de tipos semânticos dos numerais de Rothstein. Por fim, esta pesquisa propôs uma análise sintático-semântica para os numerais dessas línguas a partir do panorama brevemente apresentado aqui (para mais informações ver seção 4).

Este capítulo apresentou a metodologia aplicada nesta pesquisa, já o capítulo seguinte apresenta o panorama teórico adotado por esta pesquisa para uma análise formal dos sistemas numerais das línguas karitiana e kamayurá.

4. O panorama teórico

Este capítulo apresenta o panorama teórico que motivou e baseou o desenvolvimento desta pesquisa e conduziu a análise dos sistemas numerais do karitiana e do kamayurá apresentada no capítulo 5. Como foi exposto nos capítulos 1 e 2 anteriormente, esta pesquisa parte da consideração de que os padrões morfosintáticos em sistemas numerais consistem em um interessante objeto de estudo para a linguística formal e podem fornecer evidências sobre a ligação entre a faculdade da linguagem e a faculdade numérica.

Para Corver et al. (2007), o sistema numeral de uma língua natural pode ser avaliado como uma tecnologia cognitiva, tal qual são os sistemas de escrita, possibilitada pela faculdade numérica que é a capacidade humana de reconhecer cardinalidades no mundo e de construir expressões linguísticas para expressá-las. Sistemas numéricos e sistemas numerais são recursivos, assim como a linguagem humana, já que a partir de unidades básicas podem potencialmente produzir infinitos números e estruturas sintáticas que os expressem, i.e., numerais.

Chomsky (2005) sugere que mesmo os números naturais podem ser resultado de uma simples restrição na faculdade da linguagem. Para esse autor a faculdade numérica é uma abstração da faculdade da linguagem pois mantém a infinitude discreta, isto é, a capacidade de compor infinitas expressões a partir de um conjunto finito e segmentado de elementos. Esta pesquisa parte dessa constatação de Chomsky (2005) e considera que a análise linguística formal de sistemas numerais diversos é uma importante contribuição para esclarecer a relação entre a faculdade numérica e a faculdade da linguagem, ou ainda, para verificar se a postulação da faculdade numérica é realmente necessária. Hurford (1987) propõe que somente a faculdade da linguagem é necessária para dar conta do fenômeno do surgimento de sistemas numéricos e sistemas numerais nas sociedades humanas.

Buscou-se neste capítulo apresentar as teorias escolhidas para o desenvolvimento desta pesquisa. Por isso, o capítulo conta com três subseções. A seção 4.1 consiste na exposição dos principais pressupostos teóricos e dos estágios do desenvolvimento da teoria da Gramática Gerativa com o objetivo de apresentar o paradigma científico em que essa análise se insere. Em seguida, a seção 4.2 conta com a apresentação da teoria da Estratégia

de empacotamento, proposta por Hurford (1975, 1987, 2007) que propõe uma estrutura sintática na formação de numerais. Essa seção também conta com breves considerações sobre uma abordagem minimalista sobre essa teoria para investigações futuras, ainda que esta pesquisa não se comprometa a implementar essa formalização. Por fim, a seção 4.3 apresenta teoria semântica composicional (ROTHSTEIN, 2013, 2017) para a análise de sistemas numerais e a justificativa da escolha dessa teoria em detrimento de outras.

A seção a seguir conta com a apresentação da teoria da gramática gerativa. Por isso, apresenta seus principais pressupostos teóricos, apresenta alguns estágios da teoria ao longo do seu desenvolvimento e apresenta a resposta a algumas críticas ao modelo.

4.1 Pressupostos teóricos e desenvolvimento da teoria da gramática gerativa

Esta seção justifica a escolha da abordagem da teoria da gramática gerativa em relação ao objeto de estudo que será analisado e conta com uma breve apresentação do desenvolvimento da teoria e dos seus principais pressupostos teóricos. A pesquisa buscou uma descrição detalhada de um conjunto particular de dados com foco em analisá-los e explicá-los a partir da abordagem da linguística formal. A motivação para a definição da abordagem foi a beleza intrínseca à formação composicional dos sistemas numerais, então, o modelo de análise aplicado nesta pesquisa parte da tradição de atribuir análises composicionais para análise de expressões linguísticas, como é feito, por exemplo, na linguística gerativa e na semântica formal.

A escolha da abordagem gerativa se justifica pelos padrões encontrados nas línguas da família Tupi na formação dos sistemas numerais. Há uma sistematicidade clara entre o léxico dos numerais cardinais e sua estrutura morfossintática através da família Tupi. Os sistemas numerais das línguas dessa família parecem ser em sua maioria de base 5 e têm características peculiares, como o uso de expressões de partes do corpo como mão, dedos e pé para se referir a cardinalidades específicas. Essas características, entre outras, são recorrentes na família e sugerem um padrão que precisa ser mais bem descrito e analisado para uma melhor compreensão do conhecimento linguístico requerido para a formação dessas expressões. A fim de amenizar a bem conhecida tensão entre descrição e análise de dados linguísticos, esta investigação buscou explicar a composicionalidade e os itens lexicais e morfossintáticos envolvidos na formação de numerais a partir de uma proposta

que contemple a interface sintático-semântica possibilitada pela teoria da gramática gerativa.

Chomsky (2006) defende que é razoável supor que a principal contribuição do estudo da linguagem venha a ser a compreensão que dá sobre o caráter dos processos mentais e das estruturas que esses processos formam e manipulam. Assim, os padrões verificados na formação de numerais motivaram essa investigação já que as propriedades específicas dessas estruturas devem ser explicadas e alocadas na teoria gerativa. É preciso verificar se essas estruturas seriam fruto de princípios, parâmetros ou mesmo condicionadas por fatores não relacionados à faculdade da linguagem, como mencionado por Chomsky (2005).

Em 1957, com Chomsky, inicia-se uma mudança no domínio científico da Linguística, disciplina que antes se relacionava às dimensões histórica e social das línguas naturais. Nesse sentido, para além de descrever a história, o estilo e o uso das línguas, Chomsky considera que o objetivo da linguística deve ser descrever os processos mentais que geram as estruturas da linguagem. Nesse cenário, Chomsky desenvolveu a primeira versão da teoria gerativa, segundo a qual as estruturas linguísticas seriam criadas via aplicação de regras inconscientes determinadas pela faculdade da linguagem.

A gramática gerativa, em todos os seus estágios de desenvolvimento, sempre assumiu que os seres humanos são dotados de uma **faculdade da linguagem** que consiste em um componente da mente/cérebro¹¹ especificamente dedicado à língua. Portanto, essa é uma teoria **inatista** e **racionalista** já que assume a necessidade de um componente inato, i.e., dotado geneticamente, para explicar adequadamente fatos linguísticos como a aquisição de uma língua pelas crianças e a existência de universais linguísticos.

Nessa perspectiva teórica, os diferentes fenômenos estruturais que observamos nas línguas naturais são resultado da interação entre o léxico e a sintaxe. Portanto a variação sintática verificada em sistemas numerais através das línguas também é resultado dessa interação. Dessa maneira é possível identificar se existem regras universais ou restrições extralinguísticas compartilhadas que expliquem os padrões e a variação morfossintática

¹¹ Esta análise não contemplará o debate sobre os termos mente/cérebro, mas considera sua importância nas ciências cognitivas.

envolvidos na composição das estruturas verificadas na linguagem naturais, e no caso dos objetivos desta análise, de sistemas numerais.

De forma recorrente, o debate sobre sistemas numerais traz à mente fatos amplamente conhecidos como o resquício de um sistema numérico vigesimal no sistema numeral do francês. Esse sistema numeral apresenta expressões irregulares como *quatre-vingt* (literalmente quatro-vinte) ‘oitenta’ e *quatre-vingt-dix* (literalmente quatro-vinte-dez) ‘noventa’ em contraste com o fato de que a parte francófona da Bélgica e da Suíça tem um sistema puramente decimal com termos como *octante* (literalmente oitenta) ‘oitenta’ e *nonante* (literalmente noventa) ‘noventa’. Esse tipo de fenômeno, ainda que seja digno de investigação, não interessa à linguística gerativa. Processos diacrônicos e sociolinguísticos podem explicar essas variações, mas elas não interessam a uma abordagem linguística que universaliza a faculdade da linguagem, como a teoria gerativa. Em contrapartida, o que interessa à abordagem gerativa é o conjunto de princípios que determina a estrutura do francês e das línguas naturais, e desse modo, também interessa o conjunto de princípios, regras e/ou traços que determinam um sistema numeral de uma língua natural. Analisar aspectos específicos de uma língua, tais quais adjetivos ou perguntas, por exemplo, pode fornecer evidências para uma melhor compreensão da organização geral da faculdade da linguagem e das línguas naturais. Nessa perspectiva, esta pesquisa defende que a análise de sistemas numerais sob uma perspectiva formal também é produtiva para o entendimento geral da linguagem.

A teoria da gramática gerativa, proposta por Chomsky em 1957 e desenvolvida por inúmeros linguistas até os dias de hoje, buscou, e ainda busca, fornecer descrições e análises de línguas que pudessem explicar a **propriedade básica da linguagem**, i.e., o fato de que a linguagem humana é um sistema computacional finito que produz uma infinidade de expressões, cada uma com interpretação no interpretação nos sistemas semântico-pragmático e sensorio-motor (BERWICK & CHOMSKY, 2016). A propriedade básica da linguagem implica em um aspecto importante das línguas naturais, a **recursividade**, definida pelo fato de a língua construir expressões infinitas encaixando constituintes como pode ser observado na sentença *O João amava a Maria que amava o Pedro que amava a Joana*. A recursividade possibilita ainda outra característica importante da linguagem captada pela teoria, o **aspecto criativo da linguagem**, que consiste na capacidade especificamente humana de expressar novos pensamentos e de entender expressões novas

de pensamento via linguagem. Uma língua natural é um produto cultural sujeito a leis e princípios em parte exclusivos da língua em parte reflexos das propriedades gerais da mente.

Considerando a propriedade básica da linguagem e a constatação de padrões morfológicos óbvios em sistemas numerais, esta análise julgou mais adequado aceitar o pressuposto de que existe uma computação que conduz a formação desses itens lexicais do que assumir que são palavras isoladas, cada uma relacionada a uma cardinalidade específica, aprendidas apenas pela memorização do paradigma. Tal abordagem tradicional é verdadeira no que toca aos numerais básicos, cujo significado semântico e ordem intrínseca do conjunto lexical são aprendidos, mas não se sustenta se forem considerados os numerais compostos. Se de fato os numerais compostos fossem aprendidos como uma lista lexical ordenada, isso exigiria muito mais esforço e memória do que considerar que existe uma computação por trás da formação desses sistemas como prediz a propriedade básica da linguagem.

Ao longo do desenvolvimento desse panorama teórico percebeu-se que parte do problema da variação verificada entre as línguas naturais poderia ser melhor explicada com a descoberta de restrições sobre o sistema sobre o sistema biológico de aquisição de linguagem, ou seja, restrições sobre a **gramática universal**, conceito proposto pela teoria que representa uma capacidade inata e universal responsável pela linguagem em seres humanos, i.e. a faculdade da linguagem. É importante destacar que a gramática universal não deve ser confundida com os ‘universais linguísticos’, no sentido de Greenberg (1978), que fornecem evidências importantes sobre a linguagem humana, mas não se referem a capacidade humana de adquirir e usar uma língua. Universais linguísticos são fenômenos comuns a várias línguas naturais como marcação de caso ou marcação de número, por exemplo, mas, sozinhos, não predizem a existência de um aparato biológico universal em seres humanos. Esse órgão mental é proposto pelo modelo teórico proposto pela tradição da gramática gerativa.

Na década de 80, as descobertas linguísticas promovidas pela teoria gerativa se intensificaram e resultaram em um conjunto significativo de restrições sistemáticas sobre a

G.U¹². Esse estágio da teoria ficou conhecido como modelo de **Princípios e Parâmetros** (doravante P&P). Assim a teoria linguística assumiu um aspecto de tabela periódica com átomos se combinando em possíveis moléculas como foi observado por Baker (2002). Esse modelo propõe que a gramática universal é composta por princípios, que consistem em regras que se aplicam da mesma forma em todas as línguas, e parâmetros, que são princípios que possuem uma possibilidade de variação definida e são responsáveis pelas diferenças entre as línguas (CHOMSKY, 1981). Para ilustrar melhor essa categorização estão os exemplos do princípio do sujeito e o parâmetro do sujeito nulo, aquele prediz que toda a língua precisa ter maneiras de marcar um sujeito via as pessoas gramaticais, enquanto esse prediz que há línguas que aceitam sujeito nulo, como o português em ‘*está chovendo*’ e línguas que não aceitam, como o inglês ‘*It is raining*’.

Já nos anos 1990, como o modelo P&P já havia dado conta de uma grande variação entre as línguas, foi possível investigar sua adequação teórica a fim de reduzir as regras e restrições ao menor conjunto possível, e assim, propor condições de computação mais eficientes que exigem um aparato mental mais simples e plausível. Atualmente, a Linguística Gerativa vive o momento do **Programa Minimalista**. Essa abordagem surge com Chomsky (1995) e propõe que a computação sintática é feita via checagem de traços dos itens lexicais. Por exemplo, se um predador qualquer possui um traço nominal não checado, ele precisa se concatenar via merge com um item que possua um traço nominal, e assim a computação é realizada. As propriedades-chave da linguagem humana capturadas por pressupostos do programa minimalista são (Chomsky, 1995):

- i. A sintaxe da linguagem humana é hierárquica e cega com relação à ordem linear, as restrições de ordem linear são reservadas para a externalização das estruturas.
- ii. As estruturas hierárquicas particulares associadas às sentenças afetam sua interpretação.
- iii. Não há limite superior na profundidade da estrutura hierárquica relevante.

¹² É importante destacar que apesar dos grandes esforços empreendidos por linguistas do mundo todo, há ainda um grande conjunto de línguas que não foram analisados satisfatoriamente dentro do projeto de investigação proposto pela teoria gerativa, como é o caso das línguas indígenas sul-americanas que só muito recentemente têm sido contempladas com essas análises. De modo que a descrição e análise ainda constitui uma importante contribuição para o desenvolvimento da teoria.

A partir dessas considerações, Berwick & Chomsky (2016) argumentam que, se essas propriedades estão corretas, (i) implica que qualquer teoria linguística adequada deve ter algum modo de construir expressões hierarquicamente estruturadas, ignorando a ordem linear; (ii) implica que a estrutura de certo modo fixa a interpretação no nível do significado ; e por fim, (iii) implica que essas expressões são potencialmente infinitas. Para esses autores, essas são as propriedades mínimas que qualquer teoria sintática adequada deve possuir. É nesse sentido que esta pesquisa buscou nortear sua investigação sobre a sintaxe e a semântica na composição de numerais.

Na abordagem minimalista, e mesmo nos estágios anteriores, há uma operação responsável por construir as estruturas hierárquicas necessárias para a sintaxe da linguagem humana chamada **merge**. Essa operação consiste em pegar quaisquer dois elementos sintáticos *x* e *y* e combiná-los em uma expressão nova que é hierarquicamente menor. Nos meandros dessa abordagem existe uma distinção formal entre merge externo, que une dois objetos *x* e *y* que sejam disjuntos, e merge interno, quando um é parte do outro. Esse tipo de merge é o que resulta no fenômeno amplamente observado nas línguas naturais de deslocamento de constituintes sintáticos, como, por exemplo, ocorre em perguntas como *João comeu o que?* e *O que João comeu?*

Em termos simples, e no que toca aos interesses desta análise, merge é uma operação que forma conjuntos, desse modo, dado um objeto sintático *x* (um átomo sintático como as palavras ou um objeto que seja resultado de um merge) e outro objeto sintático *y*, a operação merge forma um novo objeto sintático $\{x,y\}$. Esse objeto resultante recebe um rótulo via um mecanismo de checagem de traços que busca o traço de núcleo, por exemplo, se *o*; núcleo é *x* e *x* é o verbo ‘escrever’, o conjunto $\{x,y\}$ será um sintagma verbal (para mais detalhes ver CHOMSKY, 1995).

Até onde a linguística moderna logrou alcançar, a faculdade da linguagem é uniforme na população humana. Além disso, a partir do estado de arte da disciplina, observa-se que as propriedades paramétricas da linguagem humana permanecem fixas, variando apenas dentro dos limites prescritos (BERWICK & CHOMSKY, 2016). A partir dessa constatação, é bastante provável que ocorra o mesmo quando o assunto são sistemas numerais, apesar das diferenças morfofonológicas significativas desses sistemas observadas nas línguas naturais. Uma teoria de gramática que busque entender o que é a linguagem, para além de descrevê-la, faz a suposição de que as línguas irão diferir pouco

quando revelarmos seus mecanismos e princípios fundamentais, apesar da considerável diversidade observada na realização superficial.

A partir da consideração de a capacidade da linguagem é a mesma para todo indivíduo humano, infere-se que as línguas naturais são variações superficiais de uma mesma capacidade. Essa abordagem exclui, portanto, noções linguísticas arcaicas de exotismo e superioridade entre as línguas e motiva a investigação sobre a natureza da faculdade da linguagem verificada apenas na espécie humana.

Esta análise que optou por sustentar uma abordagem gerativa sobre fatos de línguas indígenas. A análise defende a relevância da investigação empreendida por inúmeros linguistas durante décadas dentro do programa de investigação da linguística gerativa e pretende contemplar um assunto pouco explorado na literatura que são os sistemas numerais, a despeito de um suposto viés antropológico sugerido pela escolha de duas línguas indígenas como seu objeto de estudo.

Como argumentado em Chomsky (2006), a crença de que a linguística focada em línguas subdescritas, muitas vezes chamada de linguística antropológica, tenha ‘demolido’ os pressupostos da gramática universal não parece correta por dois aspectos importantes. Primeiro, há uma interpretação equivocada das ideias da gramática racionalista clássica, que não sustentava que as línguas seriam iguais, mas que seriam semelhantes apenas em um nível mais profundo a fim de investigar os processos que proporcionam o aspecto criativo da linguagem. Segundo, pois essa crença se baseia em uma interpretação equivocada das descobertas da linguística antropológica. Para além de uma compreensão descritiva dos sistemas numerais, ressaltando a importância desse tipo de abordagem já que boas descrições fornecem bons dados para boas teorias, esta análise buscou alocar os fatos descritos dentro de um panorama científico maior, a linguística gerativa.

É fato que novas evidências têm mostrado que a teoria da gramática gerativa tem falhas, mas parte da investigação atual ainda consiste no interesse sobre universais linguísticos que só podem ser detectados a partir de investigações minuciosas sobre línguas subdescritas. Esse é de fato o trabalho que deve ser desenvolvido pela teoria antes de assumir prematuramente seu fracasso (CHOMSKY, 2006). O programa de pesquisa da Gramática Gerativa é bastante frutífero, já que possibilita a compreensão sobre línguas

tipologicamente muito diferentes, aborda questões não antes consideradas, como inatismo e criatividade, e propõe algumas soluções plausíveis a essas questões.

Finalmente, considerando a propriedade básica da linguagem e a constatação de padrões morfológicos óbvios em sistemas numerais, esta análise julgou mais adequado aceitar o pressuposto de que existe uma computação que conduz a formação desses itens lexicais do que assumir que são palavras isoladas, cada uma relacionada a uma cardinalidade específica, aprendidas apenas pela memorização do paradigma. Tal abordagem tradicional pode até ser verdadeira no que toca aos numerais básicos, cujo significado semântico e ordem intrínseca do conjunto lexical parecem mesmo serem aprendidos, mas não se sustenta se forem considerados os numerais compostos. Se de fato os numerais compostos fossem aprendidos como uma lista lexical ordenada, isso exigiria mais esforço e memória do que considerar que existe uma computação por traz da formação desses sistemas a partir de unidades menores como prediz a propriedade básica da linguagem.

Esta seção apresentou um breve panorama do desenvolvimento da gramática gerativa, passando pelos principais estágios da teoria e apresentando os pressupostos teóricos assumidos por esta pesquisa, como a existência de um órgão cognitivo especificado para língua, i.e. a **faculdade da linguagem**, a postulação de um mecanismo recursivo que une itens em uma organização binária, i.e. **merge**, a **propriedade básica da linguagem**, a **recursividade** e o **aspecto criativo da linguagem**.

A próxima seção apresentará os fundamentos teóricos para a sintaxe de numerais adotada pela análise. Assim conta com uma apresentação da teoria da estratégia de empacotamento proposta por Hurford (1987), discutindo seus pressupostos a partir de dados de línguas como o português e o inglês, entre outras. Por fim, a seção apresenta reflexões sobre uma roupagem minimalista para a proposta de Hurford, buscando uma maneira de implementar suas regras sintáticas em uma computação via merge e checagem de traços.

4.2. A sintaxe interna dos numerais e a estratégia de empacotamento

Esta seção discute a sintaxe envolvida na formação de numerais nas línguas naturais. Para isso, apresenta a teoria da Estratégia de Empacotamento desenvolvida por Hurford (1975, 1987, 2003, 2006). Essa é uma restrição universal que ocorre na formação de estruturas sintáticas que expressam numerais nas línguas naturais. Esse conceito foi desenvolvido dentro do panorama da teoria gerativa, que pretende estabelecer regras econômicas e com adequação semântica que gerem sentenças bem formadas em uma língua. E no caso dos sistemas numerais, essa restrição atua para formar apenas expressões numerais bem formadas. A seção também conta com referências e exemplos de algumas línguas que corroboram a validade dessa teoria. Por fim, discute brevemente a possibilidade de se implementar a formalização minimalista aos fenômenos linguísticos apresentados.

A estratégia de empacotamento é amplamente observada em sistemas numéricos que expressam multiplicação e adição via estruturas sintáticas, como será apresentado em mais detalhes ao longo dessa seção (ver HURFORD, 1987; AMAECHI, 2014; KOŞANER, 2016, VIGNADO, 2017). Trata-se de um mecanismo linguístico recursivo que constrói "pacotes" a partir da unidade básica de maior valor cardinal do conjunto lexical de palavras numerais da língua. As operações sintáticas expressam operações matemáticas, como a adição e a multiplicação, que são necessárias para a formação recursiva de números.

