



Programa de
Pós-Graduação em
Linguística

**COMPLEXIDADE LINGUÍSTICA: UM PANORAMA E UM ESTUDO DE
CASO**

Ednei de Souza Leal

SÃO CARLOS

2020



Universidade Federal de São Carlos

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA

**COMPLEXIDADE LINGUÍSTICA: UM PANORAMA E UM ESTUDO DE
CASO**

Ednei de Souza Leal

Bolsista CAPES

Tese apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Doutor em Linguística, área de concentração: Descrição, Análise e Processamento Automático de Línguas Naturais.

Orientador: Prof. Dr. Renato Miguel Basso

Leal, Ednei de Souza

Complexidade linguística: um panorama e um estudo de caso / Ednei de Souza Leal -- 2020.
130f.

Tese de Doutorado - Universidade Federal de São Carlos, campus São Carlos, São Carlos Orientador (a): Renato Miguel Basso
Banca Examinadora: Márcio Renato Guimarães, Alessandro Jocelito Beccari, Rosa Yokota, Dirceu Cleber Conde, Cindy Mery Gavioli-Prestes, Oto Araújo Vale Bibliografia

1. Complexidade Linguística, Filosofia da Linguística, Historiografia da Linguística. I. Leal, Ednei de Souza. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática (SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Ronildo Santos Prado - CRB/8 7325



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Linguística

Folha de Aprovação

Defesa de Tese de Doutorado do candidato Ednei de Souza Leal, realizada em 27/08/2020.

Comissão Julgadora:

Prof. Dr. Renato Miguel Basso (UFSCar)

Prof. Dr. Márcio Renato Guimarães (UFPR)

Prof. Dr. Alessandro Jocelito Beccari (UNESP)

Prof. Dr. Dirceu Cleber Conde (UFSCar)

Profa. Dra. Rosa Yokota (UFSCar)

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Linguística.

Banca

Prof.^a. Dr.^a. Rosa Yokota

Prof. Dr. Dirceu Cleber Conde

Prof. Dr. Márcio Renato Guimarães

Prof. Dr. Alessandro Jocelito Beccari

Prof.^a. Dr.^a. Cindy Mery Gavioli-Prestes (Suplente)

Prof. Dr. Oto Araújo Vale (Suplente)

Resumo

Após décadas, a complexidade linguística volta a ser objeto de estudo na linguística. Diferentemente, porém, do século XIX, em que a complexidade linguística já havia sido objeto de estudo, agora livre dos preconceitos que vinculavam a língua à capacidade mental dos seus falantes, a complexidade linguística não é mais vista hoje como um truísmo, como transparecia na visão estruturalista até meados do século XX. Em pleno século XXI, novos entendimentos sobre a língua, trazidos por ferramentas de análise cada vez mais sofisticadas e por ideias cada vez mais claras sobre a tipologia das línguas, é finalmente possível estudar de modo pleno a complexidade, característica inerente a toda língua natural, mas que difere em suas nuances e detalhes hoje medidos por linguistas que propõem métricas para, praticamente, todos os níveis tradicionalmente estudados das línguas naturais, tais como Dahl (2004), McWhorter (2001), Culicover (2012), Kusters (2003). Nesse sentido, esta tese propõe uma visada panorâmica, mas aprofundada, acerca dos estudos sobre complexidade linguística desde a última década do século XX aos dias de hoje. Para tanto, propomos nesta tese, primeiramente, um apanhado historiográfico não exaustivo do termo “complexidade linguística” ao longo da história da linguística ou de estudos que se inclinaram a versar sobre a língua. Feito o apanhado histórico, procuramos mostrar de que forma a complexidade linguística é vista hoje em dia como objeto genuíno da ciência da linguagem. Importa ainda mencionarmos as diversas técnicas para medir complexidade linguística, nos mais variados níveis estudados das línguas. Por fim, propomos um breve exercício de análise para um fenômeno característico do português brasileiro (PB), as construções de tópico, com a finalidade de saber se este fenômeno torna a língua mais ou menos complexa com relação a sua variedade europeia, que parece não apresentar o fenômeno com tanta frequência nem do mesmo modo que o PB. Este trabalho toma como parâmetro teórico-metodológico as propostas da Filosofia da Linguística e da Historiografia da Linguística, bem como as propostas já em voga de diversos linguistas que se propõem a estudar a complexidade linguística em suas mais variadas formas.

Palavras-chave: Complexidade Linguística, Filosofia da Linguística, Historiografia Linguística, Português Brasileiro, Construções de Tópico.

Abstract

Language complexity is once again the object of study in Linguistics after being largely assumed as a resolved or uninteresting issue. Differently, however, from the 19th century perspective, during which linguistic complexity had already been the object of study, now its investigation is free from the prejudices that associated languages to the mental capacity of their speakers, and linguistic complexity is no longer seen today as a truism, as it transpired in the Structuralist view until the middle of the 20th century. In the 21st century, new understandings about the language, brought by increasingly sophisticated analytic tools and by increasingly clear ideas about the typology of languages, it is finally possible to fully study language complexity, a characteristic inherent to every natural language, but that differs in its nuances and details, and is today measured by linguists who propose metrics for practically all levels traditionally investigated by Linguistic Studies, like Dahl (2004), McWhorter (2001), Culicover (2012), Kusters (2003). In this sense, this thesis proposes a panoramic look at studies on linguistic complexity from the last decade of the twentieth century to today. We propose, firstly, a non-exhaustive historiographic overview of the term “linguistic complexity” throughout the history of linguistics or studies that were inclined to deal with language. After the historical overview, we seek to show how linguistic complexity is seen today as a genuine object of linguistic studies. It is also important to mention the different techniques for measuring linguistic complexity, at the most varied levels of languages studied. Finally, we propose a brief exercise analyzing a phenomenon characteristic of Brazilian Portuguese (BP), topic constructions, with the purpose of investigating whether this phenomenon makes the language more or less complex in relation to its European variety, which seems to present the phenomenon not as often or in the same way as BP. This thesis takes as a theoretical-methodological parameter the proposals of Philosophy of Linguistics and Historiography of Linguistics, as well as the proposals already in vogue of several linguists who propose to study linguistic complexity in its most varied forms.

Keywords: Language Complexity, Philosophy of Linguistics, Historiography of Linguistics, Brazilian Portuguese, Topic Constructions.

Dedicatória

Dedico esta tese à minha família:

A minha mãe, Dona Lair pela vida;

A minha esposa, Simone pelo amor;

A meus filhos, André e Melody pela vontade de viver;

A meus irmãos: Eli, Élcio e Elizângela pela dedicação e incentivo.

Aos meus sobrinhos: Lucas (in memoriam), Maria Izabel, Melissa, Anabela e João Pedro.

Agradecimentos

Agradeço à CAPES pela bolsa, sem a qual esta tese não teria sido realizada;

A UFSCar pelo acolhimento;

Ao meu orientador, Professor Doutor Renato Miguel Basso, pela incansável dedicação;

Aos professores da Banca, Professor Doutor Márcio Renato Guimarães; professor doutor

Alessandro Jocelito Beccari; professora doutora Rosa Yokota; professor doutor Dirceu Cleber

Conde; professor doutor Oto Vale e professora doutora Cindy Mery Gavioli-Prestes. Pelas

leituras, pelos conselhos e, sobretudo, pela amizade nessa jornada.

Ao programa de pós-graduação em Linguística da UFSCar (PPGL-UFSCar), sobretudo a Vanessa

Rodrigues pelo apoio incondicional.

Ao Grupo de Estudos e Pesquisa em Terminologia (GETerm), na pessoa da professora Gladis M.

de Barcellos Almeida, pelo uso do laboratório, onde parte desta tese foi redigida.

Aos meus colegas (e amigos) linguistas: Fábio Mesquita, Gesualda Rasia, Jorge Onodera, Letícia

Kolberg, Roger Alfredo Antunes, Thayse Ferreira, Felipe Dall'ava e Jackson Souza, pelas

calorosas discussões.

Aos meus antigos (e eternos) professores: Doutora Lígia Negri e Doutor José Borges Neto.

Aos meus amigos: João Guilherme da Silva Filho, Luiz Fernando Alves Ferreira, Rafael

Ivancheche e Harry Cowl.

Ask Joe the Plumber, Piers the Ploughman, or Tom the Piper's Son what sort of languages the half-naked tribes in the Amazonian rain forest speak, and they will undoubtedly tell you that "primitive people speak primitive languages." Ask professional linguists the same question, and they'll say something quite different. Actually, you don't even need to ask-they will tell you anyway: "All languages are equally complex." This battle cry is one of the most oft avowed doctrines of the modern discipline of linguistics. For decades, it has been professed from lecterns across the globe, proclaimed in introductory textbooks, and preached at any opportunity to the general public.

Pergunte a Joe, o encanador, Piers, o lavrador, ou Tom, o filho de gaiteiro, que tipo de idiomas falam as tribos seminuas da floresta amazônica e eles indubitavelmente lhe dirão que "os povos primitivos falam idiomas primitivos". Faça a mesma pergunta aos linguistas profissionais, e eles dirão algo bem diferente. Na verdade, você nem precisa perguntar - eles dirão mesmo assim: "Todos os idiomas são igualmente complexos". Esse grito de guerra é uma das doutrinas mais frequentemente reconhecidas da disciplina moderna da linguística. Por décadas, ele tem sido proferido em leituras por todo o mundo, proclamado em livros introdutórios de linguística e pregado em qualquer oportunidade para o público em geral.

(Guy Deutscher – Through the Language Glass)

Lista de Abreviaturas

CAF – *Complexity, Accuracy and Fluency*

CL – complexidade linguística

CP – *complementizer phrase*

CSP – construção ou construções de sujeito-predicado

CT – construção ou construções de tópico

HoL – *Historiography of Linguistics*, historiografia da linguística

IA – inteligência artificial

IP – *inflectional phrase*

L1 – primeira língua ou língua materna

L2 – segunda língua

PB – português brasileiro

PE – português europeu

PIE – protoindo-europeu

S – sentença

SLA – *Second Language Acquisition*

SN – sintagma nominal

SV – sintagma verbal

SVO – sujeito, verbo, objeto

TopP – *topic phrase*, sintagma de tópico

V – verbo

Sumário

Capítulo 0 – CONSIDERAÇÕES INICIAIS	11
0.1. Primeiras Considerações	11
0.2. Objeto do Nosso Estudo nesta Tese	13
0.3 Algumas Definições Necessárias e Algumas Propostas em Voga	14
Capítulo 1 – Historiografia da noção de Complexidade Linguística	19
1.1 Introdução	19
1.2 Século XIX.....	20
1.3 Estruturalismo – Primeira Metade do Século XX.....	26
1.4 Pós-estruturalismo – Segunda Metade do Século XX	29
1.5 Desafios para os estudos sobre Complexidade Linguística	34
1.6 Conclusão.....	37
Capítulo 2 – Definindo Complexidade	39
2.1 Introdução	39
2.2 O que a filosofia entende por complexidade.....	40
2.2.1 Tentando definir um termo nada simples	40
2.2.2 Sistemas adaptativos complexos	45
2.3 Complexidade Linguística	47
2.3.1 A complexidade linguística e suas definições de hoje	48
2.3.2 Complexidade Absoluta e Complexidade Relativa.....	51
2.3.3 Maturação de complexidade linguística e a idade das línguas	52
2.4 Conclusão.....	55
Capítulo 3 – Alguns modelos para se medir complexidade Linguística.....	56
3.1 Introdução	56
3.2 Propostas para Medir Complexidade Linguística	56
3.2.2 Pelo viés da estrutura.....	57
3.2.2 Pelo viés do uso.....	67
3.2.3 Pelo viés da aquisição	73
3.2.3.1 Método CAF e a Complexidade a serviço da SLA	75
3.2.3.2 Outros métodos disponíveis	76
3.3 Conclusão.....	82
Capítulo 4 – Construções de Tópico no Português Brasileiro medido sob o viés de um modelo de complexidade linguística: um exercício de análise.....	84
4.1 Introdução.....	84
4.2. Definição do que sejam as CT	86
4.2.1 Articulação Tema-Rema, Ruptores e outras noções	86

4.2.2. Tipos de Tópicos correntes na literatura – uma classificação	93
4.3 Dahl (2004) aplicado às CT.....	95
4.4 Exercício de interpretação conforme noções de Dahl (2004)	101
4.5 Conclusão.....	109
Capítulo 5 - Considerações Finais	111
Referências Bibliográficas	122

Capítulo 0 – CONSIDERAÇÕES INICIAIS

0.1. Primeiras Considerações

Um dos mais surpreendentes aspectos descobertos sobre a linguagem humana nos últimos tempos é o fato de que ela é uma entidade natural. Hoje em dia chega até a parecer um tanto absurdo, mas parte do senso comum admitiu (e ainda admite) que a língua natural seria a invenção de algum deus, ou mesmo de alguma mente prodigiosa em tempos imemoráveis. A língua não é invenção de uma superinteligência planejada, mas sim fruto da inteligência da humanidade, feita em sua coletividade, e também de suas pré-disposições genético-biológicas. Ao menos esta é a explicação mais aceitável hoje em dia para a linguagem humana: o que não é pouca coisa, nem tão pouco tal conclusão foi alcançada de maneira fácil. Além disso, a língua é, como aponta Humboldt já no início do século XIX, não um produto acabado dos humanos, o que ele chamou “*energeia*”, mas se encontra em construção constante (Humboldt 1836 *apud* Heidermann & Weininger 2006). Em outras palavras, isso quer dizer que a língua está em constante mudança, e não é, ao menos num sentido claro, um objeto acabado ou pronto. Portanto, ser natural e mudar constantemente são características inerentes das línguas humanas. Talvez essa seja mais uma das razões para que as línguas naturais, ao contrário do que muito se pensou, sejam um objeto de estudo tão complexo de descrever ou mesmo de explicar.

Outro fato curioso que foi descoberto a respeito das línguas – ou que ainda se trata de uma nova descoberta no caso de vários não-especialistas – é que a escrita não revela tudo aquilo que há na língua. Ao contrário da língua falada, a escrita não é natural. Na verdade, ela é mais uma dentre tantas outras tecnologias inventadas pelo homem (Auroux 1992), assim como os dicionários e a gramática¹: uma das primeiras formas de descrição da língua. Dessa forma, nem a escrita nem tampouco a gramática podem nos revelar a grandiosidade da língua em sua forma mais natural: em constante mudança, a língua viva.

Dada a persistência dos estudos gramaticais no Ocidente, como fonte quase exclusiva de metalinguagem nos estudos linguísticos, algumas questões em torno da língua se deram de forma pouco esclarecedoras durante muito tempo. A princípio, a tradição gramatical, impregnada nos currículos escolares, fez crer que a prescrição gramatical seria o ideal de língua em todas as suas manifestações. Ora, o problema de se ter a gramática como parâmetro absoluto do que seja a língua

¹ A gramática a que nos referimos aqui é aquela materializada em forma de livro ou manual, que frequentemente procura descrever a língua ou, em casos extremos, prescrever normas.

é o fato de que sua base teórica ser pré-científica² e excessivamente especulativa (cf., Leal, 2015). Além disso, sua imposição prescritiva ultrapassa os muros da escola e impregna também o senso comum, empurrando o conhecimento científico sobre as línguas naturais para os confins do conhecimento acadêmico.

A crença de que a gramática prescritiva revela tudo sobre a linguagem humana é um dos tantos mitos em torno da língua, cuja desmistificação é uma das tarefas do linguista. Seja como for, uma das tarefas da ciência, se não a principal é desfazer mistificações que o senso comum atribui à natureza. Assim, nas ciências da linguagem há, naturalmente, desacordos e equívocos que persistem ao longo de sua história (Popper, 1993 *apud* Borges Neto, 2002). Dentre estes, na área dos estudos linguísticos, está a apressada e então “politicamente correta” afirmação de que as línguas naturais, todas elas, se equivaleriam em seus **níveis de complexidade** – como esperamos mostrar, (i) há muita polêmica em tal discussão que, ainda que muitas vezes bem intencionada, tem pouco de científica, e (ii) não é nada fácil definir clara, consensual e operacionalmente a noção de complexidade linguística. A ideia “pacificadora” básica e tacitamente aceita é que haveria uma espécie de “lei de equilíbrio” misteriosa que garantiria que todas as línguas naturais tenham o mesmo nível de complexidade, fazendo uma espécie de compensação entre os tradicionais níveis de análise linguístico, como a fonética, a fonologia, a morfologia, a sintaxe, a semântica e a pragmática. Noção esta, ou mito, pois não há base científica que confirme tal lei, que se convencionou chamar de “**equicomplexidade** das línguas naturais”. Por exemplo, se uma dada língua *A* apresenta em seu paradigma maior número de flexões verbais que dada língua *B*, a tal língua *B* deverá, necessariamente, compensar esse fato com, digamos, uma ordem sintática regida por mais regras que a mesma descrição da língua *A*, de modo que, ao fim e ao cabo, as duas sejam igualmente complexas.