Hurford (1987) observa as construções sintáticas que envolvem os numerais mais comuns nas línguas naturais. Esse autor inicia sua argumentação mostrando estruturas possíveis com numerais em inglês, com exemplos de numerais e mesmo de expressões que não são numerais mas contém uma operação aritmética. É o caso da expressão *five sevens* que não é um numeral, mas é gramatical e implica em uma operação de multiplicação. Essas estruturas são apresentadas em (13) com exemplos do inglês e português.¹³

13) a. *two*

¹³ As árvores sintáticas apresentadas nessa seção foram adaptadas ou inspiradas, no caso das árvores do português, de Hurford (1987). Esse autor faz uso de árvores tripartites e esta análise optou por adaptá-las em árvores binárias como é costume da literatura linguística sobre sintaxe.

b. *vinte e dois* (expressando adição)

c. *two hundred / five sevens* (expressando multiplicação)

Os exemplos em (13) mostram as construções sintáticas com numerais mais comuns acordo com o autor. Como pode-se observar em (13c), a estrutura multiplicadora é formada por um numeral, que é o multiplicador, e um nominal que é o multiplicando. O exemplo *five sevens* não é um numeral, mas é uma expressão gramatical utilizada na situação de recitar a tabuada. Ainda que não seja um numeral, esse exemplo mostra que numerais ocorrem em arranjos sintáticos que expressam multiplicação, mas que nem todo elemento que ocorre nessa posição é fixado como uma base do sistema numeral, *seven* não é uma base, esse elemento recebe marca de plural, o que não acontece com bases multiplicadoras autênticas.

A argumentação do autor sugere que ele considera que o elemento que ocupa a posição de multiplicando é um nominal, ou seja, sugere que a classe a que as bases que ocupam esse lugar pertencem é a classe dos substantivos. Como será visto em mais detalhes a seguir, Hurford (1987) opta por dar um nome especial para as bases que compõem um sistema numeral. O autor as chama de bases morfológicas multiplicadoras. De todo modo, isso mostra que esse autor reconhece as diferenças morfosintáticas entre numerais e bases morfológicas multiplicadoras.

Em (14) são exemplificados os três tipos de expressões linguísticas geradas pelas estruturas apresentadas em (13), que são as expressões aditivas (14a); as expressões multiplicativas (14b); e as expressões complexas (aditiva e multiplicativa) (14c).

14) a. aditivas: *one and one*

b. multiplicativas: *two ones*

c. complexas: *two ones and one*

Ainda que não representem a forma habitual de contar em inglês, os exemplos em (14) são todos gramaticais nessa língua e devem ser considerados como um padrão esquemático para a formação de numerais compostos nas línguas de forma geral. Os exemplos em (15) ilustram essas estruturas com numerais do português e do inglês.

15) a. aditivas: *twenty-two, thirteen, vinte e três, dezoito*

b. multiplicativas: *two hundred, ninety, trezentos, oitenta*

c. complexas: *two hundred and three, cento e quarenta e dois*

Esse autor defende que os padrões de adição e multiplicação surgiram historicamente no desenvolvimento das línguas naturais, e que uma vez inventados e adotados tais padrões, eles foram perpetuados sucessivamente de forma geracional. Mesmo que a investigação sobre a origem dos números e dos numerais seja um tópico muito interessante, esta pesquisa não irá se aprofundar a esse respeito.¹⁴

Ainda assim, a consideração da diacronia na formação de sistemas numerais é relevante. Muitas vezes, sintagmas numerais podem conter palavras sozinhas, morfemas alterados por razões fonológicas diversas, traços de uma composição morfológica antiga e já cristalizada, conjunções expressas ou elipsadas, entre outras irregularidades. Mas os numerais e multiplicadores que constituem uma expressão numeral são sempre associados a valores cardinais em uma análise sincrônica, por exemplo, é bastante evidente que a base multiplicadora *hundred* em inglês vale 100 (10^2) independentemente de sua origem etimologia.

¹⁴ Para saber mais sobre isso, ver Menninger (2011), Hurford (1987, 2006), Epps et al (2012).

Há registros de línguas que apresentam outras operações aritméticas na formação de numerais além da adição e da multiplicação. O latim, o ainu e o yorubá (ver exemplo no capítulo 2) são línguas em que a formação de numerais compostos envolve subtração, mas essas estruturas são claramente menos comuns do que aquelas que usam adição de acordo com Hurford (1987). O numeral em (16) exemplifica a operação de subtração em ainu, uma língua tradicional do norte do Japão.

16)	<i>tupesampe</i>	tup-e-sanpe
	oito	dois.para.dez
		10-2

(HURFORD, 1987, p.87, ex. 3.1.2)

Há também línguas que utilizam a operação aritmética de divisão no sentido de que alguns numerais inteiros são expressados via a fração de outros numerais. Esse é o caso do galês, que contém o numeral *hanner cant*, que significa literalmente ‘meio cem’ e, portanto, expressa o número 50. Hurford atenta ao fato de que estruturas que utilizam essa operação são ainda mais raras do que as que utilizam subtração. Koşaner (2016) propõe um conjunto de regras para numerais que ele chama de fracionários inspirado em Hurford. Para esse autor, as estruturas numerais com essa operação têm uma projeção de fração responsável pela semântica da operação. Essa projeção é análoga à projeção de adição, por exemplo, é uma projeção sintática que carrega uma operação aritmética. Esse, porém, não é o escopo desta pesquisa.

Diante dessas considerações sobre as operações aritméticas envolvidas na formação de numerais nas línguas naturais, o autor propõe então uma abordagem linguística para a formação dos numerais considerando as operações mais frequentes nas línguas naturais, i.e., a adição e a multiplicação. A presença dessas operações em um sistema numeral não implica em um sistema decimal, elas podem ocorrer em sistemas numerais com outras bases, como mostram os exemplos de estruturas aditivas em (17) de Austin (1981) e de estruturas multiplicadoras em (18).¹⁵

¹⁵ As línguas às quais pertencem os numerais apresentados são apresentadas entre parênteses.

17) a. <i>mandu mandu</i>	‘quatro’	2+2	(diyari)
b. <i>parkuna nunara</i>	‘três’	2+1	(ngamini)
c. <i>parkulu kuna</i>	‘três’	2+1	(yarluyandi)

(AUSTIN, 1981 , p. 56)

18) a. <i>deunaw</i>	‘dezoito’	2x9	(galês)
b. <i>tri choue ch</i>	‘dezoito’	3x6	(bretão)

(HURFORD, 1987, p.241, ex. 6.1.4)

Nada prediz o uso de uma base numérica específica para um sistema numeral, mas é fato que todo sistema numeral desenvolvido tem uma base a partir da qual constrói o sistema recursivamente. É fato também que muitas vezes um sistema numeral parece utilizar mais de uma base numérica em sua composição, como o caso do francês apresentado anteriormente. Há também casos em que parece haver apenas um desvio da base do sistema, como o caso dos sistemas numerais do galês e do bretão que utilizam bases auxiliares. Os sistemas numerais dessas línguas são vigesimais, mas os dados em (18) mostram que outras bases, base 9 no caso do galês e base 6 no caso do bretão, também são utilizadas na morfologia dos numerais (HURFORD, 1987).

Alguns sistemas numerais de línguas naturais apresentam bases auxiliares na formação de numerais compostos, tal como o galês e o bretão apresentados acima. Uma base auxiliar é uma expressão associada a uma cardinalidade específica que compõe o sistema numeral de uma língua nos intervalos entre as bases multiplicadoras, base auxiliares também formam estruturas multiplicadoras. O sistema vigesimal do azteca, uma língua antiga que foi falada no México, utiliza o nome dos numerais pra 5, 10 e 15, *macuilli*, *matlacti* e *caxtulli*, respectivamente, como bases auxiliares na formação dos numerais (MENNINGER, 2011). Já o sistema do sumério é sexagesimal, i.e. de base 60, e

utiliza os numerais para 6 e 10 como bases auxiliares (MENNINGER, 2011; IFRAH, 1981).¹⁶

As bases auxiliares são utilizadas nos intervalos entre os multiplicadores formados a partir da base padrão do sistema. Por exemplo, em um sistema quinário, i.e. de base 5 deve conter bases multiplicadoras cujos valores sejam 5^1 , 5^2 , 5^3 e assim por diante. No intervalo entre 5 (5^1) e 25 (5^2) esse sistema poderia representar 10 como uma base auxiliar. Já 25 receberia um nome específico e constituiria um multiplicador no sistema. Assim, algumas multiplicações até o multiplicador seguinte, 10×2 , por exemplo, podem ser realizadas com bases auxiliares, mas não operam no sistema plenamente de forma recursiva como fazem as demais bases multiplicadoras, como por exemplo *mil* em português. Esse parece ser o caso do karitiana e do kamayurá como será apresentado no capítulo 5, os sistemas numerais dessas línguas apresentam bases auxiliares lexicalizadas que ocorrem na formação de numerais compostos.

Os conjuntos de formas para numerais sintaticamente complexos que as línguas naturais adotam seguem um padrão definido. Dessa forma, é desejável uma abordagem descritiva detalhada, geral e econômica que dê conta de padrões linguísticos sincrônicos a partir dos dados de qualquer língua. É na busca dessa abordagem que Hurford (1975, 1987, 2003, 2006) desenvolve sua teoria da Estratégia de Empacotamento. A suposição da universalidade desse mecanismo é, como qualquer outro universal linguístico já proposto, ainda que existam contraexemplos, de maneira geral, bastante bem-sucedida para descrever uma grande variedade de dados.

Quando dois números são somados ou multiplicados em uma expressão numeral para expressar um número maior, a estrutura sintática resultante não é equilibrada no sentido de que um desses números é significativamente maior que o outro, e as línguas preferem marcar bem esse desequilíbrio. É bastante natural formar numerais como *thirty-four* em inglês e *duzentos* em português, mas para uma explicação mais completa sobre isso, é preciso também responder porque expressões como **twenty fourteen* e **vintecento*, respectivamente, não são possíveis nessas línguas, já que supostamente expressam os mesmos valores. E, como observamos nesses exemplos, isso vale tanto para adição, no

¹⁶ Esse sistema numeral foi originalmente concebido a partir de símbolos, tal como os algarismos romanos. Esta pesquisa não contemplará o uso de símbolos. Sobre o sistema numeral sumério, ver mais detalhes em Menninger (2011) e Ifrah (1981).

caso de *thirty-four* vs. **twenty fourteen*, como para multiplicação, no caso de *duzentos* vs. **vinteento*.

De acordo com Hurford (1987), o princípio envolvido no fenômeno de apenas estruturas mais eficientes serem gramaticais parece bastante natural, e pode ser dito da seguinte forma: “quando formar expressões para números maiores, pegue a expressão disponível de maior valor como ponto de partida e então construa expressões a partir dela”. Para esse autor, o nome ‘estratégia de empacotamento’ vem da analogia com a ação de carregar um caminhão, independente do que se vá colocar no caminhão, é mais estratégico, e mesmo lógico, começar colocando os objetos maiores para depois utilizar os espaços menores com objetos igualmente menores. A estratégia de empacotamento é, grosso modo, uma maneira de empacotar quantidades numéricas seguindo uma lógica que parte dos elementos maiores para os menores, independentemente da ordem linear em que esses elementos se apresentem, ou seja, o que importa aqui é a hierarquia dos constituintes e não sua ordem. Isso pode ser constatado ao observar os numerais do alemão *einhundertzweiundvierzig* (literalmente *um cem dois quarenta*) ‘cento e quarenta e dois’ e do francês *cent quarante deux* ‘cento e quarenta e dois’, a ordem é diferente, nessas estruturas a ordem é diferente, mas a base *cem* está acima na estrutura que o numeral *quarenta e dois*.

A estratégia de empacotamento foi concebida por Hurford (1975). Esse mecanismo não é uma regra sintática por definição, mas sim uma restrição que atua na sintaxe e que também implica em regras de formação de sintagmas. Nesse momento da teoria, no espírito da linguística desenvolvida à época, o autor considerou que esse mecanismo era uma espécie de filtro de estrutura profunda que impedia a formação de determinadas estruturas em favor de outras que seriam mais adequadas dentro da lógica de funcionamento do sistema.

Entretanto, depois Hurford (1987) é menos categórico a respeito da natureza do mecanismo que propôs. O autor assume que a estratégia de empacotamento é uma ferramenta de descrição concebida no quadro da linguística gerativa. Essa ferramenta descritiva captura de forma elegante o fato de que algumas combinações sintáticas que expressam numerais são bem formadas e outras não, mas não explica suficientemente a razão da existência desse fenômeno.

Hurford (1987, 2006) sugere um processo histórico gradual de padronização das bases numéricas em sistemas numerais. Assim, a partir das diversas combinações que podem formar um mesmo número é definida a melhor base, i.e, aquela presente no maior número de combinações possíveis. Números maiores são mais versáteis no sentido de permitirem mais combinações em sua formação, por isso é importante selecionar qual base é mais adequada para expressar a maior quantidade de números. Esse processo é ilustrado brevemente em (19), a partir do subconjunto das possibilidades combinatórias da soma de dois números para formar números de 14-19.

19) a. $14 = \{(10+4), (9+5), (8+6), (7+7)\}$

b. $15 = \{(10+5), (9+6), (8+7)\}$

c. $16 = \{(10+6), (9+7), (8+8)\}$

d. $17 = \{(10+7), (9+8)\}$

e. $19 = \{(10+9)\}$

A recorrência do número 10 nas combinações apresentadas faz com que 10 seja a base mais eficiente para a composição recursiva desse pequeno sistema fictício. É preciso atentar para o fato de que se incluirmos no conjunto das possibilidades combinatórias possibilidades de soma de 3, 4 ou 5 números esse conjunto ficaria muito maior, o que, de certo modo, justifica porque a base numérica dos sistemas numerais varia tanto nas línguas naturais.

A hipótese de um processo de padronização gradual foi testada por Hurford (2006) com uma simulação computacional, em que números formados de maneira randômica foram dados a agentes simulados e, depois de rodar algumas vezes, o programa passava a preferir construções mais econômicas e desenvolver restrições para as formações numéricas.¹⁷ A localização desse fenômeno na arquitetura da linguagem proposta pelo

¹⁷ A origem dos sistemas numerais não é o foco desta pesquisa, para ver essa hipótese de maneira mais detalhada ver Hurford (2006).

panorama da linguística gerativa será melhor explorada no capítulo 6, onde são apresentadas as considerações finais desta pesquisa.

Hurford (2006) propõe que a explicação para a estratégia de empacotamento reflete dois princípios naturais aplicados a contagem de objetos. Os princípios são:

- i. Ir o mais longe que puder com os recursos que tem.
- ii. Minimizar a quantia de entidades com que se lida

Ademais, o mecanismo linguístico da Estratégia de Empacotamento depende dos recursos lexicais disponíveis na língua, o que explica a variação nas bases morfológicas utilizadas pelas línguas naturais para expressar a lógica numérica subjacente ao seu sistema numeral.

De qualquer forma, utilizando bases multiplicadoras ou bases auxiliares, a natureza recursiva da linguagem permite que o sistema numeral de uma língua natural encaixe uma estrutura numeral dentro de outra, e dentro de outra, e assim por diante, como por exemplo o numeral do português *dois milhões três mil quatrocentos e cinquenta e três*. É a recursividade que possibilita a potencialidade de expressar números infinitamente em uma língua natural, assim, o fenômeno dos sistemas numerais em línguas naturais deve ser investigado para além da morfologia que a língua em questão utiliza nessas expressões.

Ao elaborar as regras sintáticas envolvidas na estratégia de empacotamento, Hurford distingue duas categorias básicas na formação dos numerais: i) DIGIT, a categoria lexical de numerais básicos da língua; ii) M (MULTIPLIER), a categoria de base morfêmica multiplicadora, que corresponde a um morfema que representa um valor cardinal fixo e ocorre em estruturas multiplicadoras. Há, também, duas regras básicas para a formação dos numerais, que são repetidas abaixo na tabela 1 como apresentadas por Hurford (1987):

Tabela 1 – regras sintáticas na formação de numerais		
NUMBER →	$\left\{ \begin{array}{c} \text{DIGIT} \\ \text{PHRASE (NUMBER)} \end{array} \right\}$	INTERPRETADO COMO ADIÇÃO
PHRASE →	(NUMBER) M	INTERPRETADO COMO MULTIPLICAÇÃO

Nesta análise, optou-se por manter os nomes dos constituintes em inglês nas árvores sintáticas apresentadas para evitar ruídos no uso dos conceitos. NUMBER são os numerais do ponto de vista lexical, que podem ser formados por um DIGIT, por exemplo *dois*, ou pela adição de um PHRASE e um NUMBER, por exemplo *quarenta e quatro*. PHRASE é a estrutura interpretada como multiplicação em que um NUMBER é o complemento de um M (multiplicador). Um PHRASE sozinho é um NUMBER, como por exemplo *two hundred* que estruturalmente é uma estrutura multiplicativa mas também é usado como numeral. Em (20) vemos a estrutura sintática do numeral *six million two thousand five hundred* em inglês ilustrando as expressões numerais formadas por essas estruturas.

20)

Para melhor ilustrar as regras propostas pelo autor, a tabela 2 abaixo resume as regras da estratégia de empacotamento a partir de exemplos do português.

Tabela 2 – Exemplos das regras sintáticas na formação de numerais		
REGRAS		EXEMPLOS
M	Base morfêmica multiplicadora	{-cento}
NUMBER	DIGIT	três
	PHRASE	três x {-nta} = trinta
NUMBER (interpretado como adição)	PHRASE + DIGIT	({-cento} x dois) + três = duzentos e três
PHRASE (interpretado como multiplicação)	M x NUMBER	{-cento} x dois = duzentos

Por causa da estratégia de empacotamento, em qualquer parte de uma estrutura numeral, o constituinte M irmão de um NUMBER deve ter um valor maior que seja múltiplo da base ou uma base auxiliar. Assim, o multiplicador M deve ter uma cardinalidade maior do que a cardinalidade do NUMBER irmão e representar a maior base possível no contexto. Esse princípio se confirma nos dados de numerais como a impossibilidade **vintecento* diante de *dois mil*. O primeiro exemplo não é possível pois o sistema conta com o multiplicador *mil* que é mais eficiente e reduz o número de entidades com que se lida. Em Hurford (2007) o autor sintetiza a estratégia de empacotamento como apresentado em (20).¹⁸

20) Estratégia de empacotamento: **o constituinte irmão de um numeral (NUMBER) deve ter o maior valor possível no contexto**

Segundo o autor, nas estruturas aditivas os constituintes com maiores valores são empacotados mais próximos às posições mais altas da estrutura sintática. Assim, na estrutura os constituintes são formados de baixo pra cima do menor numeral para o maior. Portanto, elas estão à esquerda da estrutura em línguas como o português e o inglês. Mas

¹⁸ Tradução livre da autora desta pesquisa.

esse não é o caso em línguas como o karitiana e o kamayurá que são línguas de núcleo a direita.

Esse mecanismo é responsável pela distribuição das bases morfológicas multiplicadoras de um sistema. Essas bases são organizadas em uma ordem definida no sistema numeral. Ele também garante que exista apenas uma expressão bem formada para cada número, o que otimiza o sistema e impede a formação de expressões como **six thousand hundred* no inglês, por exemplo, em que *hundred* é menor e por isso não poderia ser o constituinte irmão de *six thousand*.

Abaixo, são ilustradas as estruturas sintáticas dos números *duzentos e quarenta e dois* em (21a) e *setecentos mil* em (21b) do português a partir dessa teoria.

21) a.

b.

Como mostra a estrutura em (21a), no português os constituintes de maior valor ocupam as posições mais altas na estrutura: *quarenta* é maior e está acima de *dois*, assim como *duzentos* é maior e está acima de *quarenta e dois*. Esses fatos mostram que a estratégia de empacotamento funciona plenamente na formação de numerais compostos em português.

Amaechi (2014) mostra que a estratégia de empacotamento também opera plenamente nos numerais da língua *po tangle*, língua que pertence à família afro-asiática e é uma língua minoritária falada na Nigéria (CROZIER & BLENCH 1992). Essa autora mostra que a língua possui numerais básicos até 10 e numerais de 11-19 formados pela adição do numeral *gbom̩* ‘dez’ aos outros numerais básicos com a partícula aditiva, nos termos da autora, *salai*, como, por exemplo, o numeral *gbom̩ salai d̩ok* (literalmente *dez e*

dois) ‘onze’. Ela diz ainda que a língua apresenta bases morfêmicas multiplicadoras e estruturas multiplicadoras como a mostrada na estrutura do numeral *lakikintham won payindino* ‘seiscentos mil’ no exemplo em (22) abaixo.

22)

O turco também apresenta um sistema numeral em que a estratégia de empacotamento se aplica. Koşaner (2016) mostra que essa língua possui numerais básicos até 10 e os numerais de 11-19 são formados a partir da junção do numeral *on* ‘dez’ e os demais numerais básicos como, por exemplo, o numeral *on bir* (literalmente *dez um*) ‘onze’. O autor também mostra que as regras da estratégia de empacotamento operam plenamente no turco como mostrado pela estrutura do numeral *kirk üç* (literalmente *dois cem*) ‘duzentos’ em (23a) e do numeral *beşyüzbeş* (literalmente *cinco cem cinco*) ‘quinhentos e cinco’ em (23b) abaixo.

23) a.

b.

O exemplo em (23b) mostra que turco não tem uma partícula aditiva. Sabe-se que, como será apresentado na próxima seção, a operação de adição em expressões numerais tem uma semântica própria. Por isso, esta pesquisa assume que mesmo que não haja um operador de adição marcado morfológicamente na estrutura aditiva de um numeral, há uma estrutura sintática coordenativa em toda estrutura aditiva de numerais que sustenta a semântica dessa operação.

Hurford (2003) postula que todas as línguas com sistemas numéricos expandidos usam a adição e a multiplicação como operações semânticas principais na construção dos numerais compostos. Postula também que as línguas podem usar partículas que marquem a operação de adição, e que essas mesmas partículas também que marquem a operação de multiplicação, mas esses casos são significativamente menos comuns.

No português, a estrutura de adição é possibilitada pelo uso de uma conjunção coordenativa aditiva *e* como em *vinte e três*, o que é muito comum nas línguas naturais. Ionin e Matushansky (2006) demonstraram que os marcadores da operação aritmética de adição são frequentemente conjunções e adposições, como em inglês *and*, em francês *et*, em hausa *da*, em igbo *na*, como ilustrado pelo conjunto de dados (24)-(27), abaixo.

24) Five hundred **and** eight = 508 (inglês)

cinco cem e oito

25) Vingt- **et**- un = 21 (francês)

vinte e um

26) Ashirin **da** bakwai = 27 (hausa)

vinte e sete

27) Iri ato **na** ise = 35 (igbo)

dez três e cinco

O uso de conjunções pode ser observado nos exemplos, e usualmente não há marcação para a operação de multiplicação. Assim, na maioria das línguas, a ausência de uma marca morfológica entre dois elementos dentro de uma expressão numérica indica

uma multiplicação. Ainda assim, é preciso considerar a estrutura, nos termos de Hurford, uma estrutura multiplicadora é NUMBER M, enquanto a uma estrutura aditiva é NUMBER NUMBER.

Esta pesquisa vai adaptar a estrutura aditiva proposta por Hurford (1987, 2007). Vai considerar que ao invés de ser NUMBER NUMBER, a estrutura aditiva é [NUMBER [NUMBER][COORD[NUMBER]]], adicionando, assim, uma projeção de coordenação na estrutura responsável por carregar a semântica da operação de adição nesse tipo de expressão. Em síntese, seguindo Hurford, esta análise assume as seguintes estruturas sintáticas em expressões numerais, apresentadas em representações sintagmáticas e árvores sintáticas em (28).

28) a. NUMBER:

(i) DIGIT

(ii) PHRASE

(iii) [NUMBER[NUMBER][COORD COORD[NUMBER]]]

b. PHRASE: [PHRASE[NUMBER][M]]

A categoria NUMBER se refere aos numerais do ponto de vista lexical, que, mesmo que sejam compostos, ocorrem na sintaxe sentencial como um numeral. Estruturas aditivas são desse tipo. Já PHRASE consiste na estrutura multiplicadora, i.e., no constituinte formado por um multiplicador e um NUMBER em uma relação de multiplicação. Essas são estruturas sintáticas assumidas pela análise que será apresentada no capítulo 5.

Apesar de serem poderosas ferramentas analítico-descritivas, as regras sintáticas de Hurford foram concebidas no estágio da teoria da Gramática Gerativa conhecido como Gramática Transformacional. Como foi apresentado em mais detalhes na seção anterior, atualmente a Teoria Gerativa está no estágio do minimalismo, em que ao invés de regras sintáticas, buscam-se traços gramaticais, sintáticos e semânticos, que a partir do processo de *merge*, motivado por um mecanismo de checagem de traços, formam as estruturas sintáticas geradas por uma determinada gramática.

Ainda que esta pesquisa não alcance uma formalização minimalista para a formação de numerais compostos na sintaxe, julga importante refletir sobre tal implementação. Para tal, seria necessário propor um esquema de traços formais e traços não checados para todos os elementos envolvidos na composição de numerais que fosse adequada e suficiente para i) contemplar os aspectos gramatical, semântico e fonológico dessas expressões e ii) explicar a sua derivação. Contudo, esta análise deixará esse importante passo para a descrição formal de sistemas numerais para investigações futuras, mas ressalta que essa é abordagem bastante relevante para a descrição de sistemas numerais pelo viés da linguística formal.

Em uma classificação categorial dos numerais de algumas línguas germânicas, Booij (2009) afirma que numerais são expressões linguísticas formadas por um sistema iterativo de regras que permite ao falante formar infinitas expressões numerais. Ele propõe que expressões numerais podem ser classificadas em termos de conterem ou não os traços: numeral [Num], nominal [N] e multiplicação [M]. Os exemplos do holandês em (29) ilustram essa classificação.

- 29) a. *éen* ‘um’, *twee* ‘dois’... :[+Num]
 b. *honderd* ‘cem’, *duizend* ‘mil’ :[+N, +Num,+M]
 c. *miljoen* ‘milhão’, *miljard* ‘bilhão’ :[+N, +M]

(BOOIJ, 2009, p.9)

De acordo com essa classificação, os exemplos em (29b) e (29c) carregam o traço [N] e por isso podem receber os sufixos que também se aplicam a outros nominais e também são palavras semelhantes a outros nominais. Esses exemplos também contêm o traço [M], esse traço faz com que essas expressões sejam singulares quando acompanhados de uma expressão numeral [Num]. Já os exemplos em (29c), diferentemente dos exemplos em (29b), não podem ocorrer sozinhos sem um numeral, já que não possuem o traço [Num]. Booij desenvolve uma teoria sobre a derivação dos numerais de línguas germânicas do ponto de vista da morfologia construtivista, corrente da tradição da morfologia distribuída que defende que a morfologia acontece na sintaxe. Esse autor argumenta que numerais projetam sintagmas e, portanto, tem uma estrutura sintática.

Os traços propostos por Booij (2009) sugerem uma direção para a implementação de uma abordagem minimalista na formação de numerais. É plausível assumir que traços gramaticais como nominal [N], numeral [Num], multiplicador [M], aditivo [ADD] sejam úteis a essa abordagem. Tal abordagem poderia também representar uma forma de explicar a derivação sintática em expressões numerais que envolvem outras operações, como divisão e subtração, de maneira mais uniforme e abrangente. Dessa forma, as diferentes operações aritméticas que ocorrem na formação de sistemas numerais nas línguas naturais seria um resultado dos traços formais que os itens lexicais envolvidos nesse processo contêm. Assim, por exemplo, seria plausível postular também traços de subtração [SUB] e de divisão [DIV] para abranger o fenômeno linguístico dos numerais a partir do maior número possível de línguas.

Nesta seção foram apresentadas evidências em favor de tratar numerais compostos como estruturas sintáticas complexas. Assim, foi apresentada a teoria da estratégia de empacotamento criada por Hurford, passando por exemplos que reforçam essa proposta e seus principais pressupostos teóricos. Também foram apresentadas as regras sintáticas para a formação de numerais nas línguas naturais que expressam seus numerais compostos via

estruturas aditivas e estruturas multiplicadoras. Por fim, esta seção apresentou uma breve reflexão sobre uma implementação na formalização minimalista na formação de numerais.

Diante do objetivo de propor uma análise pautada na interface da sintaxe e da semântica, a próxima conta com a teoria da semântica composicional da formação de numerais proposta por Rothstein (2013, 2017). Apresenta também outras abordagens e enumera evidências em favor dessa abordagem semântica específica em detrimento das outras.

4.3. A semântica composicional de numerais

Esta seção apresenta a semântica composicional de numerais apresentada nesta pesquisa. A investigação da interface entre a sintaxe e a semântica na formação de numerais compostos foi realizada a partir da semântica composicional para expressões numéricas proposta por Rothstein (2013, 2017). Esta seção conta com as evidências que baseiam a abordagem; os pressupostos teóricos; a justificativa da escolha dessa teoria em detrimento de outras abordagens semânticas para numerais; e a semântica dos itens envolvidos na formação de numerais compostos. Esta pesquisa defende o poder descritivo e analítico da proposta dessa autora por considerar que tratar numerais como propriedades, aos moldes de Frege (1982), é mais adequado: i) para a análise de numerais compostos por diferenciar numerais como *two* de multiplicadores como *hundred* contemplando uma descrição mais acurada do fenômeno e ii) para a análise de numerais em línguas em que a semântica sentencial de expressões numerais não se resume à modificação nominal (ver seção 5.1)

Esta pesquisa contempla os numerais cardinais, i.e., aqueles que expressam números inteiros e positivos nas línguas naturais. Rothstein (2017) chama essas expressões linguísticas de *cardinal numericals*. Para essa autora toda expressão linguística cujo significado semântico envolva a referência a um número particular é um *numerical*. Desse modo, numa língua existem *cardinal numericals*, *approximative numericals*, *ordinal numericals*, entre outros tipos de *numericals*. Ela faz essa opção pois em inglês *numeral* pode se referir a expressões linguísticas ou algarismos. Em contraste com o termo *algarismo*, esta pesquisa optou por usar o termo *numeral cardinal*, ou apenas *numeral*. Ressalta-se que, ainda que numerais ordinais, como *primeiro*; numerais derivados, como

dobro, entre outros tipos de numerais, sejam interessantes e dignos de uma investigação minuciosa, eles não fazem parte do escopo definido por esta pesquisa.

As expressões numerais que se referem a números inteiros e positivos são frequentemente chamados de numerais ou numerais cardinais (COMRIE, 2011; HAMMARSTRÖM, 2010). Numerais cardinais são palavras ou sintagmas que denotam números cardinais. Esses numerais podem ser simples, i.e. morfemas livres que denotam um número, ou compostos, i.e. aqueles que podem ser analisados a partir de suas subpartes cujos significados são combinados composicionalmente usando operações aritméticas como adição e multiplicação¹⁹. Por exemplo, *sete* é um numeral simples que denota o número 7, já *dezesete* é uma única palavra, mas não é simples, é um numeral composto já que é composto pelo prefixo *dez-* e o número *sete*.

Rothstein argumenta que como prefixos como *dez-* no português ou sufixos como *-teen* e *-ty* em inglês são morfemas presos. Assim, é plausível assumir que numerais como *dezesete* e *seventeen* são formados no léxico já que envolvem processos morfológicos, enquanto numerais como *two hundred* são formados na sintaxe. A autora ainda destaca que numerais compostos podem envolver tanto operações morfológicas quanto operações sintáticas, e que apesar de que seja plausível que alguns numerais são formados na morfologia e outros são formados na sintaxe, não é claro que tipo de operação está envolvida na formação composicional de cada numeral.²⁰ A esse respeito, a análise proposta por esta pesquisa optou por seguir Ionin & Matushanky (2006) que assumem que numerais são compostos na sintaxe a partir de itens lexicais básicos como *six* e *-ty*. Assim, optou-se nesta análise por assumir que questões de morfofonologia estariam alocadas na forma fonológica da arquitetura da linguagem e não envolvem a formação sintática estrutural das expressões numerais.

Rothstein trata numerais cardinais como propriedades, ela adota. A autora adota a *Property Theory* de Chierchia (1985), que, por sua vez, segue Frege (1982), ao assumir que propriedades têm dois modos de interpretação: (i) não saturada, onde operam como predicados $\langle e, t \rangle$, como azul em “o mar é azul” e (ii) saturada, em que operam como

¹⁹ Rothstein (2017) chama os numerais compostos de numerais complexos, mas esta pesquisa opta pelo termo numerais compostos chamando a atenção ao fato de ambas nomenclaturas se referem ao mesmo fato linguístico., i.e., numerais que envolvem operações morfossintáticas e aritméticas em sua composição.

²⁰ A autora pauta sua argumentação na morfologia a partir das noções de morfema livre e morfema preso, para ver mais sobre isso checar Rothstein (2017).

argumentos <e>, como azul em “*azul é minha cor favorita*”.²¹ Assim, seguindo a argumentação dessa autora, vai se apresentar a seguir evidências da distribuição dos numerais que reforçam a consideração de numerais como propriedades.

Numerais cardinais ocorrem em diferentes contextos, como mostram os exemplos do inglês e do português em (30) abaixo.²²

- 30) a. *six is bigger than two*
b. *one cat, three girls, four boys, six hamsters, eighteen stairs*
(ROTHSTEIN, 2017, p.16, ex. 4a-4b)
c. *dez é maior que oito*
d. *duas meninas inteligentes, um homem, três cachorros*

Em (30a) e (30c) os numerais parecem ocorrer como argumentos e em (30b) e (30d) os numerais se assemelham a determinantes ou adjetivos. Note que em (30b) e (30d) os nominais são plurais quando o numeral é maior que um, mas há muita variação translinguística relacionada a essa concordância. Por exemplo, no Árabe padrão os numerais cardinais de 3-10 são seguidos por um sintagma nominal no plural enquanto numerais de 11 para cima são seguidos por um sintagma nominal no singular (ROTHSTEIN, 2017). Essa constatação sobre a variação translinguística na concordância nominal em sintagmas com numerais é bastante pertinente no que toca a análise de línguas da família Tupi, pois essas línguas de maneira geral não expressam concordância de número no sintagma nominal, algumas línguas destacam o plural nominal via morfemas coletivizadores, como o guarani, e outras que não marcam de forma alguma, como é o caso do karitiana.

Diante da busca por análises composicionais para as línguas naturais, é bastante frequente que a semântica formal trate numerais como determinantes (BARWISE &

²¹ As notações de tipos semânticos como <e> e <e,t> são própria da formalização na semântica formal. Esse paradigma assume que para explicitar o sentido semântico da linguagem é preciso propor dois tipos semânticos básicos que são entidades cujo tipo semântico é <e> e valores de verdade cujo tipo semântico é <t>. Assim, por exemplo, nomes próprios como *João* e descrições definidas como *o menino* denotam entidades <e> e sentenças como *João come maçã* denotam valores de verdade <t>. Assim, todos os outros tipos semânticos são derivados desses tipos básico, como predicados que são de tipo <e,t> pois pegam um indivíduo <e> e leva a um valor de verdade <t>. Por exemplo, a propriedade *grande* <e,t> pega o argumento *o menino* <e>, formando *o menino grande*. Para maiores informações sobre a derivação sentencial desses tipos semânticos ver manuais de semântica como Heim & Kratzer (1998).

²² Todos os exemplos apresentados nesta seção foram adaptados ou inspirados, no caso dos exemplos do português, a partir dos exemplos de Rothstein (2017).

COOPER 1981; KEENAN & FALTZ, 1984; HURFORD, 1987; HOFWEBER, 2005). Nessa abordagem, expressões nominais como *two cars* ou *dez meninos* são tratadas como SDs, i.e., sintagmas determinantes (em inglês DPs - *determiner phrases*), assim *two* e *dez* nessas sentenças tem uma interpretação semântica similar à de quantificadores generalizados como *every* e *some* em inglês ou *todo* e *algum* em português e denotam uma relação entre conjuntos. Os exemplos em (31) ilustram esse tipo de denotação para alguns quantificadores.²³

- 31) a. *Todo*: $\lambda P \lambda Q \forall x [P(x) \rightarrow Q(x)]$
 b. *Algum*: $\lambda P \lambda Q \exists x [P(x) \wedge Q(x)]$
 c. *Quatro*: $\lambda P \lambda Q \exists x [P(x) \wedge |x| = 4 \wedge Q(x)]$
 (ROTHSTEIN, 2017, p.17, ex. 6a – 6c)

Em (31a), *todo* denota uma relação entre dois conjuntos em que o primeiro é subconjunto do segundo, ou, pega dois predicados P e Q e determina que entidade x relacionada a P está contida na entidade x relacionada a Q. Em (31b) *algum* denota uma relação entre dois conjuntos cuja interseção não é vazia, i.e, pega dois predicados P e Q e determina que a cardinalidade de x é maior que 1. Já em (31c) *quatro* denota uma relação entre dois conjuntos cuja interseção contém 4 membros, i.e, pega dois predicados P e Q e determina que a cardinalidade da entidade x relacionada a esses predicados é 4.

Nessa notação as letras maiúsculas P e Q representam predicados, assim, nessa perspectiva que assemelha numerais a quantificadores generalizados, a interpretação de sentenças com esses quantificadores é ilustrada em (32).

- 32) a. *Todo gato bebe leite*: $\forall x [\text{gato}(x) \rightarrow \text{bebe.leite}(x)]$
 b. *Alguns gatos bebem leite*: $\exists x [\text{gato}(x) \wedge \text{bebe.leite}(x)]$
 c. *Quatro gatos bebem leite*: $\exists x [\text{gato}(x) \wedge |x| = 4 \wedge \text{bebe.leite}(x)]$
 (ROTHSTEIN, 2017, p.17, ex. 6a – 6c)

Como se pode observar, (32a) afirma que todos os gatos bebem leite, ou seja, o conjunto dos gatos é um subconjunto dos bebedores de leite, (32b) afirma que alguns gatos bebem leite, ou seja, a interseção entre o conjunto dos gatos e o conjunto dos bebedores de

²³ Nas fórmulas apresentadas, P e Q representam predicados, x é uma entidade variável, $\forall$ é o quantificador ‘para todo’, $\exists$ é o quantificador pelo menos 1, $\rightarrow$ é o operador lógico ‘se então’ e $\wedge$ é operador lógico de união.

leite não é vazio e, por fim, em (32c) o numeral é tratado como um quantificador existencial enriquecido com uma restrição e afirma que quatro gatos bebem leite, ou seja, que a interseção entre o conjunto dos gatos e o conjunto dos bebedores de leite contém quatro membros.

Contudo, apesar da semelhança superficial entre numerais cardinais e quantificadores como *todo* e *algum*, Rothstein (2017) mostra que existem boas evidências que mostram que numerais não são determinantes. Essa autora defende tratar numerais cardinais como os apresentados em (30b), (30d) como modificadores predicativos. Heim (1982) e Kamp (1981) apresentaram evidências para tratar sintagmas indefinidos como *um gato* como predicados com um determinante sintático *um* que não denota um determinante.

Sintagmas cardinais, i.e., aqueles que contém um numeral cardinal, são de fato bastante parecidos com sintagmas indefinidos, como ilustrado nos exemplos em (33).

33) a. *There are two cats in the barn (but I have never seen them).*

b. *Mary has two cats asleep on her lap.*

(ROTHSTEIN, 2017, p.18, ex. 9a – 9b)

c. *Tem dois gatos na garagem, mas eu não os vi.*

d. *Maria ama seus dois gatos.*

Em (33a) e (33c), observamos que sintagmas cardinais podem ocorrer em construções com *there+be* em inglês e com *ter (existencial)* em português e que, portanto, são indefinidos. Em (33b) e (33d) mostram que esses sintagmas podem ser complementos de uma construção possessiva. Ademais, os exemplos em (34) mostram que sintagmas indefinidos e sintagmas cardinais podem ocorrer como predicados onde verdadeiros sintagmas determinantes não podem.

34) a. *O palestrante é um professor do departamento de linguística.*

b. *Os habitantes da toca são quatro gatos.*

c. # *Os hóspedes são alguns estudantes* (em uma leitura não partitiva).²⁴

d. # *Os alunos são alguns quatro estudantes.*

²⁴ O símbolo # marca a infelicidade da sentença, indica que estrutura sintática pode ser gerada mas que a sentença não tem uma interpretação natural plausível.

Em (34a) *professor do departamento de linguística* denota o conjunto de indivíduos que são professores no departamento de linguística e a sentença afirma que o professor tem essa propriedade. A sentença em (34b) tem a mesma estrutura semântica, i.e., uma propriedade sendo aplicada a uma entidade, nesse caso *quatro gatos* denota o conjunto de entidades plurais que tem a propriedade de ser gato e de ter a cardinalidade 4 e a sentença afirma que a entidade plural que são os habitantes da toca está contido no conjunto denotado por quatro gatos.

Rothstein argumenta que os dados de (30) a (34) sugerem que sintagmas quantificadores, i.e., aqueles introduzidos por um quantificador, quando são argumento de um SD denotam quantificadores generalizados enquanto sintagmas indefinidos e sintagmas cardinais são predicados genuínos. Assim, se esses sintagmas são predicados e não quantificadores generalizados, eles não são nucleados por determinantes e numerais não são determinantes. Numerais nus ocorrem como predicados em posição de predicados como mostra (35).

35) a. *My reasons for saying this are four*

(ROTHSTEIN, 2017, p.19, ex. 11)

b. *Meus motivos para essa opção são dois*

Numerais também não ocorrem necessariamente em posição de determinante, mesmo quando são parte de um sintagma nominal argumental como mostram os dados em (36).

36) a. *The four cats lay on the sofa.*

b. *Every two students got a questionnaire to fill in.*

(ROTHSTEIN, 2017, p.19, ex. 12a - 12b)

c. *Os dois cachorros estão no quintal.*

d. *A prova foi em dupla e cada dois estudantes recebeu uma folha de respostas.*

Em (36a) e (36c) a posição de determinante não é preenchida pelo numeral, mas sim pelos artigos *the* e *o*. Já em (36b) e (36d) a posição de determinante é ocupada pelos quantificadores *every* e *cada* e o numeral ocupa a posição de um adjetivo atributivo. Os dados em (35) mostram numerais em posição de adjetivo predicativo e os dados em (36) mostram que eles também podem ocorrer em posição de adjetivo atributivo. Baseada

nessas evidências, Rothstein defende que numerais cardinais são adjetivos e não determinantes.

Esta pesquisa não categorizará os numerais necessariamente como adjetivos, isso se deve ao fato de que isso não reflete alguns fatos linguísticos verificados nas línguas aqui analisadas. Esta análise, portanto, vai preferir caracterizar os numerais apenas como predicados de tipo $\langle e, t \rangle$ e não como adjetivos. Vai se evitar a inclusão dessas expressões em uma classe de palavras específica porque entende que a classificação de itens lexicais em classes abarca uma grande variação nas línguas naturais e porque a argumentação de Rothstein reproduzida aqui sugere que numerais são adjetivos baseada nos dados do inglês e em outras línguas os numerais podem pertencer a outras classes de palavras. Por exemplo, como é argumentado na seção 5.1, no karitiana há evidências em favor de considerar que os numerais funcionam como adjuntos adverbiais, e não como adjetivos.

Diante das evidências apresentadas, Rothstein (2017) propõe que numerais cardinais nascem como predicados de tipo $\langle e, t \rangle$ e que uma expressão numeral n tem a interpretação semântica apresentada em (37).

37) Cardinal n : $\lambda x. |x| = n$

(ROTHSTEIN, 2017, p.20, ex. 15b)

A formalização em (37) diz que um cardinal qualquer n tem uma semântica que pega uma entidade x e afirma que a cardinalidade dessa entidade x é n . Assim, a denotação do numeral *dois* por exemplos seria $\lambda x. |x| = 2$.

Numerais, porém, também podem ocorrer como argumentos e como numerais nus em outros contextos. Podem ocorrer, por exemplo, em uma sequência de numerais como em (38a), em afirmações matemáticas como em (38b) e (38c), em afirmações sobre a propriedades dos números como em (38d-f).

38) a. *Um, dois, três, quatro.... dez... onze, doze, treze...*

b. *Dois vezes dois é quatro.*

c. *A raiz de quatro é dois.*

d. *Um e três são os dois primeiros números ímpares.*

e. *Dois é o único número primo par.*

f. *Dezoito é o nono número par na sequência dos números naturais.*

Os exemplos em (38) mostram que numerais podem ocupar posições argumentais. Isso indica que eles têm propriedades, mais especificamente, a propriedade expressada por sua versão predicativa. Por exemplo, o numeral *dois* tem as propriedades de ser par, de expressar a cardinalidade |2|, de ser o único número primo par, de ser o menor número primo, entre outras. Assim, Rothstein diz que é natural pensar que numerais nus, como nos exemplos em (38), denotam objetos abstratos, i.e., os números. Numerais, então, são nomes para números de maneira análoga aos nomes próprios que denotam indivíduos, ou ainda, que denotam os conjuntos de propriedades que esses indivíduos possuem. Por essas razões, não se vai considerar que numerais são modificadores de predicados ou determinantes.

Contudo, existem autores que defendem o contrário, i.e., que defendem que numerais são determinantes (HOFWEBER, 2005) ou modificadores de predicado (IONIN & MATUSHANSKY, 2006). Hofweber (2005) e Ionin & Matushanky (2006) argumentam que numerais nunca denotam objetos abstratos. Seguindo Barwise & Cooper (1981), Hofweber (2005) defende que numerais são sempre determinantes e, sendo assim, não podem denotar uma propriedade abstrata. Já Ionin & Matushanky (2006) propõe que numerais sempre são modificadores de predicado de tipo $\langle e, t \rangle \langle e, t \rangle$, i.e., que são predicados que tomam um predicado $\langle e, t \rangle$ como argumento e devolvem um predicado $\langle e, t \rangle$. Ambas abordagens argumentam que numerais nus e numerais pronominais pertencem à mesma categoria sintática. Para esses autores, numerais nus como os apresentados em (38) são numerais pronominais que precedem um nominal elidido. Assim, esses autores analisam (39a) como (39b),

39) a. *dois mais dois é quatro*

b. *duas coisas mais duas coisas são quatro coisas*

Entretanto, ainda que a paráfrase em (39b) seja plausível, existem outros casos em que essa abordagem não pode ser confirmada, como mostra Rothstein (2017) elencando três argumentos. O primeiro deles consiste no fato de que os dados em (40) mostram que os verbos *count* e *contar* podem ser utilizados de duas formas, a primeira é contar quantos x existem, como em (40a) e (40b), a segunda é nomear uma sequência de números, como em (40c) e (40d).