É verdade que nem sempre se pensou dessa maneira. Até o século XIX, havia quem pensasse que, se conformando aos “hábitos primitivos” de certos povos selvagens, algumas línguas também se comportariam de maneira semelhante, ou seja, de modo simplificado ante as chamadas línguas de cultura (Kilarski, 2014). Acreditava-se que o sânscrito, o grego e o latim seriam, então, as línguas mais bem acabadas de que se tinha notícia. O parâmetro dessa afirmação se dá pela observação de antigas descrições dessas línguas, que dão conta de enxergar em suas morfologias uma riqueza tal que não são supostamente encontradas em certas línguas orientais, por exemplo. Isso talvez tenha levado ao equívoco de afirmar que o latim, o sânscrito e o grego seriam línguas mais complexas e, portanto, nesse sentido, mais ricas do que línguas que apresentam paradigmas morfológicos menos

² É importante dizer que o que é “pré-científico” aqui deve ser lido como “não se conformando aos atuais critérios de cientificidade”, como fazer previsões, não ser contraditório, ter poder heurísticos etc. É certo que a descrição gramatical nas várias épocas diferentes foi feita sob certos critérios avaliativos e críticos, como a obra de Panini, por exemplo, mas eles não atendem aos nossos conceitos recentes do que seja “científico”. Agradeço aos professores Márcio Guimarães e Alessandro Beccari por esta observação já na banca de qualificação.

densos. Um outro equívoco, que durou anos, foi usar as estruturas do latim e do grego tanto para descrever línguas que não apresentavam as mesmas estruturas quanto para servir de “métrica” de quão bem acabada ou complexa é uma dada língua, ou seja, quanto mais próxima das línguas clássicas, tanto melhor³. Na verdade, o que se ensinava até o século XVIII era o latim, as gramáticas vernaculares surgiram justamente para auxiliar no ensino do latim como segunda língua (Beccari 2013).

O estruturalismo, já na primeira metade do século XX, encontra uma alternativa plausível para a questão da complexidade linguística: a equivalência de todas as línguas naturais através da então **lei das compensações** (Hockett, 1954). Seguindo os preceitos da antropologia protagonizada por Franz Boas (2012 [1931] *apud* Kilarski, 2014), a Linguística – apressadamente, a nosso ver, pois sem o devido embasamento empírico – afirma que em todas as línguas naturais haveria equivalência de complexidade. Tal noção resolveria muitos problemas, sobretudo problemas de ordem ética e social no que tange à língua, mas, sobretudo, a seus falantes. Ora, se se segue a fórmula prevista pelos linguistas do século XIX de que certas línguas se diferem em suas complexidades, também se estaria autorizado a afirmar que falantes de línguas menos complexas também pensavam de forma mais simples, logo, que eram menos capazes, o que possibilitaria e justificaria uma série de ações atrozadas que, de fato, infelizmente, aconteceram. Nesse sentido, a linguística estruturalista procurou cessar todo tipo de discurso enviesado ou mesmo eticamente equivocados sobre as eventuais diferenças ou discrepâncias de complexidade entre as línguas naturais – um gesto louvável e necessário, absolutamente necessário, porém, com uma base empírica inexistente ou muito fraca.

0.2. Objeto do Nosso Estudo nesta Tese

Justamente o que faremos nesta tese é discutir a polêmica em torno do truísmo linguístico, que já dura décadas sobre a complexidade linguística (a qual, a partir daqui trataremos simplesmente por complexidade ou por CL, quando oportuno). Para realizar tal discussão é preciso, antes de mais nada, que tracemos algumas definições: 1) o que é complexidade; 2) o que é complexidade linguística, se se fala em diferenças de complexidade entre as línguas, é lícito que se tenha em mente algum parâmetro de medida de tais complexidades; 3) como é possível medir a complexidade linguística. Tais questões

3 A história dos estudos do português também é evidência disso. João de Barros, em sua gramática de 1540, afirmava, por exemplo, que o português tinha os mesmos casos do latim, a única diferença é que os representava por meio de preposições. Essa aproximação entre as línguas não é à toa, é uma forma de garantir status ao português. Em sua gramática, acessível em (acesso em junho de 2020): http://purl.pt/12148/6/res-5658-1-p_PDF/res-5658-1-p_PDF_24-C-R0150/res-5658-1-p_0000_capa-60v_t24-C-R0150.pdf é possível vermos, na página 28, uma tabela de declinação dos artigos do português. A ideia de que o português tinha casos (preposições) persistiu até pelo menos o século XVIII.

são envoltas em disputas ideológicas que podem ofuscar considerações mais objetivas. Nunca é demais salientar: nosso objetivo aqui não é minimizar ou criticar as consequências éticas e ideológicas positivas que seguem da ideia de que todas as línguas são igualmente complexas, mas sim buscar bases científicas, objetivas e empíricas que possam justificar em termos científicos, de preferência linguísticos, essa ideia.

0.3 Algumas Definições Necessárias e Algumas Propostas em Voga

Diferentemente das conclusões talvez um pouco precipitadas que resultaram das noções postuladas nos séculos XIX e XX, definir CL hoje exige um cabedal teórico mais robusto e interdisciplinar e, por isso, definições mais refinadas vêm surgindo desde, pelo menos, os anos 1990. Linguistas que se dedicavam à aquisição de segunda língua e ao ensino de língua estrangeira viram a necessidade de criar parâmetros para mensurar a proficiência dos falantes aprendizes – como é o caso dos trabalhos Ellis (1980 *apud* Ellis, 1997), Larssen-Freemann (1975 *apud* Ellis, 1997), Ligthbown (1980 *apud* Ellis, 1997), entre outros. Timidamente, essas reflexões suscitaram as primeiras noções de que, em certos níveis ao menos, as línguas poderiam apresentar disparidades no tocante à sua complexidade quando levamos em conta contextos de aprendizagem de línguas ou de diferenças estruturais de línguas envolvidas num processo de aprendizagem.

No âmbito da sociolinguística, por meio de ferramentas teórico-descritivas cada vez mais finas, a questão da CL veio à tona na chamada linguística teórica. Costuma-se atribuir a Trudgill (1983 *apud* Miestamo *et al.*, 2008), com seu estudo sociolinguístico sobre algumas variedades do grego moderno, a primeira indagação sobre diferenças de complexidades linguísticas, desde pelo menos os anos 50. No final dos anos 90 começam indagações mais profundas e sistemáticas, na esteira dos estudos sobre *sistemas complexos*, sobre a CL, que levam em conta critérios sociolinguísticos, como o tipo e quantidade de interações nas quais os falantes de uma dada sociedade ou comunidade fazem, bem como o tamanho e a complexidade da própria comunidade ou sociedade investigada.

No início dos anos 2000, o linguista John McWhorter publica um trabalho que levanta muita polêmica em torno da estrutura de línguas emergentes, notadamente as línguas crioulas. Ao publicar um estudo com o provocativo título de “*The world’s simplest grammars are creole grammars*”, em 2001, o autor argumenta a favor da hipótese de que os crioulos podem ser estudados pelo recorte sincrônico, e sugere que há redução de complexidade resultante de contatos linguísticos.

Além dessas contribuições iniciais, os estudos sobre complexidade linguística se valem também da mudança e história da linguística (estudos diacrônicos). A obra mais importante a contribuir neste aspecto talvez seja o livro “*The Growth and Maintenance of Linguistic Complexity*”,

de Östen Dahl, publicado em 2004. Há outras áreas de investigação que contribuem para o estudo da CL, e é importante citar também estudos da evolução linguística (Mufweme, 2013), da teoria da gramaticalização e da tipologia linguística (ver Miestamo *et al.*, 2008), principalmente. Embora o problema em torno da complexidade tenha nascido no âmbito do funcionalismo, onde se encontram talvez modelos mais robustos e refinados de se discutir e medir complexidade, vêm surgindo modelos formalistas, sobretudo baseados em postulados da gramática gerativa. Tais modelos vêm também se mostrando ferramentas eficientes para tratar do assunto, expostos em obras de linguistas como Culicover (2013) e Newmeyer & Preston (2014).

Junto a isso, já em 1998 o filósofo inglês Nicholas Reschen (1998 *apud* Miestamo *et al.* 2008) foi um dos pensadores que procurou circunscrever em termos mais precisos o que seja complexidade. Em sua obra “*Complexity: A Philosophical Overview*”, o autor dedica algumas páginas à complexidade linguística, estabelecendo uma taxonomia do que seja a complexidade das línguas; nesse trabalho, a língua é vista sob três diferentes modos: epistêmico, ontológico e funcional, e eles dariam conta de detectar e mesmo medir complexidade linguística. Tal taxonomia seria mais tarde adotada por linguistas que se inclinam a estudar o assunto. Outro filósofo que se dedica a estabelecer parâmetros para os estudos de complexidade linguística é Bruce Edmonds. Sua tese, intitulada “*Syntactic Measures Complexity*”, de 1999, propõe que cada objeto de estudo, dada sua natureza particular, apresente de forma igual suas respectivas complexidades, que lhe são inerentes. Após essa tarefa de identificar a natureza complexa de dado objeto – a língua, por exemplo – é feita uma submissão computacional para que assim se torne possível efetuar uma eventual descrição da complexidade deste objeto. O mesmo ponto de vista filosófico tem o físico Murray Gell-Mann (1995) acerca da complexidade, que dependeria justamente do ponto de vista e do objeto cuja complexidade se pretende “medir”.

Os modernos estudos sobre CL nascem de uma antiga indagação, mas tolhida, como vimos, em nome do bom senso e de uma noção geral que não foi, na época, suficientemente examinada. Segundo Kilarski (2014), não se discutiu mais o assunto, simplesmente porque a complexidade parou de ser foco das atenções, precisamente porque se tornou ponto pacífico entre linguistas – um truísmo tacitamente aceito, como notamos. Como nota Deutscher (2014), tal truísmo, apesar de indubitavelmente ser menos preconceituoso e fazer mais justiça ao nosso atual conhecimento linguístico, ainda é muito problemático – afinal, como sabemos que as línguas são igualmente complexas? Como provar que esse é o caso? Ainda não possuímos uma conceituação científica precisa sobre complexidade linguística que nos permita ancorar o truísmo de que as línguas são igualmente complexas em bases empíricas e cientificamente aceitáveis. Embora tantos desacordos, o que se entende, de maneira genérica, o que seja CL não está atrelada à capacidade dos falantes e nem propriamente à expressividade das línguas. A CL, por um lado, está atrelada a uma característica

inerente de toda língua natural e, por outro, à forma como é possível enxergar um ou outro nível de dadas línguas.

Nesse sentido, linguistas como Miestamo *et al.* (2008) acreditam que é preciso circunscrever de maneira mais clara o conceito de CL, e, para tanto, se valem das noções de *complexidade absoluta* e *complexidade relativa*. De modo geral, tais conceitos são entendidos da seguinte maneira: a complexidade absoluta seria definida como uma propriedade objetiva do sistema linguístico (seja ele qual for); e a complexidade relativa seria vista como uma espécie de “custo”, ou mesmo dificuldade vista pelos falantes de uma dada língua para operá-la. Esta última noção geralmente é “medida” conforme a dificuldade dos falantes de L2, ou também através de análises do tamanho de inventários (fonemas ou processos morfológicos, por exemplo) ou de características peculiares inerentes a cada nível linguístico (ser ou não uma língua V2, por exemplo).

Tanto para Kusters (2003), quanto para Miestamo *et al.* (2008), a complexidade absoluta diz respeito ao modo de ver ou mesmo de medir complexidade de uma língua em todos os seus subsistemas, chegando assim a uma espécie de valor global. Desse modo, para medir ou enxergar a complexidade de uma língua, deve-se observar seu sistema por inteiro, e o que definiria sua complexidade então seria a quantidade de subsistemas atrelada a esse sistema “geral” que, de maneira abstrata, descreve as línguas naturais. Sendo assim, seria supostamente mais complexa uma língua que apresente maior grau de complexidade nos seus subsistemas. Para os mesmos autores, a complexidade relativa teria a ver com medir cada um desses subsistemas (obviamente definindo, então, quais e quantos são eles) em separado, para daí se obter alguma resposta pertinente no tocante à complexidade de uma língua.

Uma solução possível para esse problema, talvez seja medir apenas complexidade relativa, ou seja, aquele tipo de complexidade na qual a junção das partes é tão importante quanto o todo. Por exemplo, quando ocorrem mudanças linguísticas, digamos, num dado traço de um sistema fonológico de uma língua, é também bastante provável que haja mudanças em outros sistemas dessa mesma língua, acarretadas pela mudança fonológica, o que geralmente acontece.

Embora a questão sobre a CL tenha sido levantada primeiramente por linguistas que estudam ensino de segunda língua, sociolinguistas e tipologistas, não são somente esses pesquisadores que se ocupam de tal debate atualmente, como vimos, e parece haver, entre as propostas de medidas de complexidade, três pontos de vista principais que culminam em medir complexidade linguística (i) pelo viés do **uso** da língua, (ii) pelo viés do **falante**, ou (iii) pelo viés da **estrutura** da língua. Veremos mais detidamente essa questão no capítulo 3

Dadas as definições e algumas propostas de medida da CL, parece que fica claro que nossas motivações para o trabalho desta tese são suficientemente importantes e mesmo prementes, visto que o assunto ainda permanece inédito no Brasil. De todo modo, mais do que a necessidade de apresentar

o assunto ao público linguista, faz-se importante os debates em torno de tal truísmo. Seja pelo ineditismo ou pela importância em debater assunto até agora pouco discutido em língua portuguesa, a CL tem se arrastado como um truísmo na linguística e, como tal, precisa ser resolvido.

Embora razoavelmente bastante difundido entre linguistas Brasil afora, as obras de McWhorter (2001, 2005, 2007) e de Deutscher (2000; 2009; 2014) não têm como foco principal as discussões em torno da CL, ainda que suas preocupações e estudos levem seus leitores a concluir que a equicomplexidade das línguas vem se tornando, ante várias novas evidências, uma noção praticamente impossível de se sustentar. Mesmo linguistas bastante competentes que procuram discutir truísmos da língua, tal como Othero (2017), ou apresentar novas tendências da linguística mundo afora, como Fiorin *et al.* (2017) é ainda comum encontrar, mesmo recentemente, pesquisadores que repetem a ideia de equicomplexidade – o que é compreensível, dado que tal dogma ainda é pouco discutido mesmo nos meios especializados. Portanto, diante de extenso material, sobretudo em língua inglesa, e diante da premência da discussão, faz-se não só útil, mas essencial nosso trabalho aqui exposto. Além disso, julgamos ser nossa tarefa a de esclarecer os usos e definições por trás desse conceito, já que também é possível, e infelizmente comum ainda nos dias de hoje, incorrer ao erro fatal de associar CL a superioridade cultural/social. Nesse sentido, seguindo o que apontamos no capítulo de historiografia (capítulo 1), procuramos mostrar quão equivocadas estavam diversos dos conceitos do que seja a complexidade das línguas. Por esses motivos é que se carece de uma boa definição do que seja complexidade linguística em termos empírico-científicos precisos.

Justamente por conta de eventuais mal entendidos que venham a surgir a partir das propostas de complexidade linguística, sobretudo de como se medir CL, é que fazemos do cerne desta tese uma discussão fundamental, calcada nos parâmetros da chamada filosofia da linguística.

A filosofia da linguística, ramo da filosofia da ciência, como qualquer outra investigação de qualquer outra ciência, não se sustenta sem um bom estudo historiográfico. Assim sendo, amparando a filosofia da linguística, que se ocupará das discussões acerca dos atuais estudos sobre CL, faremos um apanhado historiográfico amplo em torno do que se entendeu por complexidade das línguas ao longo dos séculos, mais precisamente, desde o século XIX, quando a CL ganha *status* entre linguistas, mais ou menos na mesma época em que temos a primeira proposta de tipologia linguística (Schleicher, 1856 *apud* Guimarães 2020). Posteriormente investigaremos, sob o mesmo crivo, outras importantes correntes de pensamento linguístico que se debruçaram sobre o assunto e o debateram, até o momento em que a CL se torna ponto pacífico entre os linguistas, e para de ser debatida, precisamente, na primeira metade do século XX. Este trabalho exige um aparato adequado, que é a chamada historiografia da linguística (HoL na sigla inglesa *Historiography of Linguistics*), ramo de estudo que vem se firmando no cenário dos estudos desde os anos 70. Dessa forma, tomaremos os parâmetros daqueles que julgamos ser os trabalhos mais importantes da HoL: Koerner & Asher

(1989), Auroux (1992), Altman (2012), para justificarmos posições teóricas que aqui regerão escolhas e recortes da HoL. Para tanto, nos valeremos do *princípio de contextualização*, bastante recorrente nos estudos da HoL, a partir do qual procuraremos reconstituir um *clima de opinião* para que possamos vislumbrar com mais clareza o que se entendia outrora por complexidade linguística (Koerner e Asher 1995).

A complexidade linguística e os estudos que a levam em conta têm um poder heurística a ser ainda explorada, e não apenas em sociolinguística, linguística histórica e ensino de segunda língua, mas provavelmente em várias áreas de estudo linguístico. Embora tais questões sejam pertinentes, os estudos acerca da complexidade linguística ainda não se encontram de tal modo organizados a ponto de formar um programa científico. Por conta disso, achamos que seria interessante fazermos uma revisão científica de seus parâmetros e limites: tarefa adequada, portanto, para a filosofia da linguística. Para amparar esta nossa escolha teórica, nos valeremos das ideias encontradas em Borges Neto (2002).