40) a. *I counted thirteen (books, things, animals)*

b. *I counted to thirteen (#books, #things, #animals)*

(ROTHSTEIN, 2017, p.26, ex. 28a-28b)

c. *Eu contei treze (meninos, livros, animais)*

d. *Eu contei até treze (#meninos, #livros, #animais)*

Em (40a) e (40c) observa-se o primeiro uso dos verbos e esses dados naturalmente sugerem a pergunta ‘treze o que?’. De fato, *countar quanto* é sempre contar quantos x existem de determinado tipo de objeto, sejam livros ou meninos. Esses exemplos de fato conduzem à necessidade de se assumir que o numeral nesse caso modifica uma elipse nominal. Entretanto, os exemplos em (40b) e (40d), em que *count* e *contar* são usados no segundo sentido, exigem o uso da preposição *to* em inglês e *até* em português e o numeral que segue tais preposições deve ser nu. De modo que a pergunta ‘treze o que?’ não é apropriada nesses casos. Contextos modais como em (41) mostram ainda com mais clareza esse contraste.

41) a. *I can count thirteen. – Thirteen what?*

b. *I can count to thirteen – # Thirteen what?*

(ROTHSTEIN, 2017, p.26, ex. 29a-29b)

c. *Eu posso contar cinco – Cinco o que?*

d. *Eu posso contar até 13 – # Treze o que?*

Esses dados mostram que tanto *count* em inglês quanto *contar* em português são verbos que tem dois significados diferentes que são associados a complementos numerais estruturalmente distintos.

O segundo argumento contrário à abordagem de Hofwever (2005) e Ionin & Matushansky (2006) consiste no fato de que a paráfrase apresentada em (39b) é impossível nos contextos de afirmações matemáticas apresentadas em (38b-f) como mostrado pelos dados em (42).

42) a. *two is even*

b. *# two things are even*

(ROTHSTEIN, 2017, p.27, ex. 31a-31b)

c. *três é ímpar*

d. *# três coisas são ímpares (no sentido de ser um número ímpar)*

As sentenças em (42a) e (42c) são verdadeiras enquanto as sentenças em (42b) e (42c) são no mínimo infelizes. Isso mostra que numerais e objetos plurais tem semânticas diferentes. A distribuição de construções do tipo *maior que / menor que e mais que/menos que* também evidenciam essa diferença, como pode ser observado nos exemplos em (43).

43) a. *two is smaller than three*

b. *#two things are smaller than three things*

c. *two things are fewer than three things*

(ROTHSTEIN, 2017, p.27, ex. 32a-32c)

d. *dois é menor que três*

e. *#duas coisas são menores que três coisas*

d. *duas coisas são menos que três coisas*

Já o terceiro argumento contra assumir nominais elididos na denotação de numerais nus consiste no fato de que traços de concordância fornecem evidências de que expressões como *two/dois* e *two things/duas coisas* denotam entidades de tipos diferentes. Isso pode ser observado nos exemplos em (42) e em (43) mencionados anteriormente e nos exemplos em (44), a seguir.

44) a. *There are ten balls in the box. Six are/#is red and four are/#is black*

(ROTHSTEIN, 2017, p.27, ex. 33b)

b. *Há dez bolas na caixa. Seis são/#é vermelhas e quatro são/#é pretas*

c. *Four is/#are bigger than two*

d. *Dois é/#são a raiz de quatro*

Assim, como foi ilustrado pelos exemplos em (44), se o numeral nu, exceto o numeral *one/um* que é intrinsecamente singular, é interpretado como o nome de um número ele exige concordância verbal singular, mas se é interpretado como um numeral prenominal ele exige concordância verbal plural.

Ademais, numerais podem ser contados como mostram os exemplos em (45). Rothstein (2017) destaca que o exemplo reproduzido aqui em (45a) ilustra a forma padrão de ensinar tabuada para as crianças no Reino Unido. Essa mesma observação vale para (45b) no caso do ensino de matemática básica no Brasil.

45) a. *two twos are four*

(ROTHSTEIN, 2017, p.28, ex. 35a)

b. *dois cinco somam dez*

Os dados apresentados em (38) a (45) mostram que, ainda que a abordagem ilustrada pela paráfrase em (39b) faça sentido quando o numeral modifica um nominal elidido, existem contextos em que o numeral precisa ser interpretado como um termo singular. Diante das evidências do inglês e do português que foram apresentadas, e apesar de existirem outras abordagens, esta pesquisa segue a proposta de Rothstein (2013, 2017) que considera que numerais são propriedades e nascem como predicados <et> e também podem denotar objetos abstratos de tipo <n>, os números, comportando-se assim de maneira semelhante aos nomes próprios, i.e., como nomes de números.

Mostradas as principais evidências que motivaram a escolha da abordagem semântica utilizada aqui, esta seção então apresenta a formalização semântica dessa proposta apresentada por Rothstein (2013, 2017). Para propor uma análise semântica aos numerais que trate dos dois modos de uso de um numeral, como um predicado e como um argumento, Rothstein segue a *Property Theory* de Chierchia (1985), que, por sua vez, segue Frege (1982). A autora assume que propriedades têm dois modos de interpretação: (i) não saturada, onde operam como predicados <e,t>, como *azul* em “o mar *é azul*” e (ii) saturada, em que operam como argumentos <e>, como *azul* em “*azul* é minha cor favorita”.

Como mostram os dados em (46), a forma adjetival e nominal da expressão de uma propriedade pode variar morfologicamente ainda que isso não seja uma regra.

46) a. *Mary is wise/ Wisdom is valuable*

(ROTHSTEIN, 2017, p.29, ex. 38a)

b. *João é agressivo/ Agressividade é ruim*

c. *Her house is yellow / Yellow is beautiful*

Numerais, como mostrado pelos dados (38)-(45) anteriormente, apresentam o mesmo comportamento dual das demais propriedades. Rothstein (2017), então, trata a propriedade denotada pelo numeral como um tipo especial de propriedade, a propriedade cardinal, à qual estão relacionadas entidades igualmente especiais, as cardinalidades. Chierchia (1985) assume que propriedades tem uma interpretação como predicado de tipo $\langle e, t \rangle$ que mapeia indivíduos a valores de verdade como em (47).

47) a. *wise*: $\lambda x. \text{wise}(x)$

(ROTHSTEIN, 2017, p.29, ex. 39a)

b. *agressivo*: $\lambda x. \text{agressivo}(x)$

c. *yellow*: $\lambda x. \text{yellow}(x)$

Para esse autor, adicionalmente, os predicados de tipo $\langle e, t \rangle$ são associados a um indivíduo correspondente do tipo de propriedades π . Chierchia assume que os tipos dos indivíduos, ou seja dos elementos que são saturados e não exigem argumento, se distribuem em três grupos: i) indivíduos de tipo $\langle e \rangle$, as entidades comuns como *João* ou *a menina*; ii) proposições de tipo $\langle t \rangle$, proposições são elementos saturados e tem valor de verdade, são verdadeiras ou falsas; e iii) indivíduos de tipo π , relacionados a propriedades como *yellow* em *yellow is beautiful*.

Nessa abordagem, duas operações relacionam propriedades de tipo $\langle e, t \rangle$ e propriedades de tipo π . O operador $\cap$ se aplica a um predicado $\langle e, t \rangle$ e deriva sua entidade relacionada de tipo π , enquanto o operador $\cup$ se aplica a uma entidade de tipo $\langle \pi \rangle$ e deriva um predicado de tipo $\langle e, t \rangle$. Assim, os exemplos em (48) mostram essas operações a partir da semântica das sentenças em (46a) e (46b).

48) a. *Mary is wise*

$\lambda x. \text{wise}(x) (\text{mary}) = \text{wise} (\text{mary})$

b. *Wisdom is valuable*

$\lambda x. \text{valuable}(x) (\cap \lambda x. \text{wise}(x)) = \text{valuable}(\cap \lambda x. \text{wise}(x))$

(ROTHSTEIN, 2017, p.29, ex. 40a – 40b)

Como pode-se observar, $\cap \lambda x. \text{wise}(x)$ é a denotação de *wisdom*, i.e., a propriedade $\langle \pi \rangle$ relacionada ao predicado $\langle e, t \rangle$ em questão. Assim, a denotação semântica de *wise* é $\cup \cap \lambda x. \text{wise}(x)$, ou, apenas $\lambda x. \text{wise}(x)$.

Rothstein (2013, 2017) propõe utilizar a mesma abordagem para explicar o comportamento dual dos numerais, que como foi evidenciado pelos dados em (33)-(45), operam tanto como predicados, quanto como argumentos. Da mesma forma que a interpretação do predicado *wise* é o conjunto de entidades que são sábios, a autora assume que a interpretação predicativa de um numeral n qualquer de tipo $\langle e, t \rangle$ é o conjunto de entidades plurais que contém n partes atômicas. O predicado numeral $\langle e, t \rangle$ *quatro*, por exemplo, denota o conjunto de entidades plurais que tem 4 partes atômicas. Sua denotação é apresentada em (49).²⁵

$$49) \text{Quatro}_{\langle e, t \rangle} : \lambda x. |x| = 4$$

(ROTHSTEIN, 2017, p.30, ex. 41a)

A formalização em (49) diz que o predicado *quatro* toma uma entidade x e define que a cardinalidade dessa entidade vale 4.

Já o termo singular *quatro* que ocorre como argumento de uma predicação é de tipo n e denota a propriedade relacionada ao predicado em (49) e sua denotação é apresentada em (50) abaixo.

$$50) \text{Quatro}_{\langle n \rangle} : \cap \lambda x. |x| = 4$$

(ROTHSTEIN, 2017, p.30, ex. 42a)

De forma análoga, assim como *wisdom* ou $\cap \lambda x. \text{wise}(x)$ é tudo que a propriedade que todas as pessoas que são *wise* compartilham, a propriedade relacionada ao predicado *quatro* apresentada em (50) é a propriedade compartilhada por todas as entidades plurais que contém exatamente 4 partes atômicas. Numerais também são, portanto, as entidades numéricas relacionadas a predicados como o apresentado em (49).

²⁵ A denotação de *quatro* $\langle e, t \rangle$ com a restrição do domínio de x é $\lambda x. |\{y: y \sqsubseteq \text{ATOMIC } x\}| = 4$, mas optou-se nesta seção a manter apenas as fórmulas simplificadas com a notação via cálculo lambda pois são suficientes aos objetivos da análise aqui empreendida.

É preciso ressaltar, que para facilitar a distinção das formalizações apresentadas em (49) e (50), esta análise optou por apresentar a denotação de numerais de tipo $\langle n \rangle$ com o algarismo da cardinalidade relacionada, como representado em (51).

51) *Quatro* $_{\langle n \rangle}$: 4

O uso do algarismo representa a noção de número em termos matemáticos, que é, de certa forma, a entidade numérica expressada por esses numerais. Rothstein (2017) também assume isso, pois o significado pretendido para o numeral de tipo $\langle n \rangle$ se mantém e dessa forma as derivações semânticas são facilitadas. Em síntese, (50) e (51) são sinônimas nesta análise.

Seguindo Rothstein (2013, 2017), esta análise assume que o tipo n dos numerais argumentais é um subtipo do tipo π de Chierchia. Assim, numerais do tipo $\langle n \rangle$ são as propriedades individuais abstratas correlatas dos predicados numerais de tipo $\langle e, t \rangle$. Numerais $\langle n \rangle$ são a instância de um tipo particular de propriedades, as propriedades cardinais. Existem também propriedades predicativas derivadas desse tipo, como *is a prime number* que é do tipo $\langle n, t \rangle$, já que mapeia numerais $\langle n \rangle$ em valores de verdade, assim como propriedades predicativas de tipo $\langle \pi, t \rangle$ como *is valuable property* se aplica a propriedades em geral.

Em síntese, numerais do tipo $\langle e, t \rangle$ são elementos insaturados e expressam propriedades que podem predicar argumentos, enquanto numerais do tipo $\langle n \rangle$ são saturados e podem ser argumentos de outros predicados. Dessa perspectiva, construções que expressam relações entre numerais são predicados que tomam numerais de tipo $\langle n \rangle$ como argumentos, como ilustrado pelos exemplos em (52).²⁶

- 52) a. *é menor que* $_{\langle n \langle n, t \rangle \rangle}$: $\lambda m \lambda n. n < m$
 b. *é menor que três* $_{\langle n, t \rangle}$: $\lambda m \lambda n. n < m(3) = \lambda n. n < 3$
 c. *dois é menor que três* $_{\langle t \rangle}$: $\lambda n. n < 3(2) = 2 < 3$

Outras expressões que expressam relações entre numerais também tomam numerais de tipo $\langle n \rangle$ como as expressões do inglês *plus* e *equals* na sentença *Two plus two equals four*.

²⁶ Como essa expressão pode ser utilizada com argumentos do tipo $\langle e \rangle$ como em *o rato é menor que o gato*, o tipo dessa expressão é de fato $\langle \alpha \langle \alpha, t \rangle \rangle$, mas o crucial aqui é que os dois objetos comparados por essa expressão devem ser do mesmo tipo.

Nesse caso, *two/mais* é do tipo $\langle n \langle n, n \rangle \rangle$ já que mapeia dois numerais de tipo $\langle n \rangle$ em e *equals/é igual* é do tipo $\langle n \langle n, t \rangle \rangle$ já que mapeia dois numerais de tipo $\langle n \rangle$ em um valor de verdade como mostrado pelas denotações semânticas dessas expressões e pela derivação da sentença em (53).

53) a. *plus*: $\lambda m \lambda n. n + m$

b. *plus three*: $\lambda m \lambda n. n + m (3) = \lambda n. n + 3$

c. *two plus three*: $\lambda n. n + 3 (2) = 2 + 3$

d. *two plus three equals five*: $\lambda m \lambda n. n = m (5) = \lambda n. n = 5$

$\lambda n. n = 5 (2 + 3) = 2 + 3 = 5$

(ROTHSTEIN, 2017, p.33, ex. 52a - 52d)

Como pode ser observado em (52) e (53), assim como faz Rothstein, esta pesquisa optou por restringir o domínio das funções expressadas pelos numerais de tipo $\langle n \rangle$ aos números naturais que elas expressam. Isso simplifica as derivações semânticas apresentadas na análise e não implica em uma mudança na interpretação semântica dessas expressões. Por isso, a propriedade *n* associada ao numeral predicativo será daqui pra frente expressa pela notação em algarismos dos números naturais.

Por fim, como foi apresentado nesta seção, para Rothstein (2013, 2017) os numerais operam como as demais propriedades de uma língua natural e, como essas propriedades, têm uma versão básica insaturada em que são predicados e uma propriedade cardinal relacionada expressada por sua versão saturada como argumento. Assim, os exemplos em (54) abaixo mostram as denotações semânticas dessas duas versões dos numerais como serão apresentadas nas análises propostas por esta pesquisa (ver seção 5), lembrando da restrição do domínio mencionada anteriormente.

54) a. $[[\text{quatro} \langle n \rangle]] = 4$

b. $[[\text{quatro} \langle e, t \rangle]] = \lambda x. |x| = 4$

Para além dos numerais básicos, os sistemas numerais das línguas naturais contam com numerais compostos, chamados por Rothstein (2017) e Ionin & Matushansky (2006) de numerais complexos por serem sintaticamente complexos, já que contém uma estrutura interna ainda que operem no inglês como adjetivos simples. Esses numerais são formados

via dois processos, multiplicação e adição, não necessariamente juntos. Por vezes, numerais compostos podem ser formados somente com adição como em *dezessete* ou somente com multiplicação como em *trinta*, e por vezes, podem ser formados com as duas operações como em *trinta e dois*. Construções aditivas podem ser formadas via afixação (55a), justaposição (55b, 55c) ou coordenação (55d) como mostram os exemplos abaixo.

- 55) a. *fourteen* : four + {-teen}
 4 + 10
- b. *dezoito*: dez + oito
 10+8
- c. *twenty-two*: twenty + two
 20 + 2
- d. *vinte e dois*: [vinte [e [dois]]]
 20 + dois

Vai se observar o numeral *twenty-two* (55c) para ilustrar melhor a semântica de estruturas aditivas. A denotação semântica de *twenty-two* é formada composicionalmente pelas denotações de *twenty* e de *two* mais um operador de adição elipsado, apresentado em (56a). Assim, a derivação composicional necessária para formação desse numeral é apresentada em (42).

- 56) a. [+]: $\lambda m \lambda n. n+m$
- b. *twenty-two*: $\lambda m \lambda n. m+n$ (20)
 $\lambda n. 20+n$ (2)
 $20+2 = 22$
 (ROTHSTEIN, 2017, p.35, ex. 59a – 59b)

Rothstein destaca que a operação expressa em (56a) pode ser tanto sintática quanto lexical. No entanto, essa autora prefere considerá-la lexical por razões que não serão abordadas aqui.²⁷ Para a autora, em inglês, as dezenas nos numerais são formadas via multiplicação e os numerais até 99 são formados a partir de dezenas mais numerais

²⁷ Para ver a argumentação da autora, ver Rothstein (2017).

simples. Para ela os numerais até 99 não são recursivos em inglês já que assume que são formados no léxico.

Ainda que esta pesquisa não se proponha a discutir em profundidade questões sobre as atribuições do léxico e da sintaxe, preferiu-se assumir aqui, seguindo Ionin & Matushansky (2006), que os numerais compostos são todos formados na sintaxe, mesmo que outras análises sejam igualmente plausíveis. Entendemos aqui que esse é um problema teórico que passa pela escolha de uma teoria sintática que seja mais ou seja menos lexicalista. Com os recentes avanços da teoria Gerativa, a partir do advento do programa minimalista, é notável uma abordagem cada vez mais focada em um léxico composto por traços gramaticais, fonológicos e semânticos do que um léxico composto por expressões já formadas. Assim, esta análise busca alinhar-se com essa abordagem por entender que numerais compostos são formados por partes menores na sintaxe.

Voltando à formalização semântica assumida por esta pesquisa, Rothstein (2017) afirma que assim como o inglês, muitas línguas com um sistema numeral expandido usam um sistema recursivo para expressar potencialmente uma sequência infinita de números. Para isso, a língua seleciona algumas potências do número que define a base do sistema que são combinadas com outros números em uma construção multiplicativa. A essas potências que formam um sistema numeral, como por exemplo {-cento} (10^2), mil (10^3), milhão (10^6), a autora dá o nome de *lexical powers*, que esta pesquisa optou por traduzir livremente como *multiplicador lexical*. Essa categoria é fundamental para qualquer língua que permita uma contagem sequencial potencialmente infinita.

É preciso ressaltar que os multiplicadores lexicais de uma língua variam através das línguas naturais. Por exemplo, o sistema numeral do inglês é decimal e apresenta os seguintes multiplicadores lexicais: *hundred* (10^2), *thousand* (10^3), *million* (10^6) e *billion* (10^9). Existe também *myriad* (10^4), ainda que esteja caindo em desuso, mas o fato é que essa língua não contém multiplicadores lexicais para 10^5 , 10^7 e 10^8 . No inglês também existem outros multiplicadores lexicais que não são decimais, como *dozen* (12^1) e *score* (20^1), o que mostra que o mecanismo de criar multiplicadores lexicais ajuda a língua a expressar quantidades mesmo que tais multiplicadores não se relacionem à base numérica do sistema numeral da língua em questão.

É bastante óbvio porque um sistema numeral não expressa números *ad infinitum*. Para essa autora, a potencialidade de contar infinitamente reside no fato da língua adotar potências expressadas por multiplicadores no sistema numeral, permitindo assim que potências da base numérica do sistema sejam lexicalizadas *ad infinitum* compondo o conjunto de multiplicadores lexicais desse sistema numeral. Como no caso do inglês, o sistema numeral permite que, caso seja necessário, a potência 10^{234556} que expressa um número enorme, que beira o inimaginável, possa receber um nome e se consolidar como um multiplicador lexical desse sistema.

Os multiplicadores lexicais tem diferenças morfossintáticas bastante marcadas que os distinguem dos demais numerais cardinais de uma língua. E é baseada nessas diferenças que Rothstein (2013, 2017) opta por atribuir denotações semânticas distintas a numerais predicativos, a numerais argumentais e a multiplicadores lexicais, diferentemente do que fazem Ionin & Matushansky (2006) ao propor uma teoria semântica em que numerais cardinais e multiplicadores têm a mesma denotação semântica. Essas autoras preferem uma análise semântica uniforme em detrimento de uma análise que contemple as diferenças morfossintáticas entre os elementos envolvidos na formação de um sistema numeral. Para esta pesquisa, uma teoria semântica mais alinhada à morfossintaxe dos numerais é mais adequada para uma proposta de interface entre sintaxe e semântica, já que contempla também questões não semânticas. São apresentadas a seguir as diferenças morfossintáticas que sustentam a abordagem de Rothstein.

Contrariamente aos outros numerais, como mostrado nos dados em (42)-(45), os multiplicadores lexicais nunca ocorrem nus. Em inglês eles são sempre precedidos por um numeral, ou por um determinante definido ou indefinido como mostrado em (57).

- 57) a. *Two/One/A hundred people stood in line*
b. *# Hundred people stood in line*
c. *The hundred lucky people who got tickets came to the show*

(ROTHSTEIN, 2017, p.37, ex. 62a - 62c)

Se pensarmos no numeral *two hundred*, podemos observar que essa estrutura é uma estrutura multiplicativa que expressa a operação 2×100 . Apenas multiplicadores lexicais

podem ser precedidos por numerais em um sintagma nominal, numerais cardinais não, como é observado nos dados do português e do inglês em (58).²⁸

- 58) a. *two hundred cats*
b. *two score cats*
c. *# two twenty cats*
d. *dois mil gatos*
e. *# dois vinte gatos*

Ademais, o contraste entre (58b) e (58c) mostra que trata-se de uma restrição gramatical, e não somente semântica, já que tanto *twenty* quanto *score* denotam 20. É evidente, portanto, que esses itens têm propriedades morfosintáticas diferentes. Além das diferenças morfosintáticas, multiplicadores tem ainda outra característica eles podem ocorrer também como classificadores aproximativos, i.e, podem ser utilizados em construções como classificadores que representam quantias aproximadas como em *hundreds of people read the news*. Os dados em (59) mostram que esses elementos são classificadores aproximativos em inglês quando são pluralizados e em português, também, quando são pluralizados, mas que numerais não.

- 59) a. *Hundreds/thousands/dozens of dogs*
b. *# twenties/# nineties of dogs*
c. *centenas/milhares/ dúzias de cachorros*
d. *# vintes/ # trezes de cachorros*

Os dados em (57) a (59) mostram que o tipo semântico e a denotação dos multiplicadores lexicais precisam ser diferentes do tipo e da denotação dos numerais cardinais, já que foi possível observar diferenças semânticas e diferenças sintáticas. Dessa forma, enquanto o numeral é um predicado de tipo $\langle e, t \rangle$ e pode mudar seu tipo para um numeral argumental de tipo $\langle n \rangle$ via o operador semântico n , um multiplicador lexical como *hundred* ou $\{-cento\}$ em português deve ser do tipo $\langle n \langle e, t \rangle \rangle$, i.e. uma expressão

²⁸ De acordo com Hurford (1987), as construções em (43c) e (43e) não são apenas semanticamente inapropriadas, elas são também agramaticais de acordo com as regras sintáticas que esse autor propõe, já que estruturas multiplicativas exigem um multiplicador e *twenty* e *vinte* são numerais e não multiplicadores.

que se combina com um numeral de tipo $\langle n \rangle$ em uma relação de multiplicação e devolve um predicado $\langle e, t \rangle$. Em (60) é apresentado o passo-a-passo da derivação semântica do número *two hundred thousand* cuja estrutura recursiva é $[[\text{two hundred}] \text{ thousand}]$.

$$\begin{aligned}
 60) \text{ a. } & \text{hundred}_{\langle n, \langle e, t \rangle \rangle}: & \lambda n \lambda x. |x| &= 100 \times n \\
 & \text{b. } \text{two hundred}_{\langle e, t \rangle}: & \lambda x. |x| &= 100 \times 2 \\
 & \text{c. } \text{two hundred}_{\langle n \rangle}: & {}^\cap \lambda x. |x| &= {}^\cap 100 \times 2 \\
 & \text{d. } \text{thousand}_{\langle n, \langle e, t \rangle \rangle}: & \lambda n \lambda x. |x| &= 1000 \times n \\
 & \text{e. } \text{two hundred thousand}_{\langle e, t \rangle}: & \lambda n \lambda x. |x| &= 1,000 \times n \ ({}^\cap \lambda x. |x| = 100 \times 2) = \\
 & & \lambda x. |x| &= 1,000 \times ({}^\cap \lambda x. |x| = 100 \times 2) = \\
 & & \lambda x. |x| &= 1,000 \times 200 = 200\ 000 \\
 & & & (\text{ROTHSTEIN, 2017, p.38, ex. 67a – 67e})
 \end{aligned}$$

Como mostraram os dados apresentados nesta seção, os numerais compostos podem ser construídos via adição e/ou via multiplicação. No caso das estruturas aditivas, é preciso assumir que há um operador de adição $+$ que pode ser expressado lexicalmente ou ser elipsado. Por exemplo, o numeral do inglês *two hundred **and** twenty (+) three* e o numeral do português *dois mil (+) duzentos e quatro* apresentam dois operadores de adição, um marcado e outro elipsado. Essa operação de adição tem a mesma semântica que *plus* apresentado em (53a).

Toda conjunção coordenativa é do tipo $\langle \alpha \langle \alpha, \alpha \rangle \rangle$, i.e., toma um elemento de um tipo $\langle \alpha \rangle$ qualquer, e devolve um elemento de tipo $\langle \alpha, \alpha \rangle$ que então toma outro elemento de tipo $\langle \alpha \rangle$ e devolve uma expressão de tipo $\langle \alpha \rangle$. A coordenação em um numeral composto, seja explícita ou elipsada, expressa a operação de adição, e, assim, é definida como uma relação de soma entre os elementos envolvidos. Na formação de numerais, a coordenação toma dois numerais n e m e soma as suas cardinalidades. A ordem dos elementos somados não interfere no resultado, mas os numerais normalmente se apresentam em uma ordem fixa, que é determinada pela língua em questão.

Por fim, como apontado por Rothstein (2017), não é necessário que o falante tenha uma teoria ou uma opinião sobre o que são os números para usar numerais. Ela ressalta ainda

que o que interessa para investigação linguística não é a questão filosófica ‘O que são números?’ mas sim a questão linguística ‘Como se pode modelar o significado dos numerais para poder explicar seu comportamento gramatical e linguístico?’. A autora destaca, como foi mencionado na introdução, que para todo sistema numeral de uma língua natural, é preciso investigar:

- a)** Quais são os numerais básicos e compostos?
- b)** Como numerais compostos são construídos e quais são os processos aritméticos envolvidos na sua composição
- c)** Os numerais refletem esses processos aritméticos em sua estrutura?
- d)** Essas expressões são construídas no léxico ou na sintaxe?

Essas questões orientam a investigação semântica conduzida na próxima seção sobre os numerais do karitiana e do kamayurá. Por isso, para responder (a) são descritos e analisados os sistemas numerais dessas línguas a fim de elucidar a morfossintaxe desses sistemas (ver capítulo 5). Para responder (b) e (c) são explorados os paradigmas morfossintáticos dos numerais nessas línguas a fim de elucidar as operações aritméticas envolvidas (ver capítulo 5). Para responder (d), são apresentadas considerações sobre a inserção desses fenômenos na arquitetura da linguagem proposta pela teoria da Gramática Gerativa (ver capítulo 6).

Nesta seção foram apresentadas as evidências que embasam a semântica de numerais proposta por Rothstein (2013, 2017) e justificam a escolha dessa abordagem em relação a outras que existem na literatura semântica sobre numerais e os tipos semânticos dos elementos envolvidos em sua formação. A partir das considerações apresentadas, o próximo capítulo traz as análises dos sistemas numerais das línguas karitiana e kamayurá. Para isso o capítulo é dividido em duas seções, cada uma dedicada a uma língua específica e dividida ainda em três subseções que consistem na apresentação da língua e do seu sistema numeral, na análise dos dados e nos resultados.

5. Apresentação dos dados e análise

Este capítulo apresenta os dados de numerais das línguas karitiana e kamayurá as e análises da formação dos sistemas numerais propostas para cada língua baseada na interface da sintaxe e semântica orientada pelo panorama teórico apresentado no capítulo anterior. O capítulo está dividido em duas seções, cada uma dedicada a uma das línguas analisadas, assim, a seção 5.1 trata do karitiana e a seção 5.2 trata do kamayurá. Cada seção, por sua vez, está dividida em três subseções, a primeira consiste em uma apresentação da língua e dos dados do seu sistema numeral, a segunda apresenta análise proposta por essa pesquisa e a terceira subseção sumariza os resultados das análises.

Desse modo, a próxima seção apresenta a língua karitiana. A subseção 5.1.1 traz alguns fatos linguísticos relevantes dessa língua e os dados do seu sistema numeral. Em seguida, a subseção 5.1.2 apresenta a análise proposta por esta pesquisa, e, por fim, a subseção 5.1.3 traz os resultados da análise.

5.1. O karitiana

Nesta seção são apresentadas informações sobre a língua karitiana, a subseção 5.1.1 conta com uma apresentação da língua, a discussão sobre o status do morfema {-t} nos numerais dessa língua justificando a opção em analisá-lo como uma posposição e, por fim, tabelas que apresentam os numerais simples e compostos e a decomposição morfológica que esta análise propõe a partir da descrição de outros pesquisadores (LANDIM, 1984, STORTO E ROCHA, não publicado) e da observação do comportamento sentencial dos numerais.

5.1.1. Os numerais karitiana

O karitiana é uma língua indígena brasileira cuja população vive em uma reserva localizada próxima à cidade de Porto Velho, capital do estado de Rondônia. Essa etnia abrange cerca de 350 indivíduos que conservam o uso de sua língua e são em sua maioria bilíngues. A língua karitiana pertence à família Tupi e à sub-família Arikém. Desta sub-

família, Rodrigues (1986) mostra que foram identificadas outras duas línguas extintas, o kabixiana e o arikém, sendo assim o karitiana sua única língua sobrevivente.

Essa língua possui cinco numerais básicos e mais numerais compostos até 20. Contudo, observa-se que os numerais básicos ainda são amplamente utilizados pela maioria dos falantes em seu cotidiano, enquanto os numerais compostos muitas vezes são substituídos pelos numerais do português. Assim, para expressar quantidades maiores, os falantes optam pelos numerais do português. Esse fato é bastante comum em línguas minoritárias, que acabam realizando empréstimos dos numerais das línguas oficiais dos países em que estão inseridas pelo contato da população com o comércio e com a educação formal universal.

Os numerais básicos nessa língua sempre aparecem com o sufixo {-t}, que, antecedido por um contexto nasal e bilabial, é realizado como /p/ como em *sypomp* ‘dois’ e *m̃j̃jymp* ‘três’. Muller et al. (2006) atentam para o fato de que o único contexto em que esse morfema não aparece é em repetição, em que se usa o sufixo enfático {-o}. Por exemplo, o número ‘dois’ *sypomp*, quando é resposta para uma pergunta como *O que você disse?* recebe o sufixo enfático e se torna *sypomo*. Em todos os demais contextos esse sufixo aparece ligado ao numeral necessariamente. Diante dessa característica nos numerais do karitiana, considerou-se importante justificar a seguir porque esta análise considera esse morfema um adverbializador e não um sufixo oblíquo como foi descrito anteriormente na literatura por Storto (1999).

Storto (1999) descreve o sufixo {-t} como um sufixo oblíquo e de natureza posposicional (Storto, 2003). Assim, o numeral formaria com o sufixo {-t} um sintagma posposicional de contagem. Contudo, foi verificada a existência de um sufixo adverbializador {-t} em quantificadores adverbiais como *pitat* ‘muito’ (SANCHEZ-MENDES, 2009), em outros elementos adverbiais como *asoasot* ‘rapidamente’ (VIVANCO, 2014) e em orações relativas adverbiais (ROCHA DA SILVA, 2016). Esses exemplos reforçam a hipótese de que o sufixo presente nos numerais é um adverbializador, já que é mais econômico assumir um sufixo adverbializador do que assumir que a língua apresenta a mesma forma fonológica para sufixos com funções diferentes. Ademais, Sanchez-Mendes (2009, 2014) demonstrou que não há quantificação nominal nessa língua, o que torna estranho assumir que apenas o numeral opera como um quantificador nominal.

Müller et al. (2006) mostram que não há material funcional no sintagma nominal, não há quantificação, marcação de plural, marcação de definitude e nem de gênero. Os autores demonstraram que os sintagmas nominais nucleados por nomes comuns tem denotação neutra para número. Em Karitiana, os nomes são cumulativos, se um animal *pikom* ('macaco') é somado a outro animal *pikom*, dizemos que os dois animais são *pikom*, i.e, o nome nu denota tanto entidades atômicas quanto plurais. Abaixo as sentenças com interpretação (61) singular, e (62), plural, mostram que não há flexão de plural no nominal. As sentenças (61), (62) e (62) também mostram que não há obrigatoriedade de classificadores no sintagma de contagem (MULLER et al.. 2006).

61)

Yn	naka'yt	myhint	pikom
1S	naka-'y-t	myhin-t	pikom
eu	DECL-comer-NFUT	um-ADVZ	macaco

'eu comi um macaco/ eu comi macaco uma vez' (MULLER et al., p.188 2006)

62)

Yn	naka'yt	sypomp	pikom
1S	naka-'y-t	sypom-t	pikom
eu	DECL-comer-NFUT	dois-ADVZ	macaco

'eu comi dois macacos/ eu comi macaco duas vezes' (MULLER et al., p.188 2006)

63)

Mỹjymp	'ejepo	naakat	i'orot
mỹjym-t	'ejepo	na-aka-t	i-'ot-ot-Ø
três-ADVZ	Pedra	DECL-aux-NFUT	PART-cair-DUPL-NFUT

'três quedas de pedra' (MULLER et al., p.188 2006)

Também não há na língua marca de definitude/indefinitude e/ou determinantes definidos/indefinidos, a interpretação do nome no quanto a essas diferenças depende do contexto. Na sentença (64), abaixo, tanto *taso* ('homem') quanto *boroja* ('cobra') podem ser interpretados como definidos ou indefinidos, singulares ou plurais, de modo que o sintagma nominal é neutro para número e para definitude (MULLER et al., 2006).

64)

Taso	naka'yt	Boroja
Taso	naka-y-t	Boroja
homem	DECL-comer-NFUT	DECL-aux-NFUT

'O(s)/um(s)/algum(s) homem(s) comeu/comeram a(s)/uma(s)/alguma(s) cobras'
(MULLER et al., p.189, 2006).

Os nomes em karitiana são diretamente contáveis. Entretanto, classificadores podem ocorrer na língua concordando com o numeral, como no exemplo (65) abaixo:

65)

myhint	kilot	ouro	naakat	i'orot
myhin-t	kilo-t	ouro	na-aka-t	i-'ot-ot-ø
um-	quilo-	ouro	DECL-AUX-	Part-cair-REDUPL-
ADVZ	ADVZ		NFUT	NFUT

'Um quilo de ouro caiu/ ouro caiu em quilo uma vez' (Dado de Muller)

Sanchez-Mendes & Muller (2007) demonstraram que os verbos, assim como os nomes, têm denotação cumulativa, como observado em (66):

66)

João	Iot
João	i-ot- Ø
João	PART-cair-NFUT

'João caiu (uma vez, ou um número indefinido de vezes)'

O dado em (66) ilustra a neutralidade para número do verbo. Deste modo, as possibilidades de interpretação de uma sentença, que varia em definitude e número, são

resultado da cumulatividade, i.e. da capacidade de denotar entidades ou eventos atômicos e plurais.

As posições sintáticas do adverbiais na língua em sentenças matrizes são ADVZ.S.V.O, S.V.ADVZ.O, S.V.O.ADVZ e *S.ADVZ.V.O (ver STORTO, 1999). A sentença abaixo, em (67), e suas variações em (a),(b) e (c) ilustram essas possibilidades, note que o significado não se altera.

67)	mynda	taso	Nampotporaj	ese
	mynda	taso	na-m-potpora-j	ese
	vagarosamente	homem	DECL-CAUS-ferver-FUT	água

a) t

aso nampotporaj **mynda** esse

b) tasó nampotporaj ese **mynda**

c) *tasó **mynda** nampotporaj ese

“homem(ns) ferverá(rão) água vagarosamente” (STORTO, 1999, p.139)

Os numerais respeitam as mesmas restrições sintáticas que os demais elementos adverbiais na língua, como mostra o conjunto de sentenças em (68), abaixo:

68)	Jõnso	nakam'at	ep	ombaky	sypomp
	Jõnso	naka-m-'at	ep	ombaky	sypom-t
	Jõnso	DECL-CAUS-fazer madeira	onça		dois-ADVZ

a) jõnso nakam'at **sypomp** ep ombaky

b) **sypomp** jõnso nakam'at ep ombaky

c) *jõnso **sypomp** nakam'at ep ombaky

‘mulher(es) fez(fizeram) duas onças de madeira/ onças de madeira duas vezes’
(Informante C, outubro, 2014)

O sentido da sentença não é alterado com a mudança de posição do numeral, do mesmo modo que ocorre com demais adverbiais. Além disso, as traduções possíveis para as sentenças apresentadas evidenciam que o tanto uma leitura adverbial quanto uma leitura adnominal do numeral é possível, de forma que se o numeral é um modificador de eventos, a leitura adnominal é resultado das condições de verdade em que o evento modificado pelo numeral é verdade. Assim, levando-se em conta as posições sintáticas dos numerais, que respeitam as mesmas restrições dos demais adverbiais da língua, os numerais dessa língua devem ser elementos adverbiais²⁹. Na tradução da sentença nota-se que o sintagma nominal *jõnso* ‘mulher’ pode ser singular ou plural. Em Karitiana, os nominais são cumulativos, ou seja, denotam tanto entidades plurais quanto entidades atômicas. Se o numeral estivesse modificando o sintagma nominal *jõnso* em algumas dessas condições sintáticas, seria necessário que as condições de verdade da sentença garantissem a existência de duas mulheres, e assim, uma denotação plural, o que não acontece.

Apesar de não ser o escopo desta pesquisa, a breve argumentação apresentada justifica porque, nesta análise, o sufixo em questão será glosado como um adverbializador e não como um sufixo oblíquo. Esta pesquisa optou por considerar que {-t} é um adverbializador como outros pesquisadores sugerem, ao invés de assumir que existem dois morfemas homófonos que respeitam as mesmas condições morfossintáticas mas que possuem uma semântica distinta. Assim, ao invés de analisar o sufixo {-t} presente nos numerais como um morfema específico dessa categoria de palavras, preferiu-se assumir que é o morfema adverbializador presente nos demais itens adverbiais dessa língua.

A seguir, é apresentado o sistema numeral karitiana, apresentando seus numerais básicos e seus numerais compostos. Landim foi o primeiro pesquisador a descrever o karitiana, esse autor encontrou numerais até 20 e até mesmo alguns poucos numerais isolados que representam quantias maiores. De todo modo, esta análise optou por considerar apenas numerais até 20, já que na coleta de dados realizada pela autora desta pesquisa com 6 falantes diferentes do karitiana em 2013, 2014, 2015 e 2018 não foram encontrados numerais maiores que 20 e também por não haver regularidade nos dados desse tipo apresentados por Landim (1984).

²⁹ Esse tema foi abordado em iniciações científicas desenvolvidas pela autora deste artigo, financiada pelo CNPq no período de 2012/2013 e 2015/2016, e 2014/2015 sem bolsa.

Abaixo, a tabela 3 apresenta os numerais básicos do karitiana. Essa língua apresenta numerais básicos de 1 a 5.

Tabela 3 – Numerais básicos do karitiana	
Numeral	Decomposição morfológica
<i>myhint</i> ‘um’	myhin-t um-ADVZ
<i>sypomp</i> ‘dois’	sypom-t dois- ADVZ
<i>mÿjymp</i> ‘três’	mÿjym-t três-ADVZ
<i>otadnamynt</i> ‘quatro’	ota-dna-myn-t outro-NMLZ-coisa-ADVZ
<i>yjpyt</i> ‘cinco’	yj-py-t lp.pl-mão-ADVZ

O sistema de numeral atual contém 5 numerais básicos, isto é, numerais a partir dos quais se compõem os números maiores, como por exemplo o conjunto de um a dez no português. A decomposição morfológica desses numerais sugere que apenas as palavras para os números 1, 2 e 3 são os numerais cardinais originais da língua e as palavras para os números 4 e 5 são incorporações mais recentes.

Os numerais *myhint* ‘um’, *sypomp* ‘dois’ e *mÿjymp* ‘três’ não tem estrutura interna, se referem apenas a uma cardinalidade específica. O numeral *otadnamynt* ‘quatro’ deriva da palavra *ota* ‘outro’ e contém em sua estrutura interna o afixo nominalizador {-dna-} e a palavra *mÿn*, utilizada de forma genérica como a palavra *coisa* em português em ‘me dê essa coisa’. Assim, a tradução literal do termo para 4 é algo como ‘outra coisa’. Já o numeral *yjpyt* ‘cinco’ deriva da palavra *py* ‘mão’ prefixado por *yj* pronome inclusivo de 1ª pessoa do plural, i.e ‘nossa mão’. É bastante comum nas línguas tupi que o numeral que vale 5 derive da palavra mão.

Ainda que *otadnamynt* e *yjpyt* sejam de fato aquisições mais recentes, como sua derivação sugere, esses itens operam plenamente como numerais básicos no sistema da língua como será mostrado na seção 5.1.2 a seguir. Isso é comum nas línguas naturais,

muitas vezes é possível identificar fatos diacrônicos no sistema como o caso do *eleven* e do *twelve* no inglês, como foi mostrado no capítulo 1. Esta pesquisa, entretanto, foca em uma análise sincrônica do sistema numeral do karitiana a fim de descrever e analisar seu funcionamento, a despeito da etimologia dos itens lexicais envolvidos nesse recorte.

Os numerais compostos são apresentados em séries, dos números 6 ao 10, do 11 ao 15 e do 16 ao 20, em tabelas para melhor ilustrar o paradigma da formação dos numerais em karitiana. Então, a tabela 4 a seguir apresenta os numerais de 6 ao 10.

Tabela 4 – Numerais de 6 a 10 em karitiana	
Numeral	Decomposição morfológica
<i>myhint yjpy ota ot</i> ‘seis’	myhin-t yj-py ota ot um-ADVZ 1p-mão outro pegar
<i>sypomp yjpy ota ot</i> ‘sete’	sypom-p yj-py ota ot dois-ADVZ 1p-mão outro pegar
<i>mÿjymp yjpy ota ot</i> ‘oito’	mÿjym-p yj-py ota ot três-ADVZ 1p-mão outro pegar
<i>otadnamynt yjpy ota ot</i> ‘nove’	otadnamyn-t yj-py ota ot quatro-ADVZ 1p-mão outro pegar
<i>yjpy ota tyytap</i> ‘dez’	yjpy ota tyy-tap 1p-mão outro grande(?)

Essa tabela mostra que os numerais de 6 a 9 são formados com um numeral básico e a estrutura *yjpy ota ot*. Como pode ser observado nas tabelas apresentadas nesta subseção, as palavras *yjpy* ‘nossa mão’, *yjpi* ‘nosso pé’, *ota* ‘outro’ e *ot* ‘pegar’ ocorrem de forma sistemática na formação de numerais nessa língua.

Então, o numeral *sypomp yjpy ota ot* ‘sete’ pode ser traduzido literalmente como *dois e pegue outra mão*. Como o valor cardinal de *sypomp* é 2, o valor cardinal da expressão *yjpy ota ot* ‘pegue outra mão’ deve ser 5, já que a cardinalidade total da expressão é 7. Isso é importante para a análise desenvolvida na próxima subseção 5.1.2. Já o numeral *yjpy ota tyytap* ‘dez’ apresenta uma estrutura diferente, esse numeral conta com a expressão *tyytap*.

Os consultores da língua karitiana traduzem essa expressão como *a outra mão acabou*, isso evidencia o sentido dessa expressão. As duas mãos somam 10 e o significado da estrutura desse numeral diz que ambas mãos já foram contadas.

O léxico do karitiana consultado (LANDIM, 1984; STORTO & ROCHA, não publicado) sugere que *tyy* significa ‘grande’. Essa estrutura é melhor explorada na próxima subseção.

A tabela 5 apresenta os numerais de 11 a 15 em karitiana. A partir de 11, os numerais são compostos com *yjpi* ‘nosso pé’, ao invés de *yjpy* ‘nossa mão’.