Após essa longa exposição, nesta tese ainda caberá, o que na verdade, a nosso ver, faz-se necessário, a aplicação ou análise, como um exercício, de um fenômeno linguístico sob o crivo dos estudos da CL. Tomaremos, aos moldes de Kusters (2003) e principalmente de Dahl (2004), duas variedades da língua portuguesa, e dela, debateremos um fenômeno já descrito desde Pontes (1987), qual seja: a construção de tópico do português. Nosso objetivo será tentar mostrar que tal fenômeno, que a princípio ocorre mais no português brasileiro (PB) do que em sua variedade europeia (PE), necessita de um número de descrição maior do que estruturas sintáticas que não possuem tal estratégia, ou seja, a complexidade linguística pode nos ajudar a entender certas dimensões desse fenômeno, revelando parte do seu poder heurístico que há pouco mencionamos. Procuraremos, assim, mostrar que o PB estaria mudando seu status tipológico, ao menos no tocante à relação do verbo com seu sujeito ou, no caso dos tópicos, uma nova organização sintática. Como sugere, por exemplo, Eunice Pontes (1987), o português brasileiro estaria passando por uma mudança nessa forma estrutural de suas sentenças, sendo assim, verifica-se através de *corpus* desta autora que sentenças de tópico e sentenças do tipo sujeito-predicado coexistem no PB. O resultado de tal estudo deverá ser amplamente auxiliado pelos novos parâmetros métricos da CL.

Esta tese está dividida em quatro capítulos, além destas considerações iniciais e das considerações finais. No primeiro capítulo faremos o apanhado historiográfico, com auxílio dos preceitos da HoL, conforme brevemente explicitamos acima. Nos dois capítulos seguintes, apresentaremos de forma mais detida as bases teóricas dos estudos sobre CL, bem como algumas propostas para mensurar complexidade nos seus mais diferentes níveis. No quarto capítulo faremos o estudo de caso.

Capítulo 1 – Historiografia da noção de Complexidade Linguística

1.1 Introdução

Provavelmente para boa parte do senso comum de hoje e mesmo para especialistas em línguas naturais ainda num período pré-saussureano⁴ era absolutamente aceitável, e até evidente, que havia diferenças de complexidade entre as línguas naturais conhecidas. A já sabida história de que os gregos tratavam como “bárbaras” as línguas diferentes da sua (Robins, 1979) ilustra bem essa noção, que então era corrente⁵. Kilarski (2014), por exemplo, atribui a noção de diferentes complexidades entre as línguas como base de muitos pensamentos equivocados, que poderiam levar a atitudes preconceituosas, as quais foram responsáveis por verdadeiras atrocidades, de certa forma, autorizados por afirmações de (pseudo) teorias antropológicas ou linguísticas. Essa noção serviu, então, como justificativa para se rebaixar não somente línguas, mas culturas e até populações inteiras em nome, por exemplo, do colonialismo, principalmente aquele exercido por europeus desde o século XVI. Talvez esse seja um dos principais motivos que levaram linguistas como Charles Hockett, que dedicou boa parte de sua carreira a estudar línguas nativas norte-americanas, a afirmar que não haveria discrepâncias substanciais no que tange às diferenças na complexidade entre as línguas.

Havia razões para que não se discutisse mais este assunto, dado que há décadas o que chamamos de complexidade linguística está associado a alguma espécie de valoração, muitas vezes subjetiva, da língua. Nada mais justo. Após severas e esclarecedoras discussões, notamos hoje, dados novos entendimentos e ferramentas teóricas, que os linguistas têm plena condição de voltar a tais discussões sem carregar os preconceitos do passado e pensando apenas objetiva e cientificamente sobre o que significa complexidade linguística e como medi-la. Antes de fazer essas discussões, no entanto, faz-se necessária uma justa digressão ao passado, através de uma pesquisa historiográfica circunstanciada. Portanto, neste capítulo faremos tal digressão, sem maiores polêmicas em torno do que seja a complexidade linguística. Tal assunto parece ser de fato demasiado polêmico e, até mesmo, desnecessário, no entanto, debateremos sobre ele ao longo dos próximos três capítulos a seguir.

⁴ Preferimos o termo pré-saussureano para nos referirmos aos estudos histórico comparativos e que a eles se seguiram, no século XIX. Seguimos Robins (1979), no entanto estamos cientes da polêmica em torno deste assunto (Cf. Guimarães 2018). Debatê-lo aqui, porém, fugiria do nosso objetivo. Novamente agradecemos aos professores Márcio Guimarães e Alessandro Beccari pelas preciosas observações já na banca de qualificação desta tese.

⁵ Na verdade, são vários os exemplos nesse sentido: os russos chamam a língua alemã de “немецкий язык” (nemetskiy yazyk), que se liga à raiz russa da palavra para “mudo” (aquele que não fala); em sua biografia intelectual, R. Dixon (1989), conta que, em sua viagem para a Austrália, ao dizer que iria estudar línguas locais, as pessoas diziam que ele tinha o trabalho mais fácil do mundo, porque essas línguas mal tinham gramáticas (i.e., estruturas gramaticais); e, obviamente, não podemos deixar de fora a fala comum no Brasil de que “o português é uma língua difícil”.

Passaremos, assim, a uma investigação historiográfica acerca das noções de complexidade linguística, visto que, a depender da época ou do perfil do programa científico, complexidade linguística não se refere a apenas um conceito nem chega a ter contornos tão definidos quanto gostaríamos, por exemplo, na primeira metade do século XIX não se fala propriamente em diferenças de complexidade entre as línguas, mas em perfeição de línguas ancestrais ante a paulatina degradação das línguas modernas, o que nos faz crer que havia naquele contexto uma relação de valorização entre complexidade linguística e a tal decadência que linguistas da época defendiam ser exemplificada por algumas línguas. Faremos isso tomando como base modernos estudos da historiografia da linguística, sobretudo Koerner e Asher (1995). Primeiramente exporemos alguns entendimentos acerca da complexidade linguística no século XIX, pois anterior a este período, além de não encontrarmos literatura pertinente, entendemos que a noção de diferentes complexidades entre as línguas era ponto pacífico entre estudiosos das línguas naturais⁶. É o que parece, ao menos, das leituras de Robins (1979), Mounin (1970) e Koerner e Asher (1995); manuais de história ou de historiografia da linguística, e entre estudiosos que se dedicam à complexidade linguística: Miestamo *et al.* (2008); Sampson *et al.* (2009), Joseph & Newmeyer (2012), Newmeyer & Preston (2014), dentre outros. Posteriormente, veremos como o período dos estudos linguísticos chamado estruturalista, vigente até a metade do século XX, encara este problema, e como ele é, de certa forma, transmitido ao gerativismo e mesmo para outros programas de investigação linguística atualmente em voga. Por fim, debateremos de que maneira a noção de que há diferenças de complexidade entre as línguas volta ao centro das discussões linguísticas. Espera-se, assim, que essa digressão de base historiográfica lance luz às nossas discussões seguintes acerca desse assunto que foi um truísmo na linguística durante muito tempo e que, de certo modo, ainda é.

1.2 Século XIX

Sabe-se que desde a Antiguidade, diferentes graus de complexidade entre as línguas humanas foram percebidos por diversos pensadores, estudiosos ou não das línguas (Robins, 1979). No Renascimento havia ainda a noção de que as línguas modernas europeias jamais teriam a riqueza das chamadas línguas clássicas, ou seja do grego e do latim. Riqueza, para os pensadores humanistas, dizia respeito ao grande paradigma morfológico que o grego e o latim apresentavam, notado principalmente nos sistemas verbais e nos sistemas de marcação de caso dos nomes. Essa noção, de

⁶ Na verdade Jiménez 2011 encontra uma citação de Adam Smith de 1761, enaltecendo a riqueza morfológica das línguas europeias, o que acontecia já desde a época do Renascimento, pelo menos. O assunto, entre especialistas, no entanto, passa a ser debatido efetivamente apenas no século XIX.

um modo ou de outro, perpetuou até pelo menos o século XIX, quando linguistas começaram a comparar o sânscrito a outras línguas europeias, mais precisamente, conta-se que Sir William Jones, em 1786, em Calcutá, proferiu suas suspeitas de que o sânscrito e as línguas europeias, justamente porque suas riquezas morfológicas apresentavam enorme semelhança entre si, eram aparentadas (Mounin, 1970). A tradição historiográfico-linguística conta que aí dava-se a largada aos promissores e profícuos estudos acerca das chamadas línguas indo-europeias.

Aliás, a própria fala, muito famosa, de Sir William Jones sobre o sânscrito, de 1786, é um atestado, ao mesmo tempo, (i) da provável filiação do grego e do latim (e demais línguas da Europa) ao sânscrito, (ii) à importância dada à morfologia (veja a palavras que destacamos na citação), e (iii) de que as línguas eram elencadas como mais ou menos complexas:

O sânscrito, sem levar em conta a sua antiguidade, possui uma **estrutura maravilhosa**: é mais **perfeito** que o grego, mais **rico** que o latim e mais **extraordinariamente refinado** do que ambos. Mantém, todavia, com estas duas línguas tão grande afinidade, tanto nas raízes verbais quanto nas formas gramaticais, que não é possível tratar-se do produto do acaso. É tão forte essa afinidade que qualquer filólogo que examine o sânscrito, o grego e o latim não pode deixar de acreditar que os três provieram de uma fonte comum, a qual talvez já não exista. Razão idêntica, embora menos evidente, há para supor que o gótico e o celta tiveram a mesma origem que o sânscrito (*Apud* ROBINS, 1979, p. 107).

Mais precisamente, a partir desta época inaugurou-se o que se convencionou chamar de linguística histórico-comparativa. As chamadas línguas indo-europeias, hipotetizavam então os estudiosos, provinham, de um ancestral comum, o protoindoeuropeu (PIE), que havia dominado toda a região entre Europa, norte da Índia e Oriente Próximo. Constatada no sânscrito semelhante abundância morfológica, definiu-se, portanto, como mais perfeitas as línguas que apresentavam maior inventário morfológico, e por que não, mas complexas. Aliás, o paradigma morfológico foi, junto aos mapeamentos fonético-fonológicos, o traço que ajudou a definir a família de línguas indo-europeias. August Schleicher (1821-1868), por exemplo, foi um dos grandes defensores dessa noção (Joseph & Newmeyer, 2014), tanto que foi um dos primeiros a classificar as línguas segundo critérios que hoje conhecemos como “tipológicos”. Dessa forma, observando somente a riqueza morfológica foi que se constituiu a noção de que as línguas provenientes do PIE, eram as mais bem acabadas (Mounin, 1970). A partir daí, ideias preconceituosas se expropriariam da relação entre riqueza morfológica e complexidade linguística à capacidade dos falantes.

Segundo Amsterdamska (1987, p. 38-9), na primeira geração da linguística institucionalizada, de Schlegel a Schleicher, predominou a visão de que a língua indo-europeia original, o que atualmente se chama de proto-indo-europeu, era mais orgânica do que suas descendentes e que teria entrado em

um processo de decadência na medida em que os povos que as falavam entravam na história. Essa organicidade consistia na ideia de que cada raiz indo-europeia era originalmente associada a um único significado ou função gramatical. A mudança fonética teria erodido essa organicidade, simplificando as formas das línguas descendentes. Dito de outra forma, línguas ancestrais pareciam *não* ser mais complexas seguindo critérios de hoje, mas sim mais perfeitas na correspondência raiz-significado.

No mesmo século XIX, sob a influência da filosofia romântico-nacionalista, questões sobre complexidade linguística foram delineadas, no entanto, de maneira bastante diferente da que entendemos hoje em dia. Mesmo para os parâmetros científicos da época, a noção de que havia diferenças de complexidade entre as línguas, e de que esta complexidade estava associada à qualidade de uma dada língua era já considerada, ainda que apenas por meios impressionísticos ou baseados em intuições e avaliações subjetivas, ou seja, sem a existência de efetivas comprovações minimamente verificáveis objetivamente para testar diferenças de complexidade entre línguas. De fato, alguns apontamentos linguísticos apressados levaram a impressões preconceituosas de algumas línguas e seus falantes. A preocupação, num primeiro momento, era a de comparar as línguas para que se pudesse afirmar a posição supostamente superior de um país/nação de um tempo imemorable ante outro a partir da língua falada por aquele povo, naquele território (Mounin, 1970), e, para tanto, muitas vezes uma ou outra características era tomada, quase que anedoticamente, para “provar” um dado ponto. Por exemplo, o alemão é uma língua que tem consoantes duras, ásperas porque a relação entre causa e efeito pode variar; seu povo é duro, regrado, metódico: essa consideração pode ser usada para valorizar ou minimizar o povo alemão. Assim, com vistas aos estudos posteriores, poucas das considerações descritivas feitas naquela época podem ser levadas em consideração no domínio da complexidade linguística. Isso porque, via de regra, levava-se em conta aspectos pouco relevantes para se justificar a complexidade das línguas, por exemplo, fatores externos à língua, e até mesmo minimizando o papel dos falantes sobre essas línguas supostamente menos complexas.

Esse tipo de discurso atinge inclusive estudos com algum grau de sistematicidade, como é o caso, por exemplo, do filósofo prussiano Wilhelm von Humboldt (1767-1835). Ainda que Humboldt, que teve seu pensamento linguístico reconhecido apenas décadas depois de sua morte, tivesse tocado no assunto através de noções bastante lúcidas, pois dizia ele já em 1832 que: “(...) os falares assim chamados rudimentares e bárbaros já possuem tudo o que faz parte de um uso completo (...)” (Humboldt, 2006, p. 25). Ainda na mesma obra, que leva o título *Sobre o estudo comparativo das línguas em relação com as diferentes épocas do desenvolvimento das línguas*, Humboldt admite que mesmo as nações “menos evoluídas” constroem com certa facilidade uma língua tão complexa quanto qualquer outra (Humboldt, 2006). Nesse sentido, Humboldt parece ter sido o primeiro pensador que levanta a hipótese de que a noção de complexidade é inerente a todas as línguas. Salvo isso, praticamente nenhum estudo sobre a linguagem até a primeira metade do século XIX contempla de

maneira satisfatória e aprofundada, livre de impressões pessoais e preconceitos velados ou não sobre a estrutura das línguas e a capacidade intelectual de seus falantes, a problemática da complexidade linguística. Mesmo com o possível equívoco de um anacronismo, é muito tentador considerar que o truísmo de então – de que língua claramente diferem em sua complexidade – fazia parte de uma agenda maior, aquela de identificar graus de desenvolvimento intelectual/espiritual dos povos; assim, a língua era, nesse caso, um acesso um tanto privilegiado ao grau de desenvolvimento intelectual, às vezes usado como evidência de uma conclusão já alcançada em outras bases (por exemplo, uma sociedade de caçadores-coletores, por ser, aos olhos europeus de então, muito menos complexa socialmente, também o seria em termos de sua língua).

A partir da segunda metade do século XIX, entre as principais questões que intrigavam os pesquisadores estavam: (i) como as línguas atingiram um alto grau de complexidade e (ii) como, agora, acreditavam eles, essas mesmas línguas estariam paulatinamente se degradando através, também, das mais frequentes irregularidades nos mais variados paradigmas atestados (Deutscher, 2014), e de uma franca diminuição das formas morfológicas atestadas em estágios anteriores de uma dada língua. Aliás, tais questões ultrapassam o século XIX e chegam ao século XX e mesmo ao XXI sem respostas plenamente satisfatórias, embora os esforços da linguística histórica, que procura investigar os problemas em torno das mudanças linguísticas, e de outros ramos das linguísticas descritivas⁷, já tendo nos legado importantes pistas.

Nesse sentido, termos como “irregularidade” e “paradigma” referem-se, quase sempre, ao campo de estudos da morfologia. O que os estudiosos notavam era um padrão que pode ser, muito *grosso modo*, assim ilustrado: o latim tinha uma morfologia verbal e nominal muito mais rica do que as línguas românicas; ora, como a morfologia é o que importa quando se trata de complexidade, isso significa que as línguas românicas *perderam* em complexidade se comparadas ao estágio anterior, que era o latim. Os filólogos de então notaram esse mesmo padrão para as línguas germânicas – quanto mais se volta no tempo, mais complexa (leia-se, morfológicamente rica) eram as línguas. E por que parar no latim ou no protogermânico? Muitos chegaram à conclusão de que o protoindo-europeu devia ser mais complexo (morfológicamente rico) do que todas as línguas que dele descenderam. Uma evidência de que a simplificação (perda de complexidade) estava ocorrendo há milênios. Tal conclusão, em si mesma, não tem tanto impacto, mas quando ela é combinada com a estrutura da sociedade, o resultado é um tanto incômodo. A ideia parece ser de que quanto mais complexa a sociedade, mais complexa a língua. Mas as línguas da Europa mostram exatamente o

⁷ Haja vista obras como “Language Change: Progress Or Decay?” de Jean Aitchinson, editado em 1981, cujo próprio título flagrantemente levanta a questão premente desde o século XIX, bem como outros, tais como “O Desenrolar da Linguagem” de Guy Deutscher, de 2010, traduzido para o português em 2014.

contrário: quanto mais o tempo passa, mais complexas ficam as sociedades, mas as línguas, por sua vez, ficam menos complexas. Uma aparente contradição à qual voltaremos logo mais.

E, como dissemos, os estudos de então tinham um grande apreço pela riqueza de formas (a morfologia) que uma determinada língua apresentava, e identificavam uma língua complexa (portanto, uma língua boa, importante, prestigiada) com uma morfologia rica; isso se deveu ao papel cultural do latim e do grego clássico, conforme brevemente discutimos acima. Recapitulando, o problema que se apresentava era o de que, por exemplo, as “línguas filhas” do latim perderam uma grande parte dessa riqueza morfológica, o que era sentido como uma decadência sem volta. O mesmo vale, *mutatis mutandis*, para outras famílias de línguas indo-europeias então estudadas, constituindo, para esses pesquisadores, o paradoxo ilustrado acima – quanto mais o tempo passa, pior (i.e., com menos morfologia) ficavam as línguas das civilizações, ou seja, o avanço das civilizações era um retrocesso em suas línguas. Dessa forma, tornava-se sedutora uma tipologia que classificasse línguas com mais ou menos morfologia, associando-as a serem mais ou menos “corroídas” pelo tempo.

August Schleicher, para resolver tal impasse, elaborou uma tese cômoda na época. Para ele, as “nações”, antes do advento da História (ou seja, antes da escrita), se preocuparam em elaborar e sustentar fortemente suas línguas, para depois começarem efetivamente a construir suas histórias, deixando, como consequência, a manutenção de sua língua em segundo plano. Deutscher (2014: 11), na citação abaixo, traz um resumo das ideias de Schleicher:

Assim, como quaisquer outros organismos vivos, têm as línguas um período inicial de crescimento, seguido por um período de decaimento. E o ponto de viragem entre esses dois períodos está exatamente na “aurora da história”. Como Schleicher explica, “precisamente o fato de que encontrar uma linguagem já totalmente construída pela primeira aurora da História fornece a prova de que a construção da linguagem de fato acontece antes da História. “História e construção de linguagem são duas atividades opostas do espírito humano”. Em tempos pré-históricos, de acordo com Schleicher, uma nação estava ocupada construindo sua linguagem, e “apenas quando uma nação aperfeiçoa sua linguagem, é quando ela faz a sua entrada na história”. Mas ao entrar no palco da história uma mudança radical de direção tem que ocorrer, uma vez que a partir de agora a energia de um povo é em fazer a História de vez. Portanto, uma vez a História, “aquela inimiga da linguagem”, fica em curso, a língua começa a desmoronar e entrar em decadência. (DEUTSCHER, 2014: 111)

Antes da história, ou seja, antes da escrita, os povos se preocupavam de um modo mais ou menos consciente, em manter as estruturas linguísticas; mas com o advento da história, veio também a eclosão da civilização, e então os povos passaram a investir na história, deixando a língua de lado.

Embora dono de noções basilares na época, tais como a ideia de que as línguas não são entidades estáticas e de que há níveis linguísticos observáveis, Humboldt também se utilizou da

filosofia romântica para enaltecer a língua alemã. Dado isso, o terreno está fértil para interpretações enviesadas, o que posteriormente convencionou-se chamar de “relativismo linguístico”⁸, cujos frutos geraram controversas que poderiam facilmente ter evidenciado preconceitos, não somente linguísticos. Tais noções podem ter desencadeado verdadeira aversão aos estudos da complexidade linguística (Kilarski, 2014). Visto que a associação entre riqueza linguística (ao menos do paradigma morfológico) e capacidade de seus falantes já era bastante difundida, posteriormente a simples associação entre a cultura de um dado povo, sua língua e sua capacidade, seria um passo, infelizmente perigoso, pois ameaça a “supremacia” europeia e evocava solução como a de Schleicher.

É interessante lembrarmos ainda que foi o mesmo Schleicher que procurou associar as línguas naturais à vida em seu paradigma evolucionário. Para ele, portanto, quase da mesma forma que um ser vivo, uma língua nasce, se desenvolve – provavelmente fica mais complexa – e morre. Posteriormente, Schleicher em seu *Darwinism Tested by the Science of Language* de 1869 na edição inglesa, associa os parâmetros de estudos linguísticos aos métodos da então nascente biologia evolucionista⁹. Tal associação reforçaria ainda mais a ideia de que as línguas funcionariam como organismos vivos. Assim, aliada à enorme influência de *A Origem das Espécies* (1859), a ideia de que as línguas “evoluem” de um estado mais complexo para um menos complexo, por alguma espécie de degradação promovida pelo homem, encontraria terreno fértil no contexto daquela época (Koerner e Asher, 1995).

Com o avanço do paradigma histórico-comparativo, e de conclusões incômodas como a que vimos acima, a questão da complexidade se dissipa. Lembremos que ocorreram no século XIX avanços consideráveis nos estudos linguísticos, tais como a comparação “genética” entre as línguas, bem como um entendimento mais aprofundado de como as línguas “mudam” com o tempo. É essa uma razão para que a noção histórica das línguas ganhe importância neste período. Para além da filosofia romântica, para qual cada nação tinha seu valor intelectual e cultural particular e mensurável, incluída aqui a língua dessas nações, havia a ilusão de que com o passar dos séculos, as línguas iam se “degradando” (Deutscher, 2014). Novamente, ao se observar, por exemplo, a erosão do sistema de casos latinos, que suas línguas filhas não herdaram, parecia claramente que a história das línguas derrocava para uma indelével degradação. Decorre então que, com a passagem do tempo, as línguas

⁸ Dá-se o nome de Relativismo Linguístico a uma corrente teórica – ainda que não claramente definida nesse termo – que acredita fortemente, entre outras coisas, que influências externas, como cultura e organização social de um dado povo, poderiam moldar o caráter das línguas. Tais discussões surgem na Europa no Século XIX e eclodem nos Estados Unidos, no início do século XX, sobretudo com Whorf (Robins 1979). Atualmente, essa discussão é revista. Gonçalves (2008) promove ampla discussão sobre Relativismo Linguístico, à luz de atuais parâmetros filosóficos e linguísticos para mostrar até que ponto o Relativismo Linguístico é ainda válido. Deutscher (2009) é um exemplo de trabalho recente que também se dedica a essa questão.

⁹ Embora seja pertinente falar que o método tomado como parâmetro pela linguística do século XIX tenha sido o da Biologia – veja-se, por exemplo, que o termo “morfologia” vem da biologia –, a associação entre Darwin e Schleicher, no que tange às comparações entre línguas e formas de vida, não são absolutas, mas sim frutos do *clima de opinião*. Um importante estudo de Koerner (1996) versa precisamente sobre isso, de forma bastante esclarecedora.

que outrora eram perfeitas, paulatinamente iriam perdendo sua complexidade, ficando conseqüentemente mais pobres e simplificadas. Nesse sentido, o PIE então, segundo essa lógica, só poderia ser uma língua perfeita, cujas irregularidades e idiossincrasias inexistiam. Aliada a isso, a noção schleicheriana de que as sociedades, em seu estado pré-histórico criaram a complexidade das línguas, todo um pensamento sobre este assunto notadamente se encerra sem maiores questionamentos, sobretudo no que tange à estrutura das línguas ou mesmo algum tipo de métrica de complexidade entre línguas, como podemos ver com a passagem abaixo:

Esse conceito de língua como organismo romanticamente inspirado foi combinado com uma crença no valor da língua original incorrupta. Não são as línguas contemporâneas decaídas que revelam a força vital orgânica por trás da organização linguística, mas as manifestações mais remotas da língua, que estão mais próximas do “ancestral” original (AMSTERDAMSKA 1987, p. 39).

Por outro lado, uma das questões deixadas em aberto por August Schleicher era a de que: como as línguas podem estar se degradando durante os séculos, mas ao mesmo tempo continuarem tão complexas? Ou mesmo simplesmente continuarem existindo e não desaparecem como resultado desse efeito milenar de simplificação gerado pela degradação? (cf., Deutscher, 2014). De modo geral, a percepção do próprio Schleicher de que o inglês era um alemão “degradado”, de certa forma, fazia sentido dentro de seu conjunto de crenças e ferramental descritivo, mas contrastava com a complexidade crescente da sociedade inglesa. Hoje em dia, no entanto, estamos em melhores condições de tentar resolver este mistério, cuja tensão entre complexidade aparente, custo e eficiência comunicativa sejam as respostas mais satisfatórias (Culicover, 2013).

Apresentadas alguns dos percursos historiográficos, evidências e usos sobre e da complexidade linguística até o século XIX, passaremos a investigar o paradigma estruturalista do século XX. A questão da complexidade linguística ainda não se resolve de maneira satisfatória nesta época, conforme veremos a seguir.

1.3 Estruturalismo – Primeira Metade do Século XX

Nas primeiras décadas do século XX, sobretudo depois de estabelecida a linguística estruturalista, principalmente aquela praticada nos Estados Unidos, a questão da complexidade linguística toma outras proporções, talvez a principal delas seja a ética. Ora, ao associar língua e cultura, conforme vimos anteriormente, além de termos como um dos resultados o chamado

“relativismo linguístico”, que, aliás, seria prontamente abandonado posteriormente¹⁰, havia também o perigo de associar menor complexidade linguística ao menor desempenho cultural. Há duas razões para tanto: (i) a influência positivista, que coloca como tarefa para os linguistas agora algo muito mais descritivo, objetivo e empírico e muito menos especulativo; e (ii) a forte influência das ideias linguísticas do antropólogo Franz Boas.

Sendo assim, o estruturalismo norte-americano, tanto por questões éticas quanto por questões teórico-metodológicas, propõe que todas as línguas do mundo sejam equivalentes com relação à sua complexidade. Uma das razões para tal guinada é, muito possivelmente, o fato de esses estudiosos estarem se ocupando em descrever línguas da América do Norte, que, como se sabe, possuem estruturas completamente diferentes daquelas estudadas exaustivamente na Europa, que, por mais diferentes que fossem entre si, pertenciam quase que exclusivamente à família indo-europeia¹¹. Os estudiosos rapidamente constataram a complexidade das línguas ameríndias ao descrevê-las, bem como as diferenças culturais entre os europeus e os nativos.

O próprio Humboldt (Robins 1979) já havia notado o espantoso nível de complexidade das línguas ameríndias, e mesmo antes dos estruturalistas norte-americanos as observarem. Aliadas a essas descrições, as quais renderam quatro volumosas obras descritivas¹², Franz Boas acreditava que as culturas chamadas primitivas, nada deviam – seja em seus costumes, sistemas de crenças, religião, alimentação, parentesco etc. – às chamadas culturas letradas (Boas, 2004). Junto a isso, Boas e sua equipe de estudiosos, poucos dos quais especializados em descrição linguística, oriundos da antropologia, aliavam, sem dificuldades, estudos linguísticos e culturais, logo, associando língua à cultura. Ora, se não há diferenças culturais significativas entre povos, sejam eles nômades ou cidadãos, e constatada a notável complexidade estrutural e gramatical advindas das descrições de línguas ameríndias, logo, não se pode dizer seguramente que existem línguas menos ou mais complexas do que outras – e muito menos fazer qualquer relação plausível entre desenvolvimento tecnológica e complexidade gramatical. Parece que é precisamente neste ponto que se inicia o truísmo da equicomplexidade linguística, noção que será repetidamente afirmada entre linguistas, visto que o estruturalismo norte-americano, a partir da década de 1930, se tornaria bastante influente mundo afora, inclusive no Brasil com Joaquim Mattoso Câmara Júnior (1943). No entanto, tal afirmação é mais antiga. Algumas décadas antes, o respeitado linguista norte-americano, Dwight Whitney (1827-1894), citado inclusive por Saussure em seu *Cours*, afirma já em 1875 em seu *The Life and Growth of Language* a possibilidade todas as línguas possuírem o mesmo grau de complexidade:

¹⁰ Para ser então estudado em bases neurológicas (cf. Deutscher, 2014).

¹¹ Não é por acaso que em vários dos capítulos do famoso livro “Language: An Introduction to the Study of Speech” (1921), de Edward Sapir, o autor critica várias das categorias gramaticais usadas desde a antiguidade clássica como sendo insuficientes para descrever as línguas nativas da América do Norte.

¹² Trata-se do *Handbook of American Indian Language*, editado em quatro volumes entre 1911 a 1938.

Algumas atitudes se alteraram. A erudição do século XIX afirmava firmemente que alguns dialetos, idiomas, culturas e espécies são inerentemente melhores ou "superiores" do que outros. Isso produziu uma tipologia na forma de uma escala evolutiva linear, de modo que, se duas línguas diferiam de qualquer modo essencial, só poderia ser que uma estivesse menos evoluída do que a outra. Whitney muitas vezes parece dividido entre essa atitude e o igualitarismo desinteressado da ciência natural, e o leitor moderno fica satisfeito com a frequência com que o último assume a vantagem. (HOCKETT 1979: xvi *apud* JOSEPH; NEWMEYER, 2012)¹³.

O capítulo final entre estruturalismo e complexidade linguística talvez se dê com o linguista norte-americano Charles Hockett, importante descritivista e representante do estruturalismo americano. Em sua obra *A Course in Modern Linguistic*, de 1958, Hockett discorre explicitamente sobre a questão de complexidade linguística ao longo de toda a obra, reafirmando noções inspiradas possivelmente na antropologia de que todas as línguas naturais seriam equivalentes com relação à sua complexidade, e suas diferenças estruturais seriam explicadas através de uma espécie de “lei das compensações”. São fartos os exemplos de diferentes graus de complexidade entre línguas na obra de Hockett (pp.: 95, 101, 193, etc.), sobretudo no que tange à fonética e a à fonologia, mesmo assim Hockett se esforça para mostrar que deve haver uma forma de compensação entre as línguas por ele examinadas. Por exemplo, se tomarmos o inglês e a maior parte das línguas românicas com relação à morfologia verbal, a pouca complexidade do inglês nesse aspecto, se comparado às línguas românicas, seria compensada com complexidade em outros níveis descritos dessa língua. Embora não comprovada empiricamente tal questão, Hockett tinha convicção de que em todas as línguas havia uma espécie de equiparação:

Uma medida objetiva seria difícil, mas o impressionante é que parece que a complexidade gramatical de qualquer língua, tanto a morfologia quanto a sintaxe, é aproximadamente a mesma. Isso não é surpreendente, uma vez que todas as línguas têm trabalhos igualmente complexos a fazer, e o que não é feito morfologicamente, tem que ser feito sintaticamente. O Fox, por exemplo, com uma morfologia mais complexa do que o inglês, deve ter uma sintaxe um pouco mais simples. E este é o caso (HOCKETT 1958 [1954]: 180-181)¹⁴.

¹³ No original: “Some attitudes have altered. Nineteenth-century scholarship firmly held that some dialects, languages, cultures, and species are inherently better or ‘higher’ than others. That yielded a typology in the form of a linear evolutionary scale, so that if two languages differed in any essential way it could only be that one was less fully evolved than the other. Whitney often seems torn between that attitude and the disinterested egalitarianism of natural science, and the modern reader is pleased at the frequency with which the latter gets the upper hand.” (Hockett 1979: xvi).

¹⁴ No original: “Objective measurement is difficult, hut impressionistically it would seem that the total grammatical complexity of any language. Counting both morphology and syntax, is about the same as that of any other. This is not surpqsmsg, since all languages have about equally complex jobs to do, and what is not done morphologically has to be done syntactically. Fox, with a more complex morphology than English, thus ought to have a somewhat simpler syntax; and this is the case” (HOCKETT 1958 [1954]: 180-181).

Pouco antes de Hockett, Rulon Wells, em 1956, já afirmava algo parecido (Newmeyer & Preston 2014). A semelhança entre todos esses estudiosos, no entanto, acerca da complexidade linguística é o fato de que todos são meramente especulativos em suas afirmações, por mais bem intencionados que sejam. Talvez menos especulativa seja a gramática gerativa que, no entanto, compactua com os estruturalistas a tese de que todas as línguas sejam igualmente complexas. Embora o próprio Chomsky nunca tenha, ao menos até agora, se pronunciado definitivamente sobre o assunto.