Tabela 5 – Numerais de 11 a 15 em karitiana	
Numeral	Decomposição morfológica
<i>myhint yjpi ot</i> ‘onze’	myhin-t yj-pi ot um-ADVZ 1p-pé pegar
<i>sypomp yjpi ot</i> ‘doze’	sypom-p yj-pi ot dois-ADVZ 1p-pé pegar
<i>mÿjymp yjpi ot</i> ‘treze’	mÿjjym-t yj-pi ot três-ADVZ 1p-pé pegar
<i>otadnamynt yjpi ot</i> ‘catorze’	otadnamyn-t yj-pi ot quatro-t 1p-pé pegar
<i>myhint yjpi py</i> ‘quinze’	myhin-t yj-pi py um-ADVZ 1p-pé mão

Na série de 11 a 14, os numerais do karitiana são formados a partir de um numeral básico e da expressão *yjpi ot* que significa literalmente ‘pegar pé’. Ao observar o numeral *sypomp yjpi ot* em mais detalhes, nota-se que o valor cardinal total dessa expressão é 12, e como o valor cardinal de *sypomp* é 2, então o valor cardinal da expressão *yjpi ot* deve ser 10. Já o numeral *myhint yjpi py* ‘quinze’ é formada de outra maneira.

Por fim, a tabela 6 apresenta os numerais de 16 a 20. Os numerais de 16 a 19 são formados a partir de um numeral básico e a expressão *yjpi ota ot*, que pode ser traduzida como ‘pegar outro pé’. Por exemplo, no numeral *sypomp yjpi ota ot* ‘dezessete’, *sypomp*

vale 2 enquanto a expressão *yjpi ota ot* deve valer 15, já que a expressão toda remete ao valor 17.

Tabela 6 – Numerais de 16 a 20 em karitiana	
Numeral	Decomposição morfológica
<i>myhint yjpi ota ot</i> 'dezesesseis'	myhin-t yj-pi ota ot um-ADVZ 1p-mão outro pegar
<i>sypomp yjpi ota ot</i> 'dezesesete'	sypom-p yj-pi ota ot dois-ADVZ 1p-pé outro pegar
<i>mÿjymp yjpi ota ot</i> 'dezoito'	mÿjym-p yj-pi ota ot três-ADVZ 1p-pé outro pegar
<i>otadnamynt yjpi ota ot</i> 'dezenove'	otadnamyn-t yj-pi ota ot quatro-ADVZ 1p-pé outro pegar
<i>yjpi ota tyytap</i> 'vinte'	yj-pi ota tyy-tap 1p-pé outra grande-(?)

Como mostram os dados apresentados, os numerais compostos do karitiana são formados pelas palavras *yjpy* 'nossa mão', *yjpi* 'nosso pé', *ot* 'pegar' e *ota* 'outro'. Em contrapartida, os números dez, quinze e vinte não apresentam uma regularidade tão clara. Os números para dez e vinte apresentam o termo *tyytap*, que contém a palavra *tyy* 'grande' e *tap*. O significado isolado de *tap* não foi identificado, mas ao comparar esses dados com os dados do kamayurá, parece ser o caso de que *tap* signifique 'terminar', o que está de acordo com a tradução que os consultores indígenas tentam dar para os numerais (ver 5.2.2).

Esses numerais são composto via justaposição no nível morfológico. Processos de composição por justaposição em numerais são encontrados em outras línguas, como, por exemplo, *dezesesete* (dez+sete) em português ou *tretuno* (trinta+uno) em italiano. Assim, no nível morfológico, no karitiana ocorre o mesmo, já que apesar da estrutura interna dessas palavras, a construção numérica constitui uma palavra em termos de expressar um sentido uno e não divisível, a cardinalidade da expressão.

Nesta seção foram apresentadas informações básicas sobre a língua karitiana e o conjunto de dados de numerais que serão avaliados nesta análise. A seção a seguir, propõe uma estrutura para os numerais baseada na interface sintática e semântica.

5.1.2. Análise da formação de numerais

Esta seção traz a análise dos dados do karitiana a partir do panorama teórico apresentado no capítulo 4. A seção conta com tabelas sobre a recursividade na formação dos numerais compostos dessa língua, árvores sintáticas e derivações semânticas com a finalidade de mostrar os dados dessa língua estão de acordo com as regras sintáticas e semânticas apresentadas no capítulo 4. Isso reforça a universalidade desses mecanismos e a vantagem de uma análise baseada na interface, como será visto a seguir. Optou-se, por começar essa análise pelos exemplos mais regulares desse sistema numeral e, então, vai se propor análises para os outros casos.

Os dados do karitiana mostram uma regularidade no uso das palavras *yjpy* e *yjpi* na formação dos numerais compostos. A tabela 7 abaixo apresenta o paradigma dos numerais *sypomp* ‘dois’, *sypomp yjpy ota ot* ‘sete’, *sypomp yjpi ot* ‘doze’ e *sypomp yjpi ota ot* ‘dezessete’ no karitiana para ilustrar a regularidade na formação desses numerais. Esses numerais foram escolhidos por representarem uma sequência de numerais que aumenta de 5 em 5. Na primeira linha estão os numerais, na segunda linha, a referência da expressão, e, na terceira, a lógica matemática subjacente à expressão numérica em questão.

Tabela 7 – Paradigma da formação de numerais em karitiana				
Numeral	<i>sypomp</i> dois	<i>sypomp yj-py ota ot</i> dois 1pl-mão outra pegar	<i>sypomp yj-pi ot</i> dois 1pl-pé pegar	<i>sypomp yj-pi ota ot</i> dois 1pl-pé outro pegar
Referência
Matemática	2	2+(1x5)	2+(2x5)	2+(3x5)

Como pode-se observar na esquematização feita na tabela 7, e na exposição dos dados na seção anterior, os numerais compostos *sypomp yjpy ota ot* ‘sete’, *sypomp yjpi ot* ‘doze’ e *sypomp yjpi ota ot* ‘dezessete’ apresentam, além do numeral básico *sypomp*, outras *estruturas*, em negrito na tabela, que devem ser explicadas. Esses constituintes são formados pelo morfema {*yj-*} que é a primeira pessoa do plural e as palavras *py* ‘mão’, *pi* ‘pé’ e *ot* que é a raiz do verbo pegar. Assim, essas estruturas repetidas aqui, respectivamente, podem ser traduzidas como *yjpy ota ot* ‘pegue outra mão’, *yjpi ot* ‘pegue pé’, *yjpi ota ot* ‘pegue outro pé’.

É bastante claro que os números nesse paradigma aumentam de cinco em cinco no nível matemático, enquanto a expressão linguística mostra que o valor do número depende do membro do corpo ao qual a expressão se refere se refere. Assim, por exemplo, no número *sypomp yjpi ot* ‘doze’, ao invés de usar *yjpy* ‘nossa mão’, a língua usa *yjpi* ‘nosso pé’. Faz sentido assumir que a tradução dessa expressão seja ‘dois pegue pé’, o que significa que as duas mãos, que somam dez, já foram contadas e são subentendidas na estrutura dos numerais maiores que 10.

Como já foi dito anteriormente, de acordo com Comrie (2005), a base aritmética de um sistema numérico é o valor *n* sobre o qual as expressões numéricas se constroem seguindo um padrão, e é sobre esse valor *n* que as bases multiplicadoras *M* são construídas. A lógica da recursividade numérica é expressa pela fórmula $x.n + y$, ou seja, algum numeral *x* multiplicado pela base *n* e somado a um numeral *y*.

Considerando a quantidade numerais básicos do karitiana e o paradigma exposto na tabela 7, é bastante plausível assumir que a base aritmética do sistema numeral karitiana é 5. Green (1997) afirma que os sistemas de base 20 são quinários, o que torna opaca a análise que essa autora faz desse sistema. A análise proposta nesta pesquisa mostra que a recursividade das bases morfêmicas expressam uma lógica numérica de base 5, e não de base 20, já que não existem evidências de numerais que indiquem essa possibilidade. As coletas de dados realizadas com essa língua não capturaram numerais maiores que 20. Além do mais, esse sistema numérico é restrito, já que vai até 20, e não apresenta regularidade nos poucos exemplos de números maiores que 20 de Landim (1984). Assim, considerou-se que não há evidências suficientes para argumentar que esse sistema é de base 20.

Vignado (2017) argumenta que as bases morfológicas multiplicadoras dessa língua são expressas na língua por partes do corpo que, apesar de valerem 5 cada, são distribuídas de forma fixa e hierárquica nas seguintes estruturas: *yjpy ota ot* "pegue a outra mão" > *yjpi ot* "pegue um pé" > *yjpi ota ot* "pegue o outro pé". Essa distribuição das bases está de acordo com a regra da estratégia de empacotamento que prevê que para qualquer parte de uma estrutura numérica, o irmão de um NUMBER deve ter o maior valor possível. Assim, conforme se aumenta o valor da cardinalidade da expressão, o valor da base multiplicadora também precisa aumentar, é por isso que a estrutura **sympomp yjpy yjpy ot* não é permitida, mas sim *sympomp yjpi ot*, com o uso da base *yjpi* que representa valores maiores. Ademais, se não fosse esse o caso, seria possível fazer referência a valores maiores que dez tanto por meio de *yjpy* quanto de *yjpi*, mas o uso da base *yjpi* está de acordo com essa distribuição pois ela é a maior base.

Hurford (2007) aponta que a estratégia de empacotamento depende do léxico disponível na língua. Nos numerais Karitiana as palavras *ota* "outro", *ot* "pegar" e *tyytap* "todos" são necessárias na formação dos numerais compostos. A expressão linguística dos numerais maiores que dez usa a base *yjpi* "nosso pé", que pressupõe que as mãos já foram contadas, e, para expressar números maiores que quinze, conta-se o outro pé. Por exemplo, no esquema anterior da tabela 7 pode-se observar que a palavra *ota* é usada sempre que o outro membro é contado, nos números *sympomp yjpy ota ot* "sete" e *sympomp yjpi ota ot* "dezessete". Se essa palavra está associada às bases morfológicas *yjpy* e *yjpi*, Vignado (2017) argumenta que ela deve operar sintaticamente como um NUMBER. Na tabela 8 foram organizadas as estruturas multiplicadoras que essa autora propõe no Karitiana, cuja estrutura é composta por uma base morfológica multiplicadora M e um NUMBER.

Tabela 8 – Bases morfológicas do karitiana (VIGNADO, 2017)			
<i>M (number)</i>	<i>Significado</i>	<i>valor</i>	<i>M x n</i>
yjpy (ota)	outra mão	5	5x1
Yjpi	pé	10	5x2
yjpi (ota)	outro pé	15	5x3

Essa análise, contudo, tem um problema. É esperado que um NUMBER seja um numeral, numa relação de multiplicação com um MULTIPLIER (M). Isso é, portanto, uma falha para essa análise no caso do Karitiana, já que *ota* não tem um valor cardinal fixo, como os numerais tem. A palavra *ota* indica outro membro, por exemplo, *yjpi ota* significa o ‘outro pé’. Usar a base *yjpi* significa que as mãos já foram contadas, e como são duas mãos, essa base vale 5x2. Já no caso da estrutura multiplicadora *yjpi ota*, ao contar o outro pé, são 3 membros no total. E ainda que o valor cardinal das expressões seja bastante claro, a semântica composicional não.

Como o karitiana é uma língua de núcleo final, a estrutura sintática do numeral é uma imagem espelhada da estrutura que Hurford (1987, 2007) postulou para o inglês e da estrutura do português que foi apresentada no capítulo 4 na seção 4.2. Mas como foi argumentado, a ordem linear dos constituintes não é relevante, mas sim sua hierarquia.

Esta pesquisa propõe que a palavra *ot*, raiz do verbo ‘pegar’, funciona sintaticamente como uma conjunção coordenativa na formação dos numerais, como mostram as estruturas dos números *sypomp yjpi ot* ‘doze’ em (69a) e de *sypomp yjpi ota ot* ‘dezesete’ em (69b)

69)

As estruturas propostas em (69) capturam as traduções em português que os falantes consultados deram para essas expressões. Para (69a) os falantes forneceram a tradução *dois e pega pé* e para (69b) a tradução é *dois e pega o outro pé*. A raiz do verbo pegar *ot* claramente opera como o elemento que carrega a semântica da operação de adição.

É preciso destacar que o karitiana não tem uma conjunção correspondente à palavra *e* em português ou *and* em inglês, assim, é esperado que a língua expresse a noção veiculada por essas palavras com os recursos que tem. No nível sentencial, os verbos não

ocorrem isolados em karitiana, eles sempre apresentam morfema de modo, concordância de pessoa e marca de tempo, já no caso das estruturas numerais, a raiz do verbo pegar ocorre necessariamente isolada, sem nenhuma marca. Esta pesquisa então, optou por assumir que a palavra *ot* nos numerais carrega a semântica da adição contida na interpretação dessas expressões. A semântica dessa operação foi apresentada na seção 4.3 do capítulo 4 e é repetida em (70) abaixo na versão que soma predicados de $\langle e, t \rangle$.

$$70) [[ot_{\langle e, t \rangle \langle \langle e, t \rangle \langle e, t \rangle \rangle}]] = \lambda P \lambda Q. P+Q$$

Em síntese, a palavra *ot* é a marca morfológica da operação de adição nessas estruturas. Ela soma a cardinalidade do numeral básico com a cardinalidade do restante da expressão. Em uma primeira análise, Vignado (2017) verificou que a palavra *ota* parece operar como um NUMBER. Essa autora sugere isso pois considerou que *yjpi* e *yjpy* seriam as bases morfológicas do karitiana. Dessa forma, *ot* se uniria com uma base morfológica sem qualquer partícula como é esperado nas estruturas que expressam a operação de multiplicação. Abaixo, em (71) são apresentadas as estruturas propostas pela análise de Vignado (2017).

71)

Como pode ser observado nessas estruturas, Vignado (2017) considerou que *yjpi* ‘pé’, *yjpi ota* ‘outro pé’ e, por extensão, *yjpy ota* ‘outra mão’ são PHRASES, nos termos da teoria de Hurford (1975, 1987). Para Hurford, PHRASES são compostos por um NUMBER e um M (multiplicador). Desse modo, *yjpy* e *yjpi* são considerados multiplicadores M enquanto *ota* deveria ser um NUMBER.

Para ver com mais detalhes a proposta de Vignado (2017), a estrutura sintática do numeral *sympomp yjpy ota ot* “sete” é apresentada em (72a) e a sua correspondência semântica em (72b).

72)

Se *ota* for considerado um NUMBER, é possível derivar a semântica da estrutura interpretada como multiplicação a partir da semântica de suas partes a partir dos tipos semânticos de Rothstein (2013, 2017). Na derivação semântica apresentada em (73) foi considerada a expressão *yjpy ota* na proposta de Vignado (2017).

$$\begin{aligned}
 73) \quad & [[\text{PHRASE}_{\langle e, t \rangle}]] = [[M_{\langle n \langle e, t \rangle \rangle}]] ([[\text{NUMBER}_{\langle n \rangle}]]) \quad \text{aplicação funcional} \\
 & [[\text{PHRASE}_{\langle e, t \rangle}]] = [[y_{jpy}_{\langle n \langle e, t \rangle \rangle}]] ([[ota_{\langle n \rangle}]]]) = \lambda n \lambda x. |x| = 5.n \quad (1) \\
 & [[\text{PHRASE}_{\langle e, t \rangle}]] = \lambda x. |x| = 5
 \end{aligned}$$

Ainda que a proposta da autora capture a cardinalidade dessa expressão, considerar *ota* como um NUMBER implica em considerar que essa palavra é um numeral, o que não é verdade. A palavra *ota* significa ‘outro’ e é utilizada também em outros contextos, assim, ela não carrega um valor cardinal e nem tampouco tem um significado associado a numerosidades.

Diante desse problema de análise, esta pesquisa propõe uma reanálise da composição dos numerais do karitiana baseada em Vignado (2017). Para isso, é preciso observar com mais detalhes as expressões que Vignado (2017) chamou de estruturas multiplicativas. A autora buscou privilegiar uma semântica composicional uniforme, mas desconsiderou o problema na análise dessa palavra. Como será desenvolvido a seguir, esta pesquisa mostra que de fato não há uma operação de multiplicação nessas estruturas, mas, elas carregam um valor cardinal fixo como um bloco. Como só ocorrem no sistema numeral, serão consideradas bases auxiliares.

As expressões *yjpy ota* ‘outra mão’, *yjpi* ‘pé’ e *yjpi ota* ‘outro pé’ utilizadas na formação de numerais, como pode ser observado na tabela 7 e nas tabelas apresentadas na subseção 5.1.1, não são numerais. A agramaticalidade do dado em (74) mostra que esse tipo de expressão não ocorre como um numeral.

74)	<i>*yjpi</i>	<i>ota</i>	<i>yn</i>	<i>naokyt</i>	<i>boroja</i>
	<i>yj-pi</i>	<i>ota</i>	<i>yn</i>	<i>na-oky-t</i>	<i>boroja</i>
	1p-pé outro	eu	DECL-matar-NFUT	cobra	

A semântica composicional dos numerais que contém *yjpi ota* mostra que essa expressão vale 15. Se ela fosse um numeral, a sentença em (30) significaria ‘eu matei 15 cobras’, mas esse não é o caso. A falante do karitiana consultada não conseguiu traduzir a sentença e disse que ela estava errada. Entretanto, é bastante claro que essa expressão vale 15 porque se refere a três membros que valem 5 cada, as mãos e o outro pé.

Esta pesquisa prefere considerar que as expressões formadas por *yjpi* ‘nosso pé’, *yjpy* ‘nossa mão’ e *ota* ‘outro’ são bases auxiliares. Essas expressões são a forma como a língua organiza quantidades a partir da base 5. Assim, *yjpy ota* vale 5, *yjpi* vale 10 e *yjpi ot* vale 15. Essas expressões são somadas a um numeral básico em uma estrutura de coordenação projetada pela palavra *ot* ‘pegar’. O tipo semântico dessas expressões deve ser <e,t>, já que ocorrem em uma relação de coordenação com um elemento também de tipo <e,t>.

Assim, esta análise propõe que essas expressões são propriedades cardinais de tipo <e,t> com um valor cardinal fixo definido pelo número de membros implícitos na expressão multiplicado pela base 5 (ver tabela 7). Por isso, essas expressões serão chamadas de bases auxiliares, já que não contem uma operação explícita de multiplicação e por isso não se pode afirmar que envolvem multiplicadores lexicais. Contudo, elas operam como bases auxiliares na composição dos numerais já que representam valores fixos e são utilizadas de forma hierárquica, como prediz a estratégia de empacotamento.

A estrutura sintática e semântica do numeral *sypomp yjpy ota ot* ‘sete’ nessa nova proposta é apresentada em (75).

75)

Para Hurford (1987) uma base multiplicadora pertence à classe dos substantivos, isso faz sentido já que essas bases não são numerais. Seguindo o autor, esta análise propõe que de forma semelhante a base auxiliar é um nominal que denota uma propriedade cardinal, i.e, uma cardinalidade. Essas bases são uma estratégia mais eficiente de formar numerais do que somar a base *yjpy* repetidamente. Expressões como **sypomp yjpy yjpy yjpy ot* não são geradas pela língua, mas os valores 5, 10 e 15 são representados pelas bases auxiliares *yjpy ota*, *yjpi* e *yjpi ot* respectivamente. Assim, vai-se assumir que essas expressões projetam um sintagma nominal de tipo $\langle e, t \rangle$ que denota uma cardinalidade fixa e pode ser coordenado com um numeral. Dessa forma, a denotação de *yjpy ota*, por exemplo, é apresentada em (76).

$$76) [[yjpy\ ota_{\langle e, t \rangle}]] = \lambda x. |x| = 5$$

A semântica apresentada em (76) implica que essa expressão pode ser aplicada a qualquer entidade de tipo $\langle e \rangle$, mas esse não é caso. Essas expressões ocorrem somente em numerais compostos. Por isso, é preciso assumir uma restrição do uso dessas expressões. Vai-se assumir que as propriedades $\langle e, t \rangle$ denotadas pelas expressões *yjpy ota*, *yjpi* e *yjpi ot* são propriedades numéricas, como os demais cardinais da língua, especificadas para construção de numerais compostos. É a forma com que a estratégia de empacotamento opera nessa língua tornando a composição dos numerais mais eficiente. E mesmo que essa abordagem não explique completamente a semântica composicional envolvida, é um avanço na investigação de sistemas numerais não antes estudados já que captura de forma adequada a descrição do fenômeno.

Por fim, é apresentada a derivação da semântica composicional do numeral *sypomp yjpy ota ot* ‘seven’ proposta por esta pesquisa. A estrutura sintática desse numeral foi apresentada em (75). O numeral é formado pela aplicação argumentos de tipo $\langle e, t \rangle$ *sypomp* e *yjpy ota* ‘outra mão’ à denotação de *ot* que carrega a operação de adição via uma projeção de coordenação

$$77) [[sypomp yjpy ota ot_{\langle e, t \rangle}]] = ot(yjpy ota)(sypomp) \quad \text{aplicação funcional}$$

$$[[sypomp yjpy ota ot_{\langle e, t \rangle}]] = \lambda P \lambda Q. P+Q (\lambda x. |x|=5) (\lambda x. |x|=2)$$

$$[[sypomp yjpy ota ot_{\langle e, t \rangle}]] = \lambda x. |x|=7$$

Dessa forma, os numerais compostos regulares do karitiana, i.e., as séries que representam os números 6-9, 11-14 e 16-19, são formados por uma estrutura aditiva, e não multiplicativa como foi argumentado por Vignado (2017). Nessas estruturas, duas cardinalidades são somadas, a cardinalidade do numeral básico e a cardinalidade da base auxiliar.

Os numerais *yjpy ota tyytap* ‘dez’ e *yjpi ota tyytap* ‘vinte’ foram traduzidos por um informante nativo como ‘todas as mãos’ e ‘todos os pés’, respectivamente. O significado da palavra *tyytap*, que contém a palavra *tyy* que no dicionário de Landim (1984) significa ‘grande’. Essa estrutura ainda precisa ser mais bem explorada, mas, ainda assim, esses numerais apresentam as mesmas bases morfológicas encontradas nos demais numerais. É interessante notar que a tradução “todos” dada pelos falantes remete ao termo *akatyym*, formado pela cópula *aka* e *tyy* ‘grande’ traduzido como ‘todos’. Coutinho-Silva (2008) faz uma análise desse termo que se assemelha a uma quantificação universal. Contudo, esse autor defende que não se trata de quantificação universal, mas de uma sentença subordinada que produz esse efeito de sentido, o que também corrobora a ausência de quantificação nominal nessa língua.