De um jeito ou de outro, o eurocentrismo, ao longo do século XIX, sempre procurou alternativas para colocar suas línguas num plano superior a outras ao redor do mundo, seja aparentando-as às línguas europeias, caso do sânscrito; seja vinculando-as historicamente aos europeus, caso do hebraico (Robins 1979). Seja como for, o que a linguística do século XIX conseguiu enxergar foi que o impressionante paradigma das línguas indo-europeias se mostra, indubitavelmente, como uma estrutura altamente complexa e especializada. No entanto, esta estrutura, por si só, não revela toda a complexidade que há por trás de toda e qualquer língua natural. Talvez, observando um aspecto aqui outro ali das línguas ao redor do mundo, linguistas tenham sido levados a pensar que há complexidade em todas as línguas, possivelmente de maneiras equivalentes, embora flagrantemente diferentes (Boas, 2004). Dito de outra forma, através de um dado código ético, já observado por antropólogos no início do século XX ao estudarem as formas de aglutinações sociais, seria pertinente – e talvez mesmo cômodo – para os linguistas afirmar que não havia diferenças substanciais no que concerne à complexidade entre as línguas (Hockett, 1958). É interessante notar que, mesmo com esse passo, a nosso ver, mais interessante, em direção ao que a complexidade linguística pode ser, ainda foi feita aqui a manobra fundamental de dissociar complexidade linguística da estrutura ou complexidade social e da capacidade intelectual dos falantes. Tal dissociação seria feita mais tarde.

Apresentados os argumentos estruturalistas, da primeira metade do século XX, na seção seguinte procuraremos mostrar de que forma as discussões sobre complexidade linguística foram relegadas no programa gerativista.

1.4 Pós-estruturalismo – Segunda Metade do Século XX

Na segunda metade do século XX, com o advento e a predominância da gramática gerativa transformacional como paradigma dominante nos estudos linguísticos, as hipóteses sobre o que seja uma língua natural passaram a ter um teor mais biológico do que social, e a tarefa final da linguística passou a ser a de encontrar “princípios” comuns em todas as línguas do mundo, através de sua estrutura mental, que os gerativistas têm por certo como sendo inata. Assim, o que se vê é uma

equiparação das línguas do mundo em termos de sua estrutura mental e inata. Agora a questão passa a ser a forma como as línguas são utilizadas pelos seus falantes, e menos o seu conteúdo, sobretudo aquele que a faz comum entre seus falantes. Ao contrário, para Chomsky e seus seguidores agora o que é relevante é o que todas as línguas têm em comum entre si. E isso só pode estar em dada estrutura que deve ser comum a toda língua natural, e que é convencionalmente chamada de gramática universal. Entenda-se aí “gramática”, então, como sendo esta estrutura mental, inata, comum a todos os falantes de qualquer língua natural.

Na verdade, é bem antiga a noção de que existem, em toda e qualquer língua natural, alguma estrutura ou mecanismo que seja comum a todas elas. A busca por, digamos, universais linguísticos, remonta aos estudos de Roger Bacon, já no século XIII, perpassando pelos pensadores de Port-Royal, no século XVII (Beccari, Leal 2013), quando eles compararam minuciosamente o francês ao latim e ao hebraico, justamente em busca de alguma estrutura que lhes parecesse comum. Segundo o próprio Chomsky (1966), mesmo o já citado Humboldt tinha a crença de que todas as línguas humanas tinham algo em comum: “(...) há nas línguas um ponto de completa organização a partir do qual a construção orgânica, a forma firme não mais se altera. Nelas, no entanto, enquanto produtos vivos do espírito, pode progredir até o infinito a formação mais refinada, dentro das fronteiras estabelecidas.” (HEIDERMAN e WEININGER 2006, p. 23).

Nesse sentido, é lícito pensar que, se há tal gramática, e se ela é universal, ou seja, se ela faz parte de um aparato mental, com bases biológicas, comum a toda espécie que tem a capacidade de usar essa linguagem, então parece óbvio que esse aparato não pode diferir de forma sensível, quantitativa ou qualitativamente, entre os seus usuários. Dito de outra forma: se há uma gramática universal, nela não pode haver diferenças, ao menos significativas¹⁵. Logo, a questão sobre uma possível existência de diferentes graus de complexidade entre as línguas, não faz o menor sentido, ou, na melhor das hipóteses, não é relevante. Num tal paradigma, não há espaço para uma discussão profunda e profícua sobre os diferentes graus de complexidade que as línguas do mundo supostamente apresentariam, dado que elas são estudadas e entendidas por meio de suas características universais; qualquer coisa de diferente é um problema menor.

Na busca por universais linguísticos, houve outro importante teórico na mesma década de 1960: Joseph Greenberg. Ele postula, a partir de observações de cerca de 30 línguas, muitas das quais consideradas então exóticas – algumas línguas africanas e outras americanas, como o guarani – que existiriam alguns traços em comum entre todas as línguas naturais. A partir desses traços dessas línguas então analisadas, Greenberg observou que determinados traços sempre levavam a outros. Ele

¹⁵ Se dizer que simplesmente não pode haver diferença alguma pode soar um tanto forte, dizer, por outro lado, que se há diferenças então elas não são significativas, nos obriga a ter uma métrica de diferenças, justamente para saber quando elas são significativas. Obviamente, ambas possibilidades têm vários problemas.

percebeu também que tais traços, assemelhados a algumas dessas línguas, poderiam levar a uma nova forma de se classificar as línguas naturais. Mais precisamente, o estudo de Greenberg, intitulado *Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements*, de 1963, lança as bases de uma verdadeira revolução nos estudos descritivos, pois possibilitou uma nova classificação entre as línguas, algo bem diverso do que fizeram os histórico-comparatistas, que vinculavam as línguas a famílias, cujo critério era a proximidade geográfica ou traços mais salientes apenas. Ironicamente, a disciplina que ele reinventou, agora com critérios estruturais e não somente genéticos, na segunda metade do século XX, a tipologia, iniciada, de certa forma, por Schleicher no século XIX, é uma das principais responsáveis pela eclosão das mais atuais discussões acerca da complexidade linguística. Pois é a partir das indagações suscitadas pelas observações e comparações exaustivas de muitas línguas ao redor do mundo é que a tipologia vislumbra, ainda que de maneira sutil, potenciais diferenças entre a complexidade das línguas.

Embora se suponha que tratar de complexidade linguística seja equivalente a observar a estrutura formal da língua, ou seja, seu aparato gramatical, foi no âmbito da sociolinguística – uma área dos estudos linguísticos em geral¹⁶ crítica aos postulados gerativistas – que o problema da diferença de complexidade entre as línguas do mundo ressurgiu nos moldes científicos de hoje. Segundo Newmeyer & Preston (2014), foi Peter Trudgill, em 1983, como sua obra *On Dialect: Social and Geographical Perspectives*, quem, pela primeira vez, levantou a hipótese de que pode haver diferentes graus de complexidade entre as línguas do mundo, e que essa diferença poderia explicar fenômenos linguístico-sociais (a complexidade linguística teria então poder heurístico). Precisamente num estudo de caso que o autor fez junto aos dialetos gregos nas regiões de Ática e Beócia, onde falantes albaneses majoritariamente usavam o grego moderno como língua de contato, Trudgill constatou aí reduções vocabulares, bem como simplificações no sistema morfofonêmico e redução de traços redundantes no sistema morfossintático do grego. Logo, Trudgill relacionou simplificação à redução, percebendo que, ainda que esses dialetos viessem ganhando estruturas de caso, mesmo que bastante simples, isso não justificaria uma “compensação” linguística, tal como previram os estruturalistas (Trudgill & Tzavaras 1977, *apud* Newmeyer & Preston, 2014). No entanto, as discussões pormenorizadas sobre o assunto viriam a surgir apenas no final do século XX.

Posteriormente, Kusters (2003), cujos estudos estruturais de verbos no quéchua são seminais para as indagações sobre complexidade linguística, conclui em suas obras que as diferenças de complexidade entre as línguas somente podem se dar no plano social, ou seja, as diferenças de complexidade entre as línguas, embora medidas em suas mais diversas estruturas, só podem transparecer na medida em que as línguas são efetivamente disseminadas entre seus mais diferentes

¹⁶ Há exceções, claro, como a “sociolinguística paramétrica”.

falantes e, sobretudo, nos seus mais diversos fins. Foi justamente isso que Kusters evidenciou ao demonstrar, através de comparações estruturais, que os verbos no quéchuá, enquanto língua franca, eram bem mais simples do que os usos em situações comunicacionais diferentes como, por exemplo, em contextos místico-religiosos em comunidades isoladas (Kusters, 2003). Ainda que Kusters tenha observado cuidadosamente a estrutura do quéchuá, e mesmo que seus estudos tenham se dado longe dos preceitos teóricos da sociolinguística, a conclusão a que ele chegou é que não há como definir complexidade, a não ser pelo viés social do uso da língua. Como veremos no capítulo pertinente, tal posição não é compartilhada por diversos linguistas, dentre os quais, gerativistas como Newmeyer & Preston (2014) e Cullicover (2012), ou funcionalistas como Dahl (2004). Diametralmente oposta à noção de que a complexidade só poderia ser visto no âmbito social do uso, McWhorter (2001), ao procurar uma definição estrutural, sincrônica e não social para as línguas crioulas, como se fazia até então, acabou contribuindo sobremaneira para as discussões sobre os diferentes graus de complexidade das línguas, sobretudo no que tange às línguas emergentes.

Ao comparar línguas emergentes a línguas já secularmente estabelecidas, McWhorter (2001) percebeu que (i) morfologia derivacional não-composicional, (ii) flexão e (iii) tonalidade lexicalmente distintiva¹⁷ não eram registrados em línguas crioulas, e também são as três características mais difíceis para se aprender em uma segunda língua, segundo argumenta o autor. Diz ainda McWhorter: “Um subconjunto de línguas crioulas apresenta menos complexidade gramatical geral do que as línguas mais antigas, em virtude do fato de terem nascido como pidgin e, portanto, despojado de quase todos os recursos desnecessários à comunicação¹⁸” (2001, p. 125). Dessa forma, ao investigar as estruturas das línguas emergentes, McWhorter acaba corroborando para derrubar o truísmo de que todas as línguas deveriam ser igualmente complexas. Sua proposta é uma das mais articuladas ao definir complexidade gramatical, e utilizá-la em uma “heurística positiva”, mas também evitando cair em concepções preconceituosas:

Se aceitarmos que o paradigma do passado de Kikongo e os verbos de movimento elaborados de forma pouco usual do eslavo são rastreáveis, em última instância, à elaboração do acaso, então também devemos aceitar que não podemos conceber logicamente um projeto de linguagem neural geneticamente selecionado nem um mecanismo determinado cognitivamente ou funcionalmente, isso asseguraria que as ocorrências ocasionais operassem com um grau exatamente igual em todas as gramáticas das línguas naturais na Terra. Em

¹⁷ McWhorter (2001) foi pioneiro em discutir, para além de especulações de ordem político-sociais, o que são línguas emergentes. Ou seja, com McWhorter (2001), possivelmente pela primeira vez, temos fatos estruturais concretos que mostram como funcionam as línguas emergentes. Mas, como não poderia deixar de ser, suas ideias foram e ainda são duramente criticadas, notadamente por S. Mufwene. Ver, por exemplo, Mufwene (2014).

¹⁸ No original: “[...] a subset of creole languages displays less overall grammatical complexity than older languages, by virtue of the fact that they were born as pidgins, and thus stripped of almost all features unnecessary to communication [...]” (McWhorter, 2001, p. 125).

outras palavras, nada na teoria linguística moderna implica que todas as línguas naturais sejam igualmente complexas, de modo que qualquer asserção ou mesmo suposição subjacente a esse efeito se qualifica como um truísmo ainda por ser testado e não como um fato estabelecido” (McWHORTER, 2001, p. 42)¹⁹.

Mais precisamente, McWhorter acredita que todas as gramáticas atingiram um certo grau de complexidade, mas mesmo isso é pouco para dizer que todas as gramáticas são igualmente complexas. Dessa forma, ele coloca a seguinte questão: como todas as gramáticas de todas as línguas desenvolveriam paralelamente, e igualmente, complexidades de diferentes tipos em diferentes subsistemas? A resposta seria: “[a quantidade de] complexidade de uma gramática resulta das operações de acréscimo aleatório ao longo do tempo, os crioulos exibem menos complexidade do que o resto das gramáticas naturais do mundo²⁰” (McWHORTER, 2001, p. 43). A relação entre idade da língua e sua complexidade, ou seja, os aparatos²¹ que as línguas ganham ao longo do tempo, é matéria essencial para os estudos acerca da complexidade linguística. Haja vista que importantes linguistas que atualmente lidam com a questão tomam como importante parâmetro a idade da língua: Dahl (2004) e Deutscher (2014), sobretudo. Já falamos desses autores esporadicamente ao longo deste capítulo e trataremos mais detidamente dessa relação entre idade da língua e complexidade nos próximos capítulos.

Alguns dos pontos que julgamos mais interessantes vistos nesta seção são: (i) a desvinculação da complexidade linguística com qualquer outro tipo de complexidade ou capacidade intelectual; (ii) a complexidade linguística usada numa “heurística positiva”; e (iii) a inexistência de uma relação entre complexidade linguística e potencial expressivo, ou seja, não interessa qual seja a complexidade linguística de uma língua A, ela pode dizer tudo o que seus falantes quiserem e tudo o que possa ser dito numa outra língua B. Talvez essas três características somadas definam os estudos atuais sobre complexidade linguística.

Após apresentadas as noções do que seja complexidade linguística nos estudos linguísticos na segunda metade do século XX, faremos, em seguida, um apanhado de quais desafios os estudos acerca da complexidade linguística têm em frente.

¹⁹ No original: “If we accept that Kikongo’s past-marker paradigm and Slavic’s unusually elaborated verbs of motion are traceable ultimately to chance elaboration, then we must also accept that we can conceive logically of neither a genetically selected neural linguistic blueprint nor a cognitively or functionally determined mechanism, which would ensure that chance occurrences operate to a precisely equal degree on every natural language grammar on earth. In other words, nothing in modern linguistic theory entails that all natural languages be equally complex, so any assertion or even underlying assumption to this effect qualifies as a truism yet to be tested rather than as an established fact” (McWhorter, 2001, p. 42).

²⁰ No original: “[...] grammar’s complexity results from the operations of random accretion over time, creoles display less complexity than the rest of the world’s natural grammars” (McWhorter, 2001, p. 43)

²¹ O que Dahl (2004) chama de “ornamentos”.

1.5 Desafios para os estudos sobre Complexidade Linguística

Conforme vimos nas últimas seções, as discussões em torno da complexidade linguística ora foram limitadas a discursos que poderiam levar a eventuais leituras preconceituosas ou mesmo racistas, como muitas vezes foi o caso no século XIX; ora foram relegadas a deduções imprecisas, sem maiores descrições linguísticas, ou menos sem maiores consequências, como é o caso da primeira metade do século XX, embora todo aparato de equicomplexidade inerente ao programa estruturalista norte-americano esteja presente nessa noção. Assim, mesmo sob o rigor descritivista do estruturalismo então vigente, a questão da complexidade linguística não foi suficientemente investigada, houve apenas deduções apressadas que, quando muito, estavam ancoradas no paradigma teórico-metodológico de uma determinada vertente da antropologia cultural (Guimarães, 2020).

De uma forma ou de outra, a imprecisão na observação dos fatos linguísticos levou a resultados equivocados. Há quem diga que a base do moderno preconceito linguístico teria surgido justamente das discussões mal formuladas e nas justificadas advindas do pensamento dos séculos XIX e XX, conforme acredita Kilarski (2014). Como vimos, isso se confirma em parte porque, de um modo ou de outro, a estrutura das línguas nunca tinha sido observada com o propósito de medir complexidade. Com as boas intenções advindas da antropologia do início do século XX, ao associar a língua de uma dada sociedade a sua cultura, adotou-se na linguística um truísmo paulatinamente repetido ao longo das décadas, em nome de uma postura ética, a qual sabemos hoje, um tanto sem sentido, que associa complexidade das línguas com a eventual capacidade mental de seus falantes. Por outro lado, no século XIX: “[...] as fontes da hipótese foram espalhadas pelo espectro intelectual, abrangendo a noção de que a linguagem encarna uma cosmovisão nacional e serve como uma “chave mágica” para a cultura, que pode ser atribuída a Johann Gottfried Herder (1744 -1803) e Wilhelm von Humboldt (1767-1835) [...]”²² (KILARSKI 2014: 163).

Seja como for, ao longo de muitas décadas, é compreensível que a ideia de equicomplexidade das línguas tenha sido bastante sedutora. Afinal, mesmo as fortes noções encetadas por Charles Darwin (1809-1882) em sua *Origem das Espécies* (1859), e mais atualmente atestadas pelo mapeamento do DNA humano, dão conta de que todos os seres humanos são rigorosamente iguais, portanto, teriam as mesmas capacidades. Não é disso, no entanto, que trata os estudos sobre complexidade linguística.

²² No original: “[...] the sources of the hypothesis were spread across the intellectual spectrum, encompassing the notion that language embodies a national worldview and serves as a “magic key” to culture, which can be traced to Johann Gottfried Herder (1744–1803) and Wilhelm von Humboldt (1767–1835) [...]” (Kilarski, 2012, p. 163).