Já o numeral *myhint yjpi py* ‘quinze’ parece ser uma irregularidade. Irregularidades são um fato comum em sistemas numéricos, como no Francês, por exemplo, que usa *soixante-dix* literalmente ‘sessenta-dez’ para expressar setenta, mas usa *quatre-vingt* ‘quatro vintes’ para expressar oitenta. No caso do karitiana, *myhint yjpi py* significa ‘um pé e mãos’, o que sugere que a composição aritmética subjacente é 5×3 .

Esta análise mostrou que diferentemente do que afirmou Vignado (2017), o karitiana não apresenta estruturas multiplicativas, mas representa seus numerais compostos via estruturas aditivas utilizando numerais básicos e bases auxiliares. Assim, a palavra *ota* inferir quantos são os membros envolvidos para definir o valor da base auxiliar. As bases auxiliares constituem sintagmas nominais específicos para a formação de numerais que tem um valor cardinal. Esta análise mostrou que o sistema numeral do karitiana não tem estruturas complexas, i.e., aquelas que envolvem adição e multiplicação.

Nesta seção foi apresentada a análise que defende que a estratégia de empacotamento opera em karitiana através da lexicalização de bases auxiliares, o que reforça a universalidade desse mecanismo como proposto por Hurford (1987). Isso fornece evidências para a hipótese de que a habilidade de expressar cardinalidades está relacionada à faculdade da linguagem.

Esta subseção apresentou a análise da formação de numerais em karitiana proposta por esta pesquisa. A próxima subseção sumariza os resultados obtidos.

5.1.3. Resultados

Esta análise mostrou que a língua karitiana tem apenas estruturas aditivas na composição de seus numerais. A presença de bases auxiliares, contudo, evidencia a ação da estratégia de empacotamento na formação desses numerais. Seria possível que o numeral *sypomp yjpi ota ot* ‘dezessete’ fosse construído somando a base *yjpy*, mas a forma *sypomp yjpy yjpy yjpy ot* não existe. Isso mostra que a língua lexicalizou bases auxiliares que representam os valores 5, 10 e 15, tornando a composição dos numerais mais eficiente. Optou-se pelo termo bases auxiliares pois essas expressões não envolvem multiplicadores lexicais, tem um valor cardinal fixo e só ocorrem no sistema numeral.

A consideração da semântica composicional proposta por Rothstein (2013, 2017) ajudou a elucidar a natureza das expressões como *yjpy ota*. Assumir que *ota* tem um valor cardinal que varia de acordo com o numeral em que está contido era um problema na análise prévia apresentada por Vignado (2017). Se essa expressão funcionasse de fato como um NUMBER, ela deveria variar em seu valor cardinal. Diante disso, esta pesquisa

optou por classificar essas expressões como bases auxiliares, já que claramente elas carregam um valor cardinal.

A base do sistema é 5, por isso, espera-se que esse sistema tenha multiplicadores lexicais para as potências 5^1 , 5^2 , 5^3 e assim por diante. Entretanto, esse é um sistema restrito que vai somente até vinte. A hipótese levantada é que em um estágio anterior, o karitiana contava apenas com os numerais cardinais *myhint* ‘um’, *sypomp* ‘dois’ e *mỹjymp* ‘três’. Mesmo que os numerais *otadnamynt* ‘quatro’ e *yjpyt* ‘cinco’ sejam numerais básicos no estágio sincrônico considerado nesta análise, é bastante claro que esses numerais são derivados de outras palavras, portanto, eles não são cardinais reais, mas sim, aquisições recentes. É muito comum que línguas tupi apresentem um conjunto pequeno de numerais cardinais.

Como apontado por Winter (1999), a expansão é o sentido natural no desenvolvimento dos sistemas numerais nas línguas naturais. Desse modo, esta análise propõe via contato com outras línguas e outras razões extralinguísticas, o sistema numeral do karitiana desenvolveu numerais básicos até 5, para números maiores, os numerais são compostos a partir dos numerais básicos, das bases auxiliares e da palavra *ot* que projeta uma coordenação que expressa a operação de adição. O esquema geral dos numerais regulares do karitiana é apresentado (78).

78) Numeral: [Numeral [numeral.básico] [COORD [base.auxiliar] ot]]

Esta seção apresentou os resultados da análise do karitiana. Essa análise discorda de Vignado (2017) que sugere que o karitiana tem multiplicadores lexicais e estruturas multiplicativas. Um olhar mais cuidadoso sobre a semântica composicional dessas expressões mostrou que esse não é caso, mas que a língua expressa seus numerais apenas via estruturas aditivas. Isso sugere que no desenvolvimento de sistemas numerais nas línguas naturais primeiro a língua desenvolve estratégias para compor numerais via adição

e então surgem as estruturas multiplicativas. A seção a seguir apresenta informações sobre o kamaiurá, os numerais e uma análise nos moldes da interface sintaxe e semântica proposta. A análise do kamayurá foi motivada pela análise estabelecida para o karitiana.

5.2. O kamayurá

Esta seção conta com as informações básicas sobre a língua kamayurá, a análise do seu sistema numeral e os resultados dessa análise. Assim, a subseção 5.2.1 apresenta a língua kamayurá e as tabelas dos dados de seu sistema numeral. Já a seção 5.2.2 apresenta a análise proposta por esta pesquisa e, por fim, a seção 5.2.3 sumariza os resultados da análise.

A próxima seção apresenta o kamayurá e seus dados numerais, primeiramente apresenta a língua e os seus numerais básicos. A seguir, a seção apresenta os significados dos morfemas envolvidos na formação de numerais compostos, baseados em Seki (2000) e as tabelas com os numerais compostos na língua.

5.2.1. Os numerais kamayurá

Nesta seção são apresentadas informações sobre a língua kamayurá, a descrição de seu sistema numeral na literatura e uma tabela que apresenta seus numerais e a decomposição morfológica que esta análise propõe. Essa língua foi amplamente descrita pela Gramática do kamayurá escrita por Seki (2000), todos os dados apresentados aqui foram retirados dessa gramática.

O Kamayurá é uma língua da grande, e dispersa geograficamente, subfamília Tupi-guarani. Línguas da mesma família são encontradas no Peru, Bolívia, Venezuela, Colômbia, Guiana Francesa, Paraguai, norte da Argentina e Brasil (RODRIGUES, 1986). A fonte dos dados de numerais será a gramática da língua Kamayurá, feita por Seki (2000), e que foi uma das primeiras descrições completas de uma língua indígena publicada por uma pesquisadora brasileira (MOORE e GABAS JUNIOR, 2006).

Essa língua ainda é falada e atualmente esta etnia é composta por cerca de 600 indivíduos (siasi/sesai, 2014) que habitam o parque do Xingu.

Ao descrever os numerais do kamayurá, Seki destaca as palavras e morfemas que compõem a formação dos numerais na língua. A tabela 9, abaixo, mostra essas palavras e morfemas que motivam a decomposição morfológica dos numerais proposta por esta análise.

Tabela 9 – Palavras e morfemas que compõe numerais kamayurá	
palavra/morfema	Significado
<i>jene</i>	1ª pessoa inclusiva
<i>po</i>	‘mão’
<i>py</i>	‘pé’
<i>pyaĩ</i>	‘dedo do pé’
<i>ero'yahap</i>	‘passar por, passar além de, atravessar’
{-pap}	indica que algo terminou, ‘terminar’
{mo-}	Causativo

Todos os numerais básicos dessa língua temo morfema causativo {-mo}. Por exemplo, o numeral *mokōj* ‘dois’ é formado pelo causativo e pela palavra *kōj* ‘par’, essa expressão significa literalmente ‘fazer par’ e veicula a cardinalidade 2. O significado literal dos numerais básicos pode ser parafraseado como ‘fazer n’, sendo n o valor cardinal que a expressão veicula. Mesmo assim, essas expressões operam como numerais na língua, veiculando o valor cardinal. Vai-se assumir, seguindo Rothstein (2013, 2017) que os cardinais denotam propriedades do tipo $\langle e, t \rangle$ que definem a cardinalidade de x, i.e., para qualquer numeral n sua denotação é $\lambda x. |x| = n$.

A tabela 10 apresenta os numerais básicos do kamayurá e sua decomposição morfológica de acordo com Seki (2000).

Tabela 10 – Numerais básicos em Kamayurá	
Numeral	Decomposição morfológica
<i>mojepete</i> 'um'	mo-jepe-te CAUS-um-ENF
<i>mokõj</i> 'dois'	mo-kõj CAUS-gemeos
<i>mo'apyt</i> 'três'	mo-'apyt CAUS-três
<i>mojo'irũ</i> 'quatro'	mo-jo'-irũ CAUS-?-amigo
<i>jenepomomap</i> 'cinco'	jene-po-mo-map 1p-mão-CAUS-terminar

O numeral *mojepete* também apresenta o sufixo enfático {-te}, o significado literal dessa expressão é 'fazer um mesmo'. Já a morfologia do numeral *mojo'irũ* não é tão clara quanto a dos outros numerais, mas, assim como o karitiana, esse numeral é composto a partir de uma palavra que sugere um outro elemento, ou, no caso do kamayurá, o amigo. Isso pode ser evidência de que, assim como no karitiana, os numerais para 4 e 5 sejam aquisições mais recentes do que os cardinais de 1-3.

O numeral *jenepomomap* 'cinco' também apresenta uma morfologia interessante, no caso desse numeral além o morfema causativo há a presença do morfema {-map}. Seki (2000) diz que esse morfema carrega o significado de 'acabar' ou 'terminar', assim, o significado literal dessa expressão seria 'fazer mão acabar', que representa o valor cardinal 5. Como o karitiana, o kamayurá também apresenta 5 numerais básicos.

O sistema numeral do kamayurá também conta com numerais compostos até 20, que serão apresentados nas tabelas a seguir. A tabela 11 apresenta os numerais de 6 a 11 nessa língua.

Tabela 11– Numerais de 6 a 10 em Kamayurá	
Numeral	Decomposição morfológica
<i>jenepoa wero'yahap mojepete</i> 'seis'	jene-poa wero'yahap mojepete 1p-mão passar.por um
<i>jenepoa wero'yahap mokõj</i> 'sete'	jene-poa wero'yahap mokõj 1p-mão passar.por dois
<i>jenepoa wero'yahap mo'apyt</i> 'oito'	jene-poa wero'yahap mo'apyt 1p-mão passar.por três
<i>jenepoa wero'yahap mojo'irũ</i> 'nove'	jene-poa wero'yahap mojo'irũ 1p-mão passar.por quatro
<i>jenepopap</i> 'dez'	jene-po-pap 1p-mão-terminar

Nessa série de numerais, os numerais de 6 a 9 são formados a partir de um numeral básico e da estrutura *jenepoa wero'yahap* cujo significado literal é 'passe para a nossa mão'. Como vai-se argumentar na próxima subseção, esta análise propões que a palavra *wero'yahap* 'passar por' carrega a semântica da adição projetando uma estrutura que coordena o numeral básico a expressão que se refere a uma parte do corpo e veicula uma cardinalidade específica.

É notável a semelhança entre o numeral *jenepomomap* 'cinco' e *jenepopap* 'dez'. Em (79) é exposta a decomposição morfológica dessas expressões com fins de comparação.

79) a. jene-po-mo-map
1p-mão-CAUS-terminar
'cinco'

b. jene-po-pap
1p-mão-terminar
'dez'

A morfologia desses numerais mostra que (79a) contém o morfema causativo e (79b) não. Os numerais básicos do kamayurá são todos formados com o morfema causativo, os demais não. Assim, a expressão em (79b) significa literalmente ‘as mãos acabaram’, portanto, são duas mãos que somam 10. Já a expressão em (79a) significa ‘fazer mão, acabou’, porque uma mão vale 5.

A tabela 12 apresenta a série de numerais do 11 ao 15.

Tabela 12 – Numerais de 11 a 15 em Kamayurá	
Numeral	Decomposição morfológica
<i>mojepete jene pyaĩ wero'yahap</i> 'onze'	mojepete jene pyaĩ wero'yahap um 1p.inclusiva dedo.do.pé passar.por
<i>mokõj jene pyaĩ wero'yahap</i> 'doze'	mokõj jene pyaĩ wero'yahap dois 1p.inclusiva dedo.do.pé passar.por
<i>mo'apyt jene pyaĩ wero'yahap</i> 'treze'	mo'apyt jene pyaĩ wero'yahap três 1p.inclusiva dedo.do.pé passar.por
<i>mojo'irũ jene pyaĩ wero'yahap</i> 'catorze'	mojo'irũ jene pyaĩ wero'yahap três 1p.inclusiva dedo.do.pé passar.por
<i>mojepete jene pya opap</i> 'quinze'	mojepete jene pya o-pap um 1p.inclusiva pé vt-terminar

Nessa série de numerais, ao invés de *jenepoa* ‘nossa mão’, o sistema numeral utiliza a expressão *jene pyaĩ* ‘nosso dedo do pé’. Mais especificamente, *pya* significa ‘pé’ e *ĩ* é o sufixo diminutivo, assim, literalmente, *pyaĩ* significa ‘pézinho’, mas Seki (2000) traduz a expressão como ‘dedo do pé’, e esta pesquisa vai seguir a autora. Já o numeral *mojepete jene pya opap* ‘quinze’ utiliza a expressão *jene pya* ‘nosso pé’, esse numeral parece conter uma estrutura multiplicativa *mojepete jene pya* como será exposto em mais detalhes na próxima subseção.

Por fim, a tabela 13 apresenta a os numerais de 16 a 20 e sua decomposição morfológica a partir da descrição fornecida por Seki (2000).

Tabela 13 – Numerais de 16 a 20 em Kamayurá	
Numeral	Decomposição morfológica
<i>jene pya wero 'yahap mojepete</i> 'dezesseis'	jene pya wero 'yahap mojepete 1p.inclusiva pé passar.por um
<i>jene pya wero 'yahap mokõj</i> 'dezessete'	jene pya wero 'yahap mokõj 1p.inclusiva pé passar.por dois
<i>jene pya wero 'yahap mo 'apyt</i> 'dezoito'	jene pya wero 'yahap mo 'apyt 1p.inclusiva pé passar.por três
<i>jene pya wero 'yahap mojo 'irũ</i> 'dezenove'	jene pya wero 'yahap mojo 'irũ 1p.inclusiva pé passar.por quatro
<i>Jene pya omomap mokõj</i> 'vinte'	jene pya o-mo-map mokõj 1p.inclusiva pé VT-CAUS-terminar dois

Nesses numerais, é utilizada a expressão *jenepya* 'nosso pé'. O pronome pessoal *jene* tem a mesma forma em sua versão como pronome livre e como pronome clítico. Em sua gramática, Seki (2000) apresenta os dados de numerais com *jene* separado de *poa*, *pyaĩ* e *pya* sugerindo ser um pronome livre, contudo, a autora não afirma esse status para o pronome nessa construção.

Os dados mostraram que, assim como o karitiana, a composição dos numerais exhibe padrões de 5 em 5. Isso sugere que esse sistema é de base 5. Em kamayurá, os numerais também são compostos a partir de partes do corpo. O fato de que os numerais *mojo 'irũ* e *jenepomomap* são derivados de outras palavras sugere que esses numerais são aquisições mais recentes em relação aos numerais de 1-3. É comum em línguas tupi que o numeral que vale 5 seja composto a partir da palavra mão, como, por exemplo, as expressões *iepepu* 'uma mão' em nheengatu e *xe po* 'minha mão' em tupi antigo.

Esta seção apresentou brevemente a língua kamayurá. A seção contou com a apresentação dos dados de numerais do kamayurá e sua decomposição morfológica baseada na descrição de Seki (2000). A próxima seção traz a análise desses dados proposta por esta pesquisa.

5.2.2. Análise da formação de numerais

Esta seção traz a análise dos dados do kamayurá a partir do panorama teórico proposto, assim apresenta tabelas sobre a recursividade na formação dos numerais compostos dessa língua e as estruturas sintáticas e semânticas propostas por esta pesquisa. A análise mostra que a estratégia de empacotamento opera na formação do sistema numeral da língua, o que reforça a universalidade desses mecanismos e a vantagem de uma análise baseada na interface.

Os dados do kamayurá mostram uma regularidade no uso das expressões *jene poa*, *jene pyaĩ* e *jene pya*. Assim como na análise do karitiana apresentada na seção 5.1, a tabela 14 abaixo apresenta o paradigma dos numerais *mokōj* ‘dois’, *jenepoa wero’yaha mokōj* ‘sete’, *jenepyaĩ mokōj wero’yahap* ‘doze’ e *jenepya wero’yahap mokōj* ‘dezessete’ para ilustrar a regularidade na formação desses numerais. Na primeira linha estão os numerais, na segunda linha, a referência da expressão, e, na terceira, a lógica matemática subjacente à expressão numérica.

Tabela 14 – Paradigma da formação dos numerais em kamayurá				
Numeral	<i>mokōj</i> two	<i>jene-poa wero’yaha</i> 1p-mão passar.por <i>mo-kōj</i> fazer-gêmeos	<i>mokōj jene-pyaĩ</i> dois 1p-dedo <i>wero’yahap</i> passar.por	<i>jene-pya wero’yahap</i> 1p-pé passar.por <i>mokōj</i> (17) dois
Referência
Matemática	2	2+(1x5)	2+(2x5)	2+(3x5)

Como no caso do karitiana, os números nesse paradigma aumentam de cinco em cinco no nível matemático, enquanto a expressão linguística mostra que o número é indicado pelo membro do corpo a que se refere. Nesse conjunto de dados, pode-se observar que as expressões *jenepoa*, *jenepyaĩ* e *jenepya* são utilizadas para formar numerais compostos nessa língua. Essas expressões se distribuem de forma fixa e hierárquica como foi apresentado na tabela.

Em karitiana, *ot*, raiz do verbo pegar, funciona como uma conjunção na estrutura do numeral que expressa a operação de adição. No kamayurá, a expressão *wero'yahap* 'passar por' aparece sistematicamente na formação dos numerais, esta análise propõe que essa palavra marca a operação de adição na formação desses numerais, a estrutura proposta será apresentada a seguir, o que está de acordo com Hurford (1987) que aponta que a estratégia de empacotamento depende do léxico disponível na língua.

Diferente dos numerais karitiana, os numerais kamayura apresentam uma irregularidade nos numerais de 11-14, nessa série de numerais a ordem é diferente. Esse fato é exemplificado pelo contraste entre, *jenepoa wero'yahap mokōj* 'sete' e *mokōj jenepyaĩ wero'yahap* 'doze'. Na estrutura de ambos numerais proposta por esta análise, em (80a) e (80b), pode-se observar que o numeral básico *mokōj* está sendo somado à estrutura e é irmão de COORD. Esse fato está de acordo com a observação de Hurford (1987) de que em numerais a estrutura hierárquica dos constituintes é mais importante que sua ordem linear. Assim, apesar da ordem diferente em (80a) e (80b), a hierarquia dos constituintes é a mesma.

80) a.

b.

A partir da observação da formação dos numerais compostos regulares em kamayurá, 6-9, 11-14 e 16-19, é bastante evidente que as expressões *jene poa*, *jene pyai* e *jene pya* possuem valores cardinais fixos definido a partir da base 5. Essas expressões, contudo, não estão em uma relação de multiplicação com um numeral básico, então, assim como foi estabelecido para o karitiana, esta análise propõe que essas expressões são bases auxiliares que manifestam a estratégia de empacotamento no sistema.

Assim, esta análise propõe na tabela 15 as bases auxiliares do kamayurá. É preciso recordar o conceito de base numérica proposto por Comrie (2005) em que a base numérica de um sistema numérico se dá por uma base que recursivamente é multiplicada por um

número *n* qualquer. Essa definição de base refere-se ao conceito de sistema numérico no sentido matemático, e ajuda a elucidar a recursividade morfológica do sistema numeral.

Tabela 15 – Bases morfêmicas do kamayurá

Base auxiliar	Tradução	Valor	base x n
<i>jene poa</i>	‘nossa mão’	5	5x1
<i>jene pyaĩ</i>	‘nosso dedo do pé’	10	5x2
<i>jene pya</i>	‘nosso pé’	15	5x3

O sistema numeral do kamayurá é construído a partir da base 5, assim como outras línguas Tupi como o karitiana e o nheengatu.³⁰ As bases auxiliares apresentadas na tabela 15 também são utilizadas na formação dos numerais *jenepopap* ‘dez’, *mojepete jenepya opap* ‘quinze’ e *jenepya omomap mokōj* ‘vinte’, o que reforça seu status como uma base auxiliar. A composição desses numerais, contudo, não é tão regular quanto os outros a representa problemas pra essa análise.

Como no caso do karitiana, esta análise vai considerar as bases auxiliares do kamayurá como nominais na estrutura sintática. Assim, os exemplos apresentados em (81a) e (81b) ilustram a estrutura sintática do numeral *jenepoa wero'yahap mokōj* nessa perspectiva.

81) a.

b.

³⁰ Análises sobre a base numérica dessas línguas foram propostas pela autora desta pesquisa no decorrer de uma iniciação científica realizada em 2012-2013. Foi apresentada no I Seminário “O tupi e as línguas gerais na formação da civilização brasileira”, realizado na FFLCH-USP em junho de 2013.

O exemplo em (82) abaixo apresenta a estrutura do numeral *mojepete jenepya opap* ‘quinze’ propostas por essa análise.

82)

Na estrutura em (82) parece haver uma relação de multiplicação entre *myhint* e *jenepya*. Como *jenepya* vale 15 ele é multiplicado por *mojepete* ‘um’ e resulta na cardinalidade total da expressão que também é 15.

O numeral *jenepya omomap mokōj* ‘vinte’ é um problema para esta análise. Se de fato a base auxiliar *jenepya* valer 15, a composição desse numeral não dá certo. Pois dois *jenepya* duas vezes deveriam somar 30, e não 20. Além disso, em *omomap* há um morfema causativo {-mo}. A interpretação sugerida pela morfologia dessa expressão, apresentada em (83), é ‘fazer acabar o pé duas vezes’.

83) Jenepya o-mo-map mokōj
 1p.pé VT-CAUS-terminar dois

Talvez, seja o caso de que as mãos já foram contadas e cada *jene pya* ‘nosso pé’ vale 5 nessa estrutura. Contudo, a estrutura composicional não é clara a partir do panorama teórico adotado por esta pesquisa. Uma base auxiliar não deveria estar associada a dois valores cardinais diferentes. Ainda assim, esta análise julga adequado considerar que *jene pya* é uma base auxiliar que vale 15 pois ela ocorre em mais estruturas, mas chama atenção pra essa irregularidade.

Como foi visto, o sistema numeral do kamayurá vai somente até 20. Esta pesquisa reforça a necessidade de se investigar mais os numerais de línguas cujos sistemas numerais sejam mais restritos se comparados a sistemas numerais mais bem descritos, como os das línguas indo-europeias, a fim de compreender melhor de maneira esses sistemas se expandem e como isso reflete a faculdade da linguagem.