Tanto McWhorter (2001) quanto Dahl (2004) afirmam cabalmente que, ao que se evidencia, qualquer língua natural tem todo maquinário necessário para desenvolver o mais complexo dos sistemas para ser expresso em seus mais diversos níveis: seja fonológico, morfológico, semântico, etc. Todo este entendimento, no entanto, não era evidente nos séculos anteriores:

A dissociação explícita da complexidade linguística da complexidade cultural removeu um obstáculo à ideia de que todos os seres humanos eram inerentemente iguais. Uma vez que a dissociação foi aceita, seria inútil buscar a justificação da desigualdade humana com base na suposta desigualdade linguística, removendo assim um obstáculo à conclusão de que as línguas não podem nem devem ser classificadas em termos de sua complexidade (JOSEPH; NEWMAYER 2012: 351)²³.

Posto os argumentos contra a ideia de se estudar complexidade linguística, e algumas questões que puseram novamente as discussões em torno deste assunto em voga, em seguida discutiremos algumas das questões em torno das quais a noção de complexidade das línguas pode se amparar para ajudar a responder questões adjacentes e este problema, tais como:

- (i) diferenças de complexidade implicam em diferenças expressivas?;
- (ii) a questão da tradutibilidade entre línguas é pertinente para se observar ou mesmo para se medir complexidade?;
- (iii) os chamados universais linguísticos contribuem para a questão da complexidade?;
- (iv) há uma quebra de paradigma em torno das ciências da linguagem com o recente advento das discussões em torno da complexidade linguística?;
- (v) há de fato como medir complexidade linguística?

Iniciaremos tais discussões aqui, neste capítulo, para podemos começar a pensar e avaliar respostas razoáveis para essas e outras questões adjacentes.

À questão da expressividade (questão i) está relacionada mais fortemente a questões de tradutibilidade, e de como as línguas conseguem trocar, entre si, significados que muitas vezes pertencem a aspectos bem específicos de tais línguas, tais como termos fortemente ligados à sua cultura, do que propriamente à questão da complexidade linguística. McCready (2014) lida com tais questões de uma forma que julgamos interessante, pois ele dissocia a questão da expressividade à da complexidade, chegando mesmo a propor algumas métricas formalizadas para se medir a

²³ No original: “The explicit dissociation of linguistic complexity from cultural complexity removed an obstacle to the idea that all humans were inherently equal. Once the dissociation was accepted, it would be useless to seek justification for human inequality on the basis of putative linguistic inequality, thereby removing one obstacle to the conclusion that languages cannot and should not be ranked in terms of their complexity” (Joseph & Newmeyer, 2012, p. 163).

expressividade das línguas. Segundo o autor, então, há diferenças de expressividade entre as línguas, tanto quanto há diferenças de complexidade, mas, no entanto, trata-se de questões diferentes; possivelmente porque há variados níveis de complexidade, inclusive a expressiva e a semântica. Especificamente, McCready (2014) entende que expressividade²⁴ é um conceito que se aproxima de tradutibilidade, no sentido de que dada uma sentença ou expressão X de uma dada língua A, tal expressão ou sentença, deve encontrar um equivalente semântico de X em dada língua B. Via de regra, se não há tal correspondência, diz o linguista, é possível, então, que tenhamos diferenças de expressividade entre as línguas A e B. Ou seja, para McCready (2014), a expressividade das línguas naturais podem ser medidas. Seja como for, a questão da tradução (questão ii) parece ser pertinente em ambos os casos: expressividade e complexidade. À complexidade tem interessado questões como, por exemplo, a proficiência de falantes de segunda língua para se medir complexidade entre duas línguas (Leal 2018), ou mesmo se há possibilidade ou não de traduções *ipses litteris*.

Já a relação entre os universais linguísticos e a complexidade tem trazido acaloradas discussões (questão iii), tais como a própria quebra do paradigma linguístico (questão iv) posto modernamente, desde o início dos anos 1950 com o gerativismo. Tal quebra se daria devido ao fato de que parece haver uma linha de raciocínio na qual os universais linguísticos, devido à sua natureza intrínseca, se eles de fato existem, então não pode haver diferenças de complexidade entre as línguas. Ora, desde Roger Bacon, no século XIII, supõem-se, conforme já discutimos na seção 2.1, que em todas as línguas há traços inerentes que as tornam, por assim dizer, variações de um mesmo fenômeno natural e humano. Tal questão é replicada por séculos, em movimentos historicamente pendulares (Beccari, 2013), até praticamente os dias de hoje. O que tem sido postulado nos últimos tempos é o fato de que nas línguas há diferenças em variados níveis, e que estes, sim, podem diferir, talvez, no que tange sua complexidade – algo que talvez se aproxime de Complexidade Relativa.

Um exemplo dessa quebra de paradigma vem justamente afrontar um princípio postulado pelo gerativismo, o de que a recursividade sentencial seria um traço universal entre todas as línguas naturais (Chomsky, 1965). Em Everret (2008 *apud* Miestamo *et al.*), é levantada a hipótese de que pode existir uma língua sem tal traço: trata-se do pirahã, uma língua amazônica utilizada por uma tribo isolada. Algo semelhante pensa Deutscher (2000) acerca do acadiano, língua semítica falada na antiga Mesopotâmia, e que parece, segundo seus registros apenas escritos, também não apresentar o fenômeno da recursividade, ou, pelo menos, sérias restrições ao considerarmos estratégias de

²⁴ Não nos aprofundaremos em tal debate aqui, mas é importante notar que McCready (2014) está interessado na possibilidade de tradução de termos como “fuck”, do inglês, ou “porra”, do português brasileiro, que não são estritamente falando descritivos e cujos usos estão associados a avaliações subjetivas dos falantes. A ideia, muito grosso modo, é pensar na possibilidade de tradução adequada do inventário de termos como esses em diferentes línguas. Por exemplo, em inglês, é possível termos (a) “Fuck, I forget the keys”, (b) “I fucking forget the keys” e (c) “I forget the fucking keys”, ao passo que no português brasileiro teríamos, mais naturalmente, (a’) “Porra, eu esqueci as chaves” e (c’) “Eu esqueci a porra das chaves”, mas não temos uma alternativa óbvia para traduzir (b). Seria essa uma limitação expressiva? Em resumo, são essas questões provocativas que compõem o artigo de McCready (2014).

subordinação e de encaixe sentencial. Tanto num caso, como noutro, para além das acaloradas polémicas, parece evidente que houve uma espécie de “evolução” de um estágio menos complexo para um estágio mais complexo, como o que é constatada hoje na maior parte das línguas.²⁵ Além disso, linguistas especializados no assunto, tais como Miestamo *et al.* (2008), não acham possível estudar complexidade sem o aparato da tipologia. Para ele, a descrição cada vez mais atenta de mais e mais línguas ao redor do globo, apenas corrobora a existência de diferenças de complexidade entre as línguas naturais.

A respeito da questão ‘v’ posta acima, faremos um capítulo à parte. Isso porque parece uma das questões mais prementes em torno dos estudos sobre a complexidade linguística, tanto por fatores externos, tais como questões éticas que envolvam diferentes valorações às línguas, quanto por fatores internos, tais como a própria possibilidade de, descritivamente, se medir a complexidade inerente nas línguas. Seja como for, já existem ferramentas para tais medidas, amparadas em bases teóricas sólidas, e críticas a elas já podem ser feitas em relação à sua eficácia e precisão e não com relação à sua própria possibilidade de existência. Aliás, apresentaremos os principais parâmetros de medidas, bem como suas ferramentas teóricas, no capítulo quatro. Esta é a razão, inclusive, de não nos alongarmos no assunto aqui.

Para além desses desafios, talvez um dos maiores seja o de colocar os estudos sobre complexidade linguística em seu devido lugar dentro da comunidade científica. Essa parece ser uma tarefa que cabe à filosofia da linguística, que está espalhada em dois capítulos (3 e 4) que ajudam a explicar não somente o fenômeno, mas as motivações e bases teórico-metodológicas dos estudos em torno da complexidade linguística. Prosseguiremos no programa proposto, depois da devida conclusão deste capítulo.

1.6 Conclusão

Se, por um lado, temos o eurocentrismo do século XIX, para o qual havia diferenças de complexidade entre as línguas naturais, baseadas nos diferentes níveis culturais dos povos com os quais os europeus travavam contatos, por outro, temos a bem intencionada noção antropológica do século XX, que insiste em afirmar que todas as línguas naturais são igualmente complexas, numa espécie de “lei das compensações”. Diante disso, podemos, então, afirmar que encontramos novamente um exemplo que reforça a conclusão de Thomas Kuhn de que forças sociais, éticas ou

²⁵ Tomamos tais exemplos (Everett e Deutscher) como meramente ilustrativos, pois em ambos os casos não há evidências suficientes para atestar um estágio menos complexo das línguas naturais, ao contrário do que se tem com pidgins, por exemplo. Seja como for, afirmações e crenças desses dois linguistas vêm sendo paulatinamente rebatidas.

mesmo histórico-tradicionais, obviamente não científicas, podem anuviar as verdadeiras nuances observáveis nos fenômenos naturais (Kuhn, 1965).

Na linguística, isso também apareceu claramente. Sobretudo na última década, em que questões como complexidade linguística e evolução das línguas estão sendo fortemente suscitadas, mas que outrora eram relegadas ao lugar de mero mito. Mesmo com um olhar crítico, podemos notar que outrora, as especulações filosófico-linguísticas e, sobretudo sociais, foram um tanto infelizes ao associar deliberadamente à língua a cultura e a capacidade cognitiva de seus falantes. Não tanto pela associação em si, mas pela forma com que foi abordada. Seja como for, o *princípio de contextualização*, advindo das teorias encetadas pela historiografia da linguística (Koerner e Asher, 1995; Swiggers, 2009), nos fez avaliar todas as noções do que seja complexidade linguística de um ponto de vista privilegiado. Privilegiado por procurar se colocar fora do alcance de críticas temporais, que não devem inebriar nosso entendimento de como tais e tais programas teóricos encaravam o fenômeno da complexidade linguística. Foi justamente o que procuramos fazer neste capítulo.

Seja como for, procuramos cumprir nossa tarefa historiográfica de mapear, através de uma leitura histórica e cronológica, o entendimento e mesmo a mudança sobre o conteúdo e o papel da noção de complexidade linguística. Tal tarefa foi embasada nos princípios descritos no capítulo 1 desta tese, ou seja, nos princípios da historiografia linguística, princípios dos quais nos propomos a seguir. No entanto, como bem lembra Swiggers (2009), não há como fazer historiografia da linguística sem fazer uma filosofia da mesma matéria. Dito de outra forma: a tarefa da historiografia estaria relegada à mera crônica, caso não se investigasse também, e minuciosamente, as fontes teóricas dos programas linguísticos. Por isso, junto à historiografia linguística, ao longo desta tese, nos valem também dos preceitos da filosofia da linguística.

Capítulo 2 – Definindo Complexidade

2.1 Introdução

No capítulo anterior mostramos o percurso histórico da ideia de complexidade linguística e alguns de seus usos e consequências, através de investigação historiográfica, sobretudo através de modernos preceitos da historiografia linguística (Leal 2014, Beccari 2013, Koerner e Asher 1995, Swiggers 2009), procurando deixar implícito que o que se chamava outrora complexidade linguística não tem, nem de longe, o mesmo sentido que atribuímos hoje a esse conceito. Basta lembrarmos da classificação prototipológica de Schleicher, para quem, no século XIX, as línguas de morfologia rica eram também as mais complexas, e isso era um reflexo de sua organização social, que também deveria ser complexa. Já no século passado, dadas diversas polêmicas, convencionou-se tratar complexidade linguística como um tabu que não deveria ser mais tocado, como uma questão agora resolvida, cujo passado era inclusive motivo de um certo embaraço. Seja como for, convém agora debatermos o que se entende atualmente por complexidade linguística, em sua reabilitação e reavaliação contemporâneas. Antes, no entanto, procuraremos mostrar quais as definições utilizadas hoje em dia sobre a própria ideia de complexidade, sobretudo no campo da filosofia.

Complexidade tem sido um termo bastante debatido, em diversas áreas do conhecimento, especialmente na filosofia, principalmente nos últimos 50 anos, e é amplamente identificado como algo que se relaciona com teorias da complexidade, teorias do caos, bem como em outras áreas, como a cibernética e as correntes filosóficas pós-modernas. Na verdade, há ainda certo *frison* em torno do que seja complexidade, do mesmo modo que termos como *não-determinístico* ou *não-linear* tiveram em épocas recentes. De qualquer forma, dados os longos e produtivos debates realizados principalmente a partir de nomes como Gell-Mann (1995, 1996), Rascher (1998 *apud* Miestamo *et al.* 2008) e Edmonds (1999), apenas para citar alguns dos pensadores cujas ideias tiveram algum impacto recente²⁶, iremos estabelecer neste capítulo um parâmetro minimamente fiável do que seja complexidade para, então depois, tentarmos definir complexidade linguística. Nenhuma das duas tarefas – quais sejam: definir complexidade e definir complexidade linguística – são simples. No entanto, sendo um assunto há muito debatido, procuraremos expor aqui quais melhores conceitos de complexidade que se têm estabelecidos, pelas quais poderemos nos guiar. Faremos o mesmo com a noção geral do que seja complexidade linguística. Ainda que os linguistas não tenham chegado a um consenso sobre como medi-la, o que seja complexidade linguística já é mais ou menos estabelecido,

²⁶ Para uma introdução mais acessível sobre várias dos temas que envolvem complexidade, cf. Gleick (2013).

antes mesmo de qualquer investigação séria mais aprofundada acerca do assunto. Obviamente, para podermos medir alguma coisa é preciso, antes de mais nada, defini-la, e assim é um fato relevante reconhecer que já haja um certo consenso, como veremos, sobre o que é complexidade linguística.

Assim, nossa tarefa aqui é procurar definir/estabelecer o que seja complexidade, pelo menos “operacionalmente”, para os objetivos da presente tese. Depois, procuraremos uma definição consistente do que seja complexidade linguística. Serão nesses termos, então, que este capítulo está organizado. Além disso, é importante frisar que o que faremos aqui, bem como o que faremos no capítulo seguinte, é estabelecer alguns parâmetros do que seja complexidade segundo fontes de especialistas. Dessa forma, este capítulo é construído numa pesquisa bibliográfica, que, no fim das contas, é posta de forma bastante interpretativa. Seja como for, sabe-se que as opiniões, ainda que sutilmente impostas, sempre permeiam a superfície textual. Mesmo assim, procuramos ser o mais imparcial possível.

2.2 O que a filosofia entende por complexidade

Nesta seção, segundo então a pesquisa bibliográfica que realizamos, procuraremos definir de modo bastante conciso o que vem a ser complexidade, ou seja, de que maneira cientistas e filósofos entendem o que seja o fenômeno da complexidade que vem mesmo acarretando uma possível mudança de paradigma nos modelos científicos.

A noção de complexidade ganhou tanta força nas últimas décadas, que foi mesmo elevada a uma ciência autônoma, e passa a ser não apenas um termo antônimo à simplicidade. Aliás, o físico teórico Murray Gell-Mann (1994 e 1995), que é um dos importantes estudiosos da complexidade, refuta veementemente a noção de que complexidade é o contrário de simplicidade. Mais efetivamente, tais noções, conforme veremos, têm mais relação com variados pontos de vista e, segundo Edmonds (1999), também têm relação com o ambiente. A seguir, esperamos deslindar a noção de complexidade e seus eventuais desdobramentos. Diferentemente do capítulo 1, não faremos aqui regressos históricos, apenas discorreremos sobre o conceito como é entendido hoje.

2.2.1 Tentando definir um termo nada simples

Para Heylighen, Cilliers e Gershenson (2007), estudar complexidade é um ponto de vista praticamente inverso ao modo newtoniano de se pensar a ciência. De fato, identificamos no “mundo newtoniano” uma visão segundo a qual as coisas no universo apresentam uma complexidade aparente,

assim, poderíamos explicar o Universo com leis simples (e elegantes): “A ciência da complexidade baseia-se em uma nova maneira de pensar que contrasta fortemente com a filosofia subjacente à ciência newtoniana, baseada no reducionismo, determinismo e conhecimento objetivo²⁷” (Heylighen, Cilliers e Gershenson, 2007: 117). Na verdade, para esses autores, mas não apenas para eles, as respostas a muitas questões trazidas pela natureza, seriam respondidas pelas chamadas ciências da complexidade²⁸. E algo semelhante acontece ao pensarmos no meio social: “[a] complexidade é talvez a característica essencial da nossa presente sociedade²⁹” (Heylighen, Cilliers e Gershenson, 2007: 117). Parece, então, que a propriedade de “ser complexo”, além de ser inerente a vários sistemas da natureza, é um de seus traços mais relevantes e, possivelmente, definidor. Seja como for, a noção de complexidade é definida, hoje em dia, mediante diferentes traços, os quais procuraremos explorar a seguir. Talvez o principal deles seja o saber.

A definição de *saber* tem proximidade à da complexidade, no sentido justamente do que se observa – leia-se, mede – para procurar descrever aspectos da natureza. Ou seja, para se enxergar complexidade, é lícito que se pense em medir a quantidade de traços ou aspectos que definem um dado objeto observacional. Dado isso, é lícito também que, para podermos observar quais e quantos são os traços que descrevem dado objeto, é preciso *saber* quais eles são. Para Gell-Mann (1995), *saber* pode ser visto, a princípio, de duas formas: um intuitivo e um outro adquirido *a posteriori*. Metaforicamente, poderíamos dizer que essas duas maneiras de *saber* podem ser exemplificados de acordo com saberes linguísticos, tais como os chamados saber *metalinguístico* e saber *epilinguístico* (Swiggers, 2009). Enquanto este diz respeito ao saber inato ou intuitivo, o outro só é obtido através do ensino regular e formal. Ou seja, todo falante sabe utilizar sentenças com orações encaixadas, mas nem todos conseguem identificá-las num discurso ou em textos, se se trata de uma oração encaixada, ou subordinada, ou coordenada.

Voltemos a nada simples noção do que se entende atualmente por complexidade, segundo o mesmo Gell-Mann (1995):

Algumas das quantidades, como a complexidade computacional, são medidas de tempo (ou espaço). Seus estudiosos estão preocupados com o tempo que levaria (ou a quantidade de capacidade que seria necessária), no mínimo, para que um computador universal padrão execute uma tarefa específica. A própria complexidade computacional está relacionada ao

²⁷ No original: “The science of complexity is based on a new way of thinking that stands in sharp contrast to the philosophy underlying Newtonian science, which is based on reductionism, determinism, and objective knowledge” (Heylighen, Cilliers e Gershenson 2002, p. 117).

²⁸ Optamos por utilizar o termo “ciências da complexidade” para fins expositivos, pois há, na verdade, diferentes desdobramentos do que sejam os programas científicos cuja base é a complexidade. Tais distinções e detalhes, no entanto, não são necessários para nossa discussão aqui.

²⁹ No original: “Complexity is perhaps the most essential characteristic of our present society” (Heylighen, Cilliers e Gershenson 2002).

menor tempo (ou número de etapas) necessário para realizar uma determinada computação (GELL-MANN 1995: 1)³⁰.

Junto a este conceito mesurável, há o conceito que diz respeito à percepção do que seja complexidade, para o mesmo Gell-Mann (1995): “Assim, algo quase inteiramente aleatório, com praticamente nenhuma regularidade, teria uma complexidade efetiva quase zero” (Gell-Mann 1995: 2)³¹. O mesmo aconteceria para algo absolutamente regrado, com total regularidade, ou seja, também teria complexidade zero. Isso porque – e parece ser consenso entre aqueles que se inclinam a estudar complexidade – é bastante intuitivo associar complexidade a aleatoriedade e, como veremos nos casos das línguas naturais, a idiossincrasias ou “exceções”. Dessa forma, a complexidade seria, então, inerente a diversos sistemas na natureza, não propriamente em suas características menos regulares, que tendem ao caos; nem mesmo nas suas características mais regulares – caos e regularidade absolutos têm, ambos, complexidade zero. Dito de outra forma, para Gell-Mann (1995), a complexidade estaria situada entre essas duas “forças” da natureza: o caos e a perfeição. Novamente, podemos perceber que complexidade é, antes de tudo, um ponto de vista, pois depende do tipo de “saber” e, principalmente, do “lugar” de onde se observa dado objeto de estudo.

Não por acaso, essa ideia se aproxima daquela que mencionamos brevemente acima de que a complexidade linguística tem a ver, por exemplo, com um balanço entre redundância e informações repetidas (regularidade) e exceções (caos). Dahl (2004) oferece um exemplo bastante interessante ao mobilizar uma parte da teoria da informação para medir complexidade (ver capítulo 4). Ao procurar explicar as premissas da teoria de Kolmogorov, aplicada a dados informacionais, Dahl (2004) fornece os seguintes exemplos numéricos:

(a) 185185185

(b) 187351873518735

(c) 098754870897

Dadas tais sequências, o autor explica que (a) é menos complexo que (b) dado seu comprimento, ou seja, o número de algarismos repetidos, mas também dada a possibilidade de tal repetição ser “resumida”³² em apenas três algarismos que se repetem. Diferentemente de (b), que ao apresentar

³⁰ Original: “Some of the quantities, like computational complexity, are time (or space) measures. They are concerned with how long it would take (or how much capacity would be needed), at a minimum, for a standard universal computer to perform a particular task. Computational complexity itself is related to the least time (or number of steps) needed to carry out a certain computation” (Gell-Mann, 1995, p. 1).

³¹ No original: “Thus something almost entirely random, with practically no regularities, would have effective complexity near zero” (Gell-Mann, 1995, p. 2).

³² Utilizamos aqui o termo “resumo”, mas Dahl (2004) originalmente fala em “zip”, um termo comum utilizado ao lidarmos com informação computacional, que nada mais são do que “compactadores” de informação. Nesse sentido, ao

mais Algarismos, apresentaria também um “resumo” mais longo desses Algarismos caso repetidos, o exemplo em (c), diz Dahl, embora muito complexo, não apresenta um padrão, o que resultaria numa impossibilidade de medi-lo, dado sua aleatoriedade³³.

São também inerentes à definição de complexidade, as noções de *emergência* (emergentismo) e holismo. Para Heylighen, Cilliers e Gershenson (2007), emergente é uma propriedade cuja inteireza não pode ser dissociada em partes. Por exemplo, não dá para dissociar ritmo, harmonia e melodia de propriedades que são da música. Já holismo, que remete a filósofos do início do século XX – como Bergson, Teilhard, Whitehead e em particular Smuts (1926) (Heylighen, Cilliers e Gershenson 2007: 5) –, e pode ser definido como a tendência de um todo ser maior (e mais importante) que a simples soma de suas partes. Dessa forma, presume-se, a partir dos mesmos Heylighen, Cilliers e Gershenson (2007) que complexidade é uma característica que só pode ser vista ou medida a partir de um dado sistema como um todo. Veremos, adiante, que tal noção não é compactuada por todos os estudiosos da complexidade, sobretudo entre linguistas. Segundo, então, esses autores, diferentemente dos newtonianos, o sistema utilizado para observar a natureza (a ciência) não é necessariamente regido por leis determinísticas de sistemas fechados, mas aberto. Isso quer dizer que sistemas abertos são dependentes de seus ambientes. Segundo essa caracterização, não se pode olhar e descrever um dado sistema (ou super sistema) sem levar em conta todos os seus subsistemas e a interação entre eles. Mesmo assim, considera-se o sistema como um todo. Caso contrário, a descrição desse dado super sistema não será válido aos parâmetros das teorias da complexidade. O que explicaria estes fenômenos é o que estudiosos chamam de teorias dos sistemas gerais:

O método científico tradicional, baseado na análise, no isolamento e na coleta de informações completas sobre um fenômeno, é incapaz de lidar com essas complexas interdependências. A emergente ciência da complexidade oferece a promessa de uma metodologia alternativa que seria capaz de resolver tais problemas (HEYLIGHEN, CILLIERS, GERSHENSON 2007: 1)³⁴.

compactar (ou resumir) uma informação, o que os softwares que “zipam” arquivos fazem, nada mais é do que abreviar informações ou termos repetidos, por essa razão, quanto mais termos repetidos, não importa seu tamanho, menos complexo é o objeto.

³³ Saber que algo é, de fato, aleatório é uma questão filosófico-matemática extremamente complexa. Dado que número podem ter tamanho arbitrários e infinitos, como saber que uma dada sequência é realmente aleatória? Por exemplo, o valor de π , em 2019, já foi calculado para além de 31,4 bilhões de dígitos, sem nenhum sinal de repetição ou padrão. Mas como se trata de um número irracional, infinito, pode ser que, por exemplo, a cada 33 bilhões de dígitos comece um padrão de repetição. Portanto, mesmo diante de algo dessa magnitude, não podemos afirmar definitivamente que o valor de π seja aleatória. De qualquer forma, não entramos a fundo nesta questão e admitiremos, como é prática comum e aceita, que a aleatoriedade de fato existe na natureza.

³⁴ No original: “The traditional scientific method, which is based on analysis, isolation and the gathering of complete information about a phenomenon, is incapable of dealing with such complex interdependencies. The emerging science of complexity [1–3] offers the promise of an alternative methodology that would be able to tackle such problems” (Heylighen, Cilliers e Gershenson 2007: 1).

Já para Edmonds (1999), a complexidade “[...] é a propriedade de um modelo que dificulta a formulação de seu comportamento geral em uma determinada linguagem, mesmo quando recebidas informações razoavelmente completas sobre seus componentes atômicos e suas inter-relações.” (Edmonds 1999: 72)³⁵. Ou seja, Edmonds argumenta que “A complexidade representa o fosso entre o conhecimento dos componentes e o conhecimento do comportamento global (ou emergente)” (Edmonds 1999: 72)³⁶. Além disso, Edmonds (1999) lembra que existem, no mínimo, dois tipos de complexidade, uma oriunda de um *modelo objetivo*, outra oriunda de um *modelo geral*. Entenda-se por objetivo o modelo de complexidade cujas partes são tomadas *per se*, e o chamado modelo geral, por sua vez, contempla todo um sistema, seja ele qual for. Mas é importante notar que tal definição não é unânime.

O mesmo Edmonds (1999) ainda lembra que seu conceito de complexidade só pode ser aplicado a modelos de descrição científica e não a sistemas naturais, como as línguas. Seja como for, esquematicamente ele elenca a possibilidade de se trabalhar com complexidade dadas as seguintes características:

- aplica-se a modelos e não a sistemas naturais;
- a complexidade se distingue da falta ou carência de informação;
- é relativo à linguagem de modelagem na qual é expresso;
- é relativo à identificação de componentes e comportamento geral;
- complexidade é uma característica global de um modelo;
- há diferentes tipos de complexidades e diferentes tipos de dificuldade, ou seja, ambos não são necessariamente relacionados;
- complexidade representa a lacuna entre o conhecimento componente e o conhecimento do comportamento global (ou emergente). (Edmonds 1999: 72).

Finalmente, Edmonds (1999) parece convencido de uma firme definição do que seja a complexidade (ou os diferentes tipos de complexidade, para além do simples binômio: geral *versus* objetivo), segundo ele, há alguns poucos parâmetros que podem fazê-lo:

- a complexidade dos dados: a dificuldade de codificar um modelo de dados de forma compacta, dada uma linguagem de codificação;

³⁵ No original: “[...] is that property of a model which makes it difficult to formulate its overall behaviour in a given language, even when given reasonably complete information about its atomic components and their inter-relations” (Edmonds, 1999, p. 72).

³⁶ No original: “Complexity represents the gap between component knowledge and knowledge of global (or emergent) behaviour” (Edmonds, 1999, p. 72).

- a complexidade do modelo informal (mental): a dificuldade em fazer uma previsão informal a partir do modelo, dadas as condições hipotéticas;
- a complexidade do uso do modelo formal para prever aspectos do sistema em estudo, dadas certas condições (Edmonds 1999: 74).

Vimos com isso, que os termos entendidos como complexidade atualmente são diversos, mas convergem a estudos palpáveis que têm mesmo a pretensão de modificar o chamado “paradigma científico newtoniano”, cujas principais características são o determinismo³⁷ e a linearidade. Além disso, complexidade é fortemente associada a pontos de vista (pois além de depender das diferentes concepções de diferentes estudiosos, depende do tipo e lugar em que o objeto em estudo é visto) e influências do ambiente (a relação de um sistema com o meio e com seus subsistemas), além de se associar a arbitrariedade e, mesmo, é tomada, a partir de investigações mais acuradas, como parâmetro essencial para explicar diversos fenômenos da natureza que apresentam, aliás, em suas definições, propriedades que lhe são inerentemente complexas.

Dadas as definições mais salientes do que seja complexidade, na próxima seção tentaremos, preliminar e paulatinamente, antes de adentrarmos a questão da língua, associar teorias da complexidade e como esta corrente enxerga as línguas naturais.

2.2.2 Sistemas adaptativos complexos

Há, nos estudos acerca da complexidade, uma série de conceitos cujas observações acuradas vêm confirmando a eficiência das metodologias empregadas pelas teorias da complexidade. Dentre as propostas de observação da natureza, estão, ainda que sutilmente colocadas, questões acerca de fenômenos sociais, leia-se: ciências humanas:

[...] o comportamento de um indivíduo é controlado não apenas pela neurofisiologia do cérebro, mas pelas regras da sociedade à qual ele pertence. Devido às dependências entre os componentes, as propriedades desses componentes não podem mais variar independentemente: elas precisam obedecer a certos relacionamentos. Isso torna irrelevantes muitas das propriedades individuais, enquanto muda o foco para o estado de seu

³⁷ Não confundir com “determinismo” conforme usado em “determinismo geográfico”, por exemplo. Aqui, “determinismo” se refere ao fato de que, sabendo o estado inicial de um dado sistema mais um conjunto de leis, é possível determinar seu estado posterior.

relacionamento, que agora definirá um novo tipo de propriedade “emergente” (HEYLIGHEN, CILLIERS, GERSHENSON 2002: 122)³⁸.

E, apesar de algumas investidas negativas, há possibilidade de uso para os estudos linguísticos das teorias e conceitos da Complexidade:

Até o momento, as aplicações da teoria da complexidade às ciências sociais não têm sido muito produtivas. Pode haver várias razões para isso, mas pode-se argumentar que muitos teóricos sociais foram introduzidos na complexidade por meio do trabalho realizado por cientistas de complexidade “dura” [...] (Gell-Mann 1994: 16).

Na verdade, tal cenário vem mudando drasticamente: em linguística têm-se utilizado no ensino de línguas estrangeiras os preceitos das teorias da complexidade³⁹.

Dentre esses modelos dentro dos parâmetros “humanísticos” há uma definição das chamadas teorias da complexidade para as línguas naturais, os chamados *sistemas adaptativos complexos*. Tais sistemas, dos quais a língua natural seria o exemplo mais evidente e usual, são entendidos como parte da natureza biológica, cujas leis são regidas por parâmetros arbitrários, e como parte da natureza social, porque pode sofrer modificações através do aprendizado de um dado código:

A característica comum de todos estes processos é que em cada um deles um sistema adaptativo complexo adquire informação sobre seu meio ambiente e sobre sua própria interação com este meio ambiente, identificando regularidades naquela informação, condensando estas regularidades em um tipo de “esquema” ou modelo, e atuando no mundo real com base neste esquema. Em cada caso, há vários esquemas competindo, e os resultados da ação sobre o mundo real retroalimentam o esquema e influenciam a competição entre eles (GELL-MANN 1996: 38).

Segundo, então, estudiosos da complexidade (sobretudo Gell-Mann 1995 e 1996, e Rescher 1998), a língua seria um sistema adaptativo complexo, cujos pré-dispositivos inatos não são suficientes para sua eclosão. Junto a estes, deve-se considerar o ambiente e as condições sociais, ainda que estas também sejam herdadas. Um tipo de herança, aliás, diferente da genética, a que Dawkins

³⁸ No original: “[...] the behaviour of an individual is controlled not only by the neurophysiology of her brain, but by the rules of the society to which she belongs” (Heylighen, Cilliers e Gershenson 2002: 122).

³⁹ Não traremos aqui quais são os modelos da/na Linguística que se utilizam das Teorias da Complexidade em sua metodologia, pelo fato de que i) fugiria ao nosso escopo, que é o de falar pontualmente sobre o fenômeno da complexidade linguística de ii) porque há já diversas discussões sobre a adaptação de termos de um ramo da ciência para outro. Sokal e Brickmont (2006) e Borges Neto (2009) mostram como essa questão pode ser desastrosa. De qualquer forma, não é nosso escopo, ainda que façamos menção à essa expropriação de termos de outros ramos da ciência ao longo desta tese.

(1976) chamou de *meme*⁴⁰. Ou seja, ao invés de uma herança genética, os sistemas adaptativos complexos se valem de heranças sociais. Dada esta característica, há uma tendência geral desses sistemas em se modificar mais rapidamente, isso porque os sistemas adaptativos complexos têm como principal característica absorver, através do aprendizado junto à sociedade que os circunda, o conhecimento acumulado por essa sociedade. Assim, diferentemente de outros sistemas, o sistema adaptativo complexo é capaz de “aprender”⁴¹ e, com isso, se modificar. Justamente como fazem as línguas naturais: “Observamos como o sistema adaptativo complexo percebe regularidades no fluxo de informações, escolhendo-as a partir de características consideradas incidentais ou arbitrárias, e as condensando em um esquema sujeito a variações” (Gell-Mann 1996: 27). Talvez, dada sua natureza arbitrária, este tipo de sistema pareça tão adequado a fins linguísticos. Aliás, é importante lembrar que vários linguistas (como Culicover (2013), por exemplo), vêm determinando traços mais salientes de complexidade justamente em idiosincrasias dos sistemas, sobretudo da morfologia e da sintaxe.

De modo geral, a ideia de um sistema adaptativo complexo é bem atraente, isso porque além de trazer uma definição mais precisa do que seja complexidade, parece definir também mais claramente o que são línguas, sobretudo dentro do modelo geral de uma teoria da complexidade. De qualquer forma, por alguma razão, linguistas, como regra geral, não vêm adotando nenhum modelo das chamadas teorias da complexidade, ao menos não na linguística teórica. O que é adotada pela linguística são conceitos gerais, provindos das teorias da complexidade. Talvez isso seja feito por duas razões: i) os conceitos do que seja complexidade estão bem definidos e são totalmente compatíveis com o tipo de complexidade operado pelas línguas naturais; ii) a linguística já possui diversos modelos teóricos bastante eficientes, inclusive para medir complexidade linguística (cuja eficiência ainda está para ser averiguada).

Na próxima seção, iremos discutir o que se entende por complexidade linguística, segundo, principalmente, linguistas.

2.3 Complexidade Linguística

Ainda que não seja atualmente evidente a diferença entre correntes formalistas e correntes funcionalistas nos estudos sobre complexidade linguística, Newmeyer & Preston (2014) e Culicover (2013) insistem em debater modelos formais que podem contribuir para esta discussão. Mas mesmo

⁴⁰ *Meme* é um termo cunhado por Richard Dawkins em seu famoso livro “O Gene Egoísta” de 1976. Sua definição está ligada à herança culturalmente transmitida pelos humanos, como ideias, sons, desenhos e mesmo a língua natural. Nos últimos anos tal termo ganhou outros significados, mas o qual suscitamos aqui é este criado por Dawkins.

⁴¹ Obviamente, “aprender” aqui não é usado no mesmo sentido que aquele aplicado a humanos, mas tem a ver com adaptação a pressões ambientes e sua transmissão via e para seus usuários.

estes, digamos, mais afeitos a correntes formalistas, já vêm admitindo que o conjunto de fenômenos responsável pela complexidade nas línguas naturais tem mais relações externas aos sistemas, pois que estas, segundo ainda Culicover (2013), apresentam forte tendência à simplificação da língua. Possivelmente dado o caráter pragmático das línguas naturais, ou seja, aquilo que diz respeito à língua enquanto processo comunicativo, que numa tensão entre forma e conteúdo possam estar causando – há séculos, diga-se – a complexidade inerente das línguas naturais. De um modo geral, acredita-se que as diversas ferramentas especializadas em comunicação e expressão das línguas possam ter tornado as línguas cada vez mais complexas, como argumentam McWhorter (2001) e Dahl, (2004). Seja como for, vimos que a questão de diferença de complexidade entre as línguas eclode das correntes funcionalistas e, ainda hoje, os métodos mais destacados para medir complexidade linguística, privilegiam a informação e sua transmissão.

Aliás, é importante que se registre aqui que entendemos por correntes formalistas da linguística aquelas que estariam mais preocupadas com as razões internas à língua, de caráter sistemático que dizem respeito à sua estrutura, o que linguistas como Faraco (2005) chama de “linguística imanente”. Já por correntes funcionalistas, entendemos aquelas que se detêm mais aos aspectos pragmáticos e comunicativos, portanto, externos à língua (Borges Neto 2009): “Podemos distinguir, *grosso modo*, duas grandes concepções de linguagem: uma (a mais forte em linguística) a considera como um objeto autônomo; outra a considera como um objeto intrinsecamente ligado a realidade social, histórica e cultural de seus falantes” (Faraco 2005, p. 103). Ditas essas primeiras palavras sobre a complexidade linguística, vamos, na próxima seção, tentar entender melhor o que os linguistas definem por complexidade linguística.

2.3.1 A complexidade linguística e suas definições de hoje

Apesar de ser uma noção quase intuitiva que todas as línguas naturais são, em algum sentido, complexas, não há consenso, tanto na linguística quanto na filosofia, do que seja “complexidade linguística”.

Edmonds (1999) argumenta, por exemplo, que o tipo de complexidade linguística é diferente da complexidade de outros fenômenos naturais, primeiramente porque possui em sua estrutura interna uma provável disposição biológica inata, como já apontado por Chomsky (1965), mas também porque a língua possui um certo tipo de relação externa, ou seja, sua complexidade (ou ausência dela) depende em larga medida de fatores sociais, visto que a função comunicativa da língua é primordial e uma de suas características inerentes, e também, possivelmente, o que lhe garante, em parte, a noção de complexidade. Além disso, o mesmo Edmonds (1999) lembra que o tipo de complexidade

linguística é estrutural e biologicamente aleatório, e cultural e socialmente regrado. Isso quer dizer que as prováveis mudanças que levaram à complexidade linguística se deram inicialmente na estrutura de forma aleatória, mas sua disseminação se deu de modo regrado em âmbito social. A exemplo disso, a eclosão das construções de tópicos, conforme veremos mais detidamente no capítulo 4, que surgem possivelmente por alguma característica expressiva, e hoje estão amplamente disseminadas na fala do PB.

Nos próximos parágrafos desta seção, veremos algumas definições diferentes do que vem a ser “complexidade linguística”, tendo como foco as ideias de Dahl (2004), Kusters (2013), Culicover (2013) e Newmeyer & Preston (2014).

Para Dahl (2004), uma boa definição de complexidade se aproxima de noção de *custo*, que, por sua vez, se refere ao esforço dispensado para a execução de uma dada tarefa. Por exemplo, uma língua que possui marcação redundante de plural no SN, teria mais “custo” do que uma língua que não possui tal redundância para executar uma tarefa semelhante. Por outro lado, Dahl também está preocupado com o nível de “informatividade” transmitido pelas línguas. Nesse sentido, para ele, complexidade diz respeito à relação entre os processos gastos pela língua para realizar suas tarefas gramaticais e o que isso gera de custo de processamento para que as informações sejam transmitidas por essa língua. Dessa forma, o contrário de “custo”, ou seja, a “economia”, pode significar ineficiência comunicacional na língua. Ilustrativamente, se usarmos o mesmo exemplo, poderíamos dizer que a marca redundante de indicação de plural no SN, como faz o português padrão, garantiria mais clareza de informação a toda cadeia sentencial, apesar de ser menos econômica do que uma marcação não redundante.

Por exemplo, uma sequência como ‘os meninos altos chegaram’ traz quatro vezes a informação do plural, como mostra o sublinhado; assim, se o ouvinte não ouve a informação de um dado item por conta, por exemplo, de ruídos no ambiente, a marcação repetida em outros itens garantirá a informação do plural. A redundância, apesar de pouco econômica, garantiria a preservação da informação. Por sua vez, a sequência ‘os menino alto chegou’, traz igualmente a informação do plural, porém, agora, marcada num único item. Ou seja, ela é muito mais econômica, pois faz a mesma coisa que a outra sequência com muito menos “custo”. Contudo, a falta de redundância pode fazer com que a informação do plural se perca mais facilmente – por exemplo, suponha que o ouvinte não tenha escutado, por algum motivo que seja, o determinante. Trata-se, na verdade, de um balanço ideal entre “sinal” e “ruído”, conforme exemplificado acima. A ideia é definir complexidade gramatical justamente na dinâmica custo-benefício dos sistemas linguísticos, sem levar em conta, obviamente, julgamentos normativos que simplesmente não vêm ao caso. Ou seja, concordância e informação devem ser entendidas num sentido estritamente estrutural e computacional.

O autor deixa claro, ainda, que complexidade não deve ser vista como sinônimo de dificuldade, mas como uma espécie de “medida de quantidade de informação”:

As línguas humanas são sistemas complexos não porque precisam expressar pensamentos complexos. A complexidade que estou interessando aqui, no entanto, não tem preocupações com o que se pode dizer com uma língua – o poder expressivo da língua – como costume dizer. Complexidade é vista aqui não como sinônimo de “dificuldade”, mas como uma propriedade objetiva do sistema – uma medida de quantidade de informação necessária para descrevê-lo ou reconstruí-lo (DAHL 2004: 12)⁴².

Para Kusters (2003), complexidade linguística diz respeito à maneira como determinados falantes julgam dadas línguas, e assim, segundo esse autor, uma boa forma de medir complexidade é a aquisição (ou aprendizado), sobretudo de segunda língua, com foco nas estruturas linguísticas. Desse modo, as estruturas mais complexas de uma língua residem justamente nas áreas da gramática em que são observadas as maiores dificuldades dos falantes aloglotas. Note que, neste caso, complexidade linguística é algo que deve ser relativizado aos falantes em situação de aprendizagem de língua estrangeira, ou, pelo menos, diante de sua atitude a línguas estrangeiras em comparação com sua(s) língua(s) nativa(s).

Para Culicover (2013) e para Newmeyer & Preston (2014), por sua vez, complexidade diz respeito ao número de regras pelas quais a gramática de uma dada língua pode ser descrita. Um pouco diferente da visão de Dahl (2004) ou de Kusters (2003), essa abordagem parece tender para uma linguística mais formalista, ou seja, Culicover (2013) e Newmeyer & Preston (2014) não relacionam complexidade à língua em uso (ou seu aprendizado) ou à transmissão de informação ou conteúdo linguístico, mas única e exclusivamente a análise estrutural imanentista das línguas.

Independente da definição de complexidade, é ponto pacífico entre todos os estudiosos que não há uma relação entre falta de complexidade e baixa capacidade cognitiva ou cultural de seus falantes, muito menos entre complexidade e potencial expressivo (i.e., o que pode ser dito numa dada língua). Por outro lado, sabe-se também que não foram constatadas línguas surgidas de processos naturais que possuíssem um nível muito elementar de complexidade, ou seja, uma língua de complexidade zero⁴³. Seja como for, é impossível negar o fato de que a cognição humana responsável

⁴² Original: “Human languages are complex systems, not least because they need to express complex thoughts. The complexity I am interested in here, however, does not so much concern what one can say in a language — the expressive power of the language — as how it is said. Complexity is here seen, not as synonymous with “difficulty” but as an objective property of a system—a measure of the amount of information needed to describe or reconstruct it”.

⁴³ Nos termos de Gell-Mann (1995), uma língua ou sem exceções ou sem regras. É interessante notar aqui casos que envolvem línguas que foram artificialmente criadas, como o Esperanto. Em sua criação, o Esperanto não possuía exceções, contudo, essa foi a língua materna de alguns indivíduos que, durante o curso de sua vida, inovaram as regras do Esperanto e criaram exceções. Pode-se argumentar que o fato de o Esperanto ser a língua nativa de alguns e de ele se

pelo processamento da língua possui um “maquinário” suficiente para dar conta de toda e qualquer língua natural, independente de quão complexo seja qualquer um dos seus sistemas, como atestam as línguas atualmente faladas no planeta. Dito de outra forma, mesmo que atestado que uma dada língua é, em termos absolutos, menos complexa que outra – sem ainda termos total clareza do que isso significa –, o processamento cognitivo de seus falantes é tão eficiente quanto o de outro falante que se utiliza de uma língua eventualmente mais complexa, em termos relativos ou absolutos.

Passemos, na sequência, na seção 2.2.2, a investigar melhor a ideia de tipos diferentes de complexidade linguística, com o objeto de chegar a uma definição mais razoável dessa noção, ou, pelo menos, entendermos as dificuldades envolvidas nessa tarefa. Investigaremos também, na seção 2.2.3, a ideia de maturação da complexidade linguística, que relaciona de maneira diretamente proporcional a idade de uma língua à sua complexidade.

2.3.2 Complexidade Absoluta e Complexidade Relativa

Para lidar com a difícil tarefa de definir complexidade linguística, vários autores (cf., Dahl 2004, Kusters 2003, Miestamo *et al.* 2008, Deutscher 2009 e Culicover 2013) propõem distinguir entre *complexidade linguística absoluta* e *complexidade linguística relativa*. De modo geral, tais conceitos são entendidos da seguinte maneira: a complexidade absoluta seria definida como uma propriedade objetiva do sistema linguístico (seja ele qual for); e a complexidade relativa seria vista como uma espécie de “custo”, ou mesmo dificuldade vista pelos falantes de uma dada língua para “operá-la”.

Tanto para Kusters (2003) quanto para Miestamo *et al.* (2008), complexidade absoluta diz respeito ao modo de ver ou mesmo de medir a complexidade de uma língua em todo seu sistema. Assim, para medir ou enxergar a complexidade de uma língua, deve-se observar seu sistema por inteiro, e o que definiria sua complexidade então seria a quantidade de subsistemas atrelada a esse sistema “geral” que, de maneira abstrata, descreve a língua. Desse modo, seria mais complexa uma língua que apresenta maior grau de complexidade nos seus subsistemas. Para os mesmos autores, a complexidade relativa seria dada pela medida de cada um desses subsistemas em separado (obviamente definindo, então, quais e quantos são eles), para daí se obter alguma resposta pertinente a respeito da complexidade de uma dada língua.

Uma solução possível para este problema é medir apenas a complexidade relativa, que seria, num certo sentido, mais acessível, pois corresponderia em parte aos conhecidos níveis de análise

comportar como um sistema complexo o torna uma língua natural (quanto tratamos dos falantes que o adquiram; cf. Okrent, 2009).

linguística, assim, poderia medir em separado os níveis morfológicos, sintáticos etc., mas depois avaliar, ante todo sistema, se há mais ou menos complexidade em dada língua⁴⁴. De fato, para vários autores, medir a complexidade absoluta de uma língua seria uma tarefa impossível. Além disso, pragmaticamente, dado o nível de descrição disponível hoje sobre diversas línguas pelo mundo, a análise dos mais diferentes níveis das línguas é mais produtiva⁴⁵.

2.3.3 Maturação de complexidade linguística e a idade das línguas

Para além das definições de complexidade absoluta e relativa, para Dahl (2004), o que torna uma língua complexa é seu tempo de “maturação”. Nesse sentido, Dahl se alinha a McWhorter (2005) para quem línguas jovens (pidgins e crioulos, por exemplo) não têm o mesmo nível de ornamento, e, portanto, de complexidade, do que línguas mais antigas.

Ornamentos são estratégias de marcação formal que a princípio não fariam falta para o bom trânsito na semântica das línguas. No entanto, na medida em que uma língua envelhece, ela vai ganhando ornamentos – o que antes não era uma marcação formal obrigatória passa a ser. É impossível não ver aqui ecos da teoria da gramaticalização – e isso faz sentido, pois a gramaticalização é um dos mais importantes processos que tornam certas marcações obrigatórias. Poderíamos ainda, segundo as noções de Dahl (2004), dizer que os ornamentos são traços que as línguas ganham com o tempo, e que podem servir como reforço para, por exemplo, transmitir informações no uso comum da língua. Basicamente, os ornamentos surgem para que a transmissão de informação seja garantida. Mais uma vez, a marca redundante de plural em SNs seria um ornamento. A ‘superespecificação’ de morfemas verbais, marcas morfológicas redundantes de tempo, por exemplo, nos termos de McWhorter (2005), seriam outros exemplos de ornamento.

Nesse sentido, nas línguas “velhas” haveria uma espécie de “manutenção” de aparatos complexos, como redundâncias. A redundância marcada em alguns sistemas linguísticos, por exemplo, ao passo que torna dada língua computacionalmente menos eficiente, a faz também mais

⁴⁴ Um pressuposto que merece ser questionado aqui e que, na prática, nunca é, é o seguinte: os níveis de análise linguística tradicionais que, em princípio, definiriam os subsistemas de uma língua, foram postulados com o objeto de descrever uma dada língua, levando em conta certas teorias, baseadas em aceções sobre como as línguas se estruturariam, e também que certos níveis são diferentemente estruturados (a famosa ideia de “dupla articulação” das línguas naturais). Contudo, nada garante que, para medirmos complexidade relativa, esses mesmos níveis sejam de fato relevantes. Pode ser que precisemos, por exemplo, postular um novo nível de análise que seja a interface entre a prosódia e a sintaxe, ou níveis de análise que enfatizem a interface informativa-interacional das línguas naturais. A crítica aqui, em resumo, é: não podemos ser ingênuos ao falar de subsistemas ou níveis de análise linguísticas ao pensarmos em complexidade linguística, simplesmente usando noções e definições prévias como se elas servissem para essa nova tarefa. É preciso repensar o que é mais relevante e faz mais sentido considerando uma dada noção de complexidade linguística.

⁴⁵ Östen Dahl (2004), embora engajado com Miestamo (2008) e Kusters (2008) quanto à posição teórica, discorda da noção de complexidade absoluta. Para Dahl, a complexidade absoluta se resume simplesmente a: quanto mais complexo é cada um dos subsistemas de uma língua, mais complexa ela é.

complexa na comparação com línguas que não possuem tal traço em alguns de seus sistemas. Por outro lado, segundo Dahl (2004), mesmo a redundância, vista sob a perspectiva da teoria de comunicação, é um traço necessário para garantir “informatividade”, ou seja, a preservação da informação, como vimos. Nesse sentido, a complexidade parece ser não somente um aspecto inerente, mas necessário às línguas.

Em suma, o que parece tornar uma língua mais complexa é seu tempo de “vida”. Segundo Dahl (2004), a língua só ganha seus ornamentos (que a tornam mais complexa) com o tempo. Tais ornamentos podem ser vistos tanto no nível discursivo (retórico), quanto no nível gramatical; mais especificamente, são nuances que podem ser vistas como traços linguísticos sem utilidade comunicativa aparente: seja a marcação redundante de plural, sejam pronomes neutros, partículas expressivas de realce ou mesmo marcações de caso que podem se tornar “fósseis”, ou “expressões cristalizadas”, mais do que são, de fato, recursos da língua propriamente. Esse autor, na verdade, propõe noções significativas para que se possa enxergar