No caso dos numerais *jenepomomap* ‘cinco’, *jenepopap* ‘dez’, *mojepete jenepya opap* ‘quinze’ e *jenepya omomap mokōj* ‘vinte’, há também a presença do morfema {-pap}, que foi descrito por Seki (2000) com o significado de ‘acabar, terminar’. É interessante notar que em karitiana foi observado nesses mesmos numerais um morfema que parece ser cognato {-tap} por sua fonologia e por seu significado, como exemplificado com os numerais de ambas línguas na tabela 16 abaixo.

Tabela 16 – Acabar, karitiana vs kamayurá				
	Karitiana		Kamayurá	
‘dez’	yjpy	ota ty- tap	jene-po- pap	
	1p-mão	outro grande-terminar	1p-mão- terminar	
‘vinte’	yj-pi	ota ty- tap	jenepya	o-mo- map mokōj
	1p-pé	outra grande-terminar	1p-mão	VT-CAUS- terminar dois

Considerando que karitiana e kamayurá são línguas aparentadas, mas distantes na genealogia da família, é esperado apresentam sistemas numerais semelhantes, i.e., sistemas numéricos de base 5. Os numerais básicos de 1-3, aqueles que parecem ser cardinais genuínos, entretanto, são diferentes. Por fim, a comparação entre essas línguas pode ajudar na elucidação do significado da construção *tyytap* que ocorre em alguns numerais karitiana. Esta análise sugere que essa expressão signifique ‘acabar todas’, o que está de acordo com o significado dos numerais em que ocorre.

Esta subseção apresentou a análise proposta para os numerais do kamayurá. Mesmo que essa análise ainda apresente falhas, configura uma primeira tentativa de explicar formalmente os numerais dessa língua. A próxima subseção sumariza os resultados desta análise.

5.2.3 Resultados

A exposição da sintaxe e da semântica composicional dos numerais do kamayurá na subseção anterior mostrou que, com certeza, essa língua expressa seus numerais compostos via estruturas aditivas. Alguns numerais sugerem que o sistema tem estratégias de multiplicação como foi mostrado pelo caso do numeral *mojepete jenepya opap* ‘quinze’, entretanto, não é claro como isso funciona a partir da semântica das partes envolvidas.

A presença de bases auxiliares, contudo, evidencia a ação da estratégia de empacotamento na formação desses numerais, já que é uma maneira de tornar a composição dos numerais mais eficiente. Assim como no karitiana, lexicalizar os valores 5, 10 e 15 a maneira como o sistema numeral compõe números maiores. Esse sistema apresenta numerais básicos até cinco e numerais com maior valor são compostos a partir dos numerais básicos, das bases auxiliares e da palavra *wero'yahap* que veicula a semântica da operação de adição.

A comparação entre alguns numerais do karitiana e do kamayurá sugere que o morfema {-tap} do karitiana e o morfema {-pap} do kamayurá são cognatos. Já os numerais básicos de 1-3, os cardinais genuínos, são diferentes, o que sugere uma origem diferente. Isso se justifica pela distância entre essas línguas na genealogia da família Tupi.

A estrutura sintática geral dos numerais regulares de 6-9 e 16-19 em kamayura é apresentado (84).

84) Numeral: [Numeral [COORD [base.auxiliar] [wero'yahap]] [numeral.básico]]

Já a série de numerais de 11-14 apresenta outra ordem, então, sua estrutura sintática é apresentada em (85)

85) [Numeral [numeral.básico][COORD [base.auxiliar] [wero'yahap]]]

Nesta seção foi apresentada a análise que defende que a estratégia de empacotamento opera na formação de numerais em kamayurá através da lexicalização de bases auxiliares que tornam o sistema numeral mais eficiente. Essa análise juntamente com a análise do karitiana reforça a universalidade desse mecanismo como proposto por Hurford (1987) e mostra que a faculdade numérica está relacionada à faculdade da linguagem. A seção a seguir apresenta as considerações finais desta pesquisa.

6. Considerações finais

As análises do karitiana e do kamayurá evidenciaram que os sistemas numerais dessas línguas não são ótimos como sistemas numerais mais desenvolvidos como o do português e do inglês. Esses sistemas não são plenamente recursivos, pois além de expressarem numerais somente até 20, não apresentam multiplicadores lexicais, como *hundred*, por exemplo. Mesmo assim, ambas línguas têm sistemas numerais que contêm numerais compostos que exibem um padrão morfossintático em sua composição. Nessas línguas as partes do corpo são fundamentais para cardinalidade do numeral, por isso é preciso considerar a cardinalidade que cada parte do corpo representa para fornecer uma análise que contemple tanto o aspecto linguístico quanto o aspecto matemático das expressões numerais. Predomina na literatura a descrição de línguas indo-europeias faladas em contextos de industrialização. Sendo assim, em geral, sabe-se pouco sobre sistemas de numerais e por isso é importante analisar os sistemas numerais das línguas brasileiras.

Esta análise apresentou o sistema numeral de ambas línguas, com seus numerais básicos e seus numerais compostos. Os dois sistemas apresentam numerais básicos até 5 e são sistemas de base 5. A análise foi baseada na sintaxe e na semântica das estruturas numéricas de duas línguas indígenas brasileiras a partir do princípio da estratégia de empacotamento de Hurford (1987), que busca fornecer regras que gerem economicamente e com adequação semântica todas as expressões numéricas bem formadas de uma língua, e da proposta de tipos semânticos para uma semântica composicional de numerais elaborada por Rothstein (2013, 2017).

Mesmo que os dados não correspondam com exatidão ao modelo semântico aplicado na análise, esta pesquisa conclui que confrontar a teorias com os dados dessas línguas foi bastante produtivo para elucidar a lógica de sua composição. Problemas na análise de Vignado (2017) motivaram o uso do conceito de base auxiliar para explicar melhor os dados. Já que se observou que essas não apresentam multiplicadores lexicais como *hundred* ou *thousand* que se compõe recursivamente com os numerais básicos da língua. Esta análise assume que as bases auxiliares nessas línguas são nominais que denotam propriedades cardinais de tipo $\langle e, t \rangle$ específicas para a formação do sistema numeral.

Sobre o karitiana, a análise dos dados mostrou que o sistema numeral karitiana é de base 5 e que os numerais compostos são formados via numerais básicos, bases

auxiliares e coordenação, como era previsto pelo panorama teórico. Foram identificadas três bases auxiliares que se distribuem de forma fixa e hierárquica *yjpy ota*, *yjpi* e *yjpi ota* que valem 5, 10 e 15 respectivamente. Assim como outras línguas naturais, o karitiana apresenta uma marca morfológica nas estruturas interpretadas como adição. Nessa língua, a raiz do verbo “pegar” *ot* projeta uma estrutura coordenativa que carrega a semântica da operação de adição. Esses fatos linguísticos seguem a tendência dos sistemas numéricos das línguas naturais, como foi apresentado previamente.

Diferentemente de Vignado (2017), esta análise mostrou que o sistema numeral do karitiana apresenta bases auxiliares, e não multiplicadores lexicais no sentido de Hurford (1987) e de Rothstein (2017). Em síntese, esta análise mostrou que: (i) o sistema numérico karitiana é de base 5; (ii) a estratégia de empacotamento opera nos numerais karitiana através da lexicalização de bases auxiliares com valores cardinais fixos; (iii) o verbo *ot* funciona como uma conjunção coordenativa na formação de numerais.

Sobre o kamayurá, a análise de dados também mostrou que o sistema numeral kamayurá é de base 5 e que os numerais compostos são formados via os numerais básicos, as bases auxiliares e coordenação, como era previsto pela teoria. Até o momento foram identificadas três bases morfêmicas multiplicativas *jenepoa*, *jenepyaĩ* e *jenepya*. O kamayurá apresenta uma marca morfológica nas estruturas interpretadas como adição, seguindo a tendência de outras línguas naturais. Nessa língua, o verbo *wero 'yahap* ‘passar por’ funciona como uma conjunção coordenativa. No caso do kamayurá, parece haver uma estratégia de expressar multiplicação. Essa análise, entretanto, não logrou explicar esse fenômeno completamente. O numeral *jenepya omomap mokōj* ‘vinte’ é uma irregularidade que representa um desafio a descrição e análise das estruturas multiplicativas na língua.

Em síntese, esta análise mostrou que: (i) o sistema numérico kamayurá é de base 5; (ii) a estratégia de empacotamento opera nos numerais kamayurá através das bases auxiliares; (iii) *wero 'yahap* opera como uma conjunção coordenativa expressando a operação de adição.

Os fatos linguísticos apresentados mostram que é possível analisar numerais de línguas indígenas formalmente. No karitiana e no kamayurá, observa-se que as bases auxiliares são representadas por partes do corpo, o que poderia sugerir que esses sistemas não sejam

numéricos, ou seja, que não expressem uma lógica numérica subjacente. A análise contudo, mostrou que de fato os sistemas dessas línguas se organizam sobre a base 5.

Contrariamente, esta análise se propõe a não ignorar os padrões linguísticos dos dados. Por isso, ressaltando a importância de se considerar a linguística e a matemática dos numerais, esta pesquisa buscou entender o sistema numérico subjacente aos sistemas numerais indígenas auxiliando a esclarecer assim sua recursividade. Como pretendeu-se mostrar nas análises, ainda que os sistemas numerais utilizem partes do corpo como bases, essas bases apresentam valores cardinais que se organizam de forma hierárquica, como se espera de um sistema numeral.

Ainda que as análises apresentadas para os sistemas numerais do karitiana e do kamayurá não sejam ideais e apresentem falhas, esta pesquisa defende que a abordagem proposta constitui um passo inicial importante para a descrição e análise de sistemas numerais de línguas indígenas e de línguas subdescritas. Não há na literatura linguística outra tentativa de explicar a estrutura interna de numerais de línguas tupi a partir da interface entre semântica e sintaxe. Não há também trabalhos que busquem explicar a composição dos numerais dessas línguas dentro da perspectiva da linguística formal. E justamente nessa lacuna que reside a relevância da investigação empreendida por esta pesquisa.

Por fim, a análise dessas línguas contribui para corroborar a hipótese da universalidade da estratégia de empacotamento a partir de dados de línguas não indo-europeias. Analisar línguas tipologicamente subdescritas suporta a verificação de teorias que propõem universais linguísticos. Os sistemas numéricos são tecnologias cognitivas geradas por nossa faculdade numérica, que, segundo Chomsky (1998), é um subproduto da arquitetura da linguagem natural. Assim, investigar numerais de línguas não indo-europeias de uma perspectiva formal e gerativa é potencialmente esclarecedor, e ajuda a avançar a compreensão da origem e desenvolvimento de sistemas numerais nas línguas naturais, e da relação da faculdade da linguagem e da faculdade numérica.

Chomsky (2005) sugere que existem restrições linguísticas que não são inatas, mas que são criadas via um desenvolvimento histórico e social que culmina em regras passadas via aquisição. Para o autor, essas restrições formam um conjunto que ele chama de terceiro fator da linguagem. De acordo com a argumentação de Hurford (2006) sobre o

desenvolvimento gradual dos sistemas numerais, esta pesquisa sugere que a estratégia de empacotamento é esse tipo de restrição a que Chomsky se refere. Contudo, reconhece que ainda é preciso investigar mais a adequação teórica dessa afirmação.

Por fim, esta pesquisa propõe que os sistemas numerais do karitiana do kamayurá são um retrato sincrônico de um sistema numeral em desenvolvimento, já que tem uma extensão limitada em relação a outros sistemas maiores como o do português e o do inglês e não apresenta multiplicadores lexicais e estruturas multiplicativas plenamente eficientes. Isso, de certo modo, justifica as irregularidades e imperfeições desses sistemas. Por isso, optou-se por assumir que ambos sistemas contam com bases auxiliares, mas não com multiplicadores lexicais reais como *hundred*, por exemplo.

Essas bases auxiliares evidenciam um processo de expansão do sistema sobre uma lógica quinária que não alcançou lexicalizar multiplicadores. Isso sugere que na expansão de um sistema numeral primeiro surjam as estruturas aditivas e depois as estruturas multiplicativas. Esta pesquisa ressalta que ainda que esses sistemas numerais *não* sejam plenamente recursivos, eles têm padrões morfossintáticos bastante claros que fornecem evidências de como os sistemas numerais se desenvolvem nas línguas naturais.

É importante ressaltar o que se considera universal sobre numerais. Concluiu-se que, assim como propõe Hurford (1987), a estratégia de empacotamento não é uma restrição inata à gramática universal, ela é uma restrição que vem do uso motivada pela eficiência do sistema numeral. Dessa forma, o componente inato relacionado aos sistemas numerais é a sintaxe envolvida na construção de estruturas que expressem uma semântica que carrega operações matemática, e não a estratégia de empacotamento em si.

É predominante na literatura línguas que apresentam as operações de adição e multiplicação, sendo que a adição parece ser a operação mais básica das duas. Ainda assim, não é claro se esse fato se justifica pela natureza das operações aritméticas em questão ou se reflete apenas o fato de que há mais descrições e análises de sistemas numerais de línguas indo-europeias faladas em contextos industrializados. Então, de modo geral, não se entende muito sobre sistemas numerais, já que apenas uma parcela das línguas naturais teve esse tema explorado em mais detalhes.

Por fim, análise defendeu que numerais compostos são formados na sintaxe a partir de partes menores, a análise considera importante se investigar a sintaxe envolvida do

ponto de vista do programa minimalista, mapeando os traços envolvidos na derivação dessas expressões. Ainda que este trabalho não tenha alcançado essa implementação formal, reconhece a importância dessa abordagem para investigações futuras.

7. Referências bibliográficas

ALFORD, M. 1987. Developing Facilitative Reading Programmes in Third World Countries, *Journal of Multilingual and Multicultural Development*, Vol. 8(6):493-511.

ANDERSON, L. 1958. Vocabulario breve del idioma Ticuna, *Revista Peruana de Cultura*. Cuzco, Peru, Vol.8(21):53-68.

AMAECHI, Mary Chimaobi. (2014). “The Syntax of Po Tangle Numerals”. *Journal of Universal Language* 15-2 p. 35–54.

AUSTIN, P. 1981 : A Grammar of Diyari, South Australia. University Press, Cambridge.

BARWISE, J. and R. COOPER: 1981, ‘Generalized Quantifiers and Natural Language’, *Linguistics & Philosophy* 4, 159–219.

BOOIJ, G. 2009. Constructions and Lexical Units: An Analysis of Dutch Numerals. *Linguistische Bericht*, 19, 1-14.

BOSWOOD, J. 1978. Quer falar a língua dos Canoeiros? Brasília, Summer Institute of Linguistics. 108 p.

CHIERCHIA, G.: 1985, ‘Formal Semantics and the Grammar of Predication’, *Linguistic Inquiry* 16, 417–443.

COMRIE, Bernard (2005). Numeral Bases. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology

CORVER, N. ; DOETJES, J.; ZWARTS, J (2007). Linguistic perspectives on numerical expressions: Introduction. *Lingua*, vol. 117, iss. 5, pp. 751-757.

CHOMSKY, N. Perspectivas sobre a linguagem e a mente. Sobre natureza e linguagem. Trad. de Marylene Pinto Michael. São Paulo: Martins Fontes, 2006.

_____. (1998) *Language and Problems of Knowledge*. MIT Press, Cambridge, Massachusetts.

_____. *Syntactic Structures*. The Hague: Mouton, 1957.

- EPPS, P. ; SALANOVA, A. P (2012). “A Linguística Amazônica Hoje”. *Liames* 12: 07–37.
- EVERETT, D. L. (2005). Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology*, 46(4), 621–646.
- ESIRI, Ejoro. (2011). Numeral System of Yoruba: Focus on the Dialect of Oyo B.A Dissertation
- FARGETTI, Cristina Martins (2007). Estudo fonológico e morfossintático da língua juruna. Muenchen: LINCOM-EUROPA.
- FARGETTI, C.M.; SUMAIO, P.A (2015). Numerals in Juruna. *LIAMES: Línguas Indígenas Americanas*, v. 15, n. 2.
- FREGE, G. (1974). The foundations of arithmetic: A logic-mathematical enquiry into the concept of number (2nd ed. rev., J. L. Austin, Trans.). Oxford, England: Blackwell. (Original work published 1884)
- GREEN, Diane (1997). Diferenças entre termos numéricos em algumas línguas indígenas do Brasil. *Boletim do Museu Paraense Emílio Goeldi-CNPq. Série Antropologia* 13(2): 179-207. Disponível em: <http://www.sil.org/americas/brasil/publens/ling/NumBrasi.pdf>.
- GRIFFITHS, G. 1975. Numerals and demonstratives in Kadiwéu, *Arquivos de Anatomia e Antropologia* . Instituto de Antropologia Professor Souza Marques, Vol.1:63-77.
- HAMMARSTRÖM, H. (2008). Complexity in numeral systems with an investigation into pidgins and creoles. In M. Miestamo, K.Sinnemäki, & F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 287-305). Amsterdam, The Netherlands: John Benjamins Publishing Company.
- HANLON, C. (1988) The emergence of set-relational quantifiers in early childhood. In: *The development of language and language researchers: Essays in honor of Roger Brown*, ed. F. S. Kessel. Erlbaum. [aLJR]
- HEIM, I. 1982. The Semantics of Definite and Indefinite Noun Phrases. Dissertação de doutoramento, University of Massachusetts, Amherst.

HOFWEBER, T. (2005) Number determiners, numbers, and arithmetic. *The Philosophical Review* 114:179–225.

HUFORD, J. R. (2010). *The linguistic theory of numerals*. Nova York: Cambridge University Press.

_____ (2006). A performed practice explains a linguistic universal: Counting gives the Packing Strategy. *Lingua*, 117(5):773--783.

_____ (2003). The interaction between numerals and nouns. F. Plank (ed.) *Noun Phrase Structure in the Languages of Europe*, pp. 561-620.

_____ (1987). *Language and number: the emergence of a cognitive system*. Basil Blackwell, Cambridge.

IONIN, T.; MATUSHANSKY, O. The Composition of Complex Cardinals. *Journal of Semantics*, v. 23, n. 4, p. 315–360. 2006.

KAKUMASU, J. & KAKUMASU, K. 1988. comp. *Dicionário Por Tópicos Urubú-Kaapor -Português*. Brasília, Summer Institute of Linguistics. 348 p.

KAMP, H. 1981. 'A Theory of Truth and Semantic Representation', in: Jeroen A. G. Groenendijk, T.M.V. Janessen & Martin B.J. Stokhof (eds.), *Formal Methods in the Study of Language*, Mathematical Centre Tract 135, - Reprinted in: Jeroen A. G. Groenendijk, T. M. V. Janesse (eds.), *Truth, Representation and Information*, GRASS series n° 2, Dordrecht, pp. 277- 322.

KARAJÁ, I., FORTUNE, D. & FORTUNE, G. eds. 1989. *Pré-Matemática para Falantes da Língua Karajá. Como Transmitimos o Conhecimento Cultural às Nossas Crianças: Como as Crianças Aprendem a Contar e a Calcular*. 2ª ed. Brasília, Summer Institute of Linguistics. 111 p.

KEENAN, E. L. and L. FALTZ: 1984, *Boolean Semantics for Natural Language*, Reidel, Dordrecht.

KOŞANER, Özgün. (2016) "Numerals in Turkish." *Open Journal of Modern Linguistics* 6.2 p. 131–147.

IONIN, T., O. MATUSHANKY (2006). The Composition of Complex Cardinals. *Journal of Semantics*, 23,4, p. 315–360.

LANDIN, D. 1983. Dicionário e Léxico Karitiâna - Português. Brasília, Summer Institute of Linguistics. 197 p.

LEITE, Yonne e FRANCHETO, Bruna. 500 Anos de Línguas Indígenas no Brasil-MS,

LIMA, Suzi. In press. Numerosidade em línguas indígenas brasileiras. Índio não fala Tupi. Bruna Franchetto (ed).

MAZAUDON, Martine. 2009. Number building in Tibet-Burman languages. *Nosth East Indian Linguistics*, Vol. 2, ed. By Stephen Morey & Mark W. Post, 117-148. New Delhi: Cambridge University Press India.

MCMAHON, April; MCMAHON, robert. Language classification by numbers. oxford: oxford University Press, 2005.

MEDEIROS, Nádia M. J.; MIRANDA, Maxwell G. Análise preliminar dos quantificadores numéricos e não-numéricos em Krenak (Família Botocudo) In Dermeval da Hora (org.). Anais do VI Congresso Internacional da ABRALIN, pp. 3944-3949. João Pessoa: Idea Editora. 2009

MOORE, Denny; GABAS, N. O Futuro das Línguas Indígenas Brasileiras. In: Louis Forline; Ima Vieira; Rui Murrieta. (Org.). Amazônia além dos 500 Anos. Belém: Museu Paraense Emílio Goeldi, v. , p. 433-454. 2006

MULLER, Ana ; STORTO, L. ; COUTINHO-SILVA, T. . Número e a Distinção Contável-Massivo em Karitiana. *Revista da ABRALIN*, v. 5, p. 185-213, 2006.

RODRIGUES, A. Línguas brasileiras: para o conhecimento das línguas indígenas. São Paulo, Loyola. 1986

ROCHA, Ivan. A Estrutura Argumental da Língua Karitiana: desafios descritivos e teóricos. Master's Thesis. Universidade de São Paulo. 2011

ROTHSTEIN, Susan. A Fregean semantics for number words. In Aloni, Maria, Franke, Michael and Roelofsen, Floris (eds.), *Proceedings of the 19th Amsterdam Colloquium*, pp. 179–86. 2013´

_____. Numericals and How They Work. In *Semantics for Counting and Measuring (Key Topics in Semantics and Pragmatics*, pp. 9-48). Cambridge: Cambridge University Press. 2017

SANCHEZ-MENDES, L. ; MULLER, Ana . The Meaning of Pluractionality in Karitiana. In: *Semantics of Under-Represented Languages in the Americas 4*, 2007, São Paulo. UMOF 35: Proceedings of the 4th Conference on the Semantics of Under-Represented Languages in the Americas. Amherst, MA, USA : GLSA. v. 35. p. 247-257. 2007

SEKI, Lucy (2000b). Gramática do Kamaiurá. Língua Tupi-Guarani do Alto do Xingu. Campinas, SP: Imprensa Oficial e Editora da UNICAMP.

VIGNADO, J.N. (2017). A sintaxe e a semântica de expressões numéricas do karitiana . *Revista de Letras da UFPR*.

WELLMAN H. M., & MILLER K. F. (1986). Thinking about nothing: Developmental concepts of zero. *British Journal of Developmental Psychology*, 4, 31–42.

WILLIAMS, J. 1932. Grammar Notes and Vocabulary of the Language of the Makuchi Indians, S. Gabriel-Modling near Vienna, Austria, *International Review Anthropos*. p. 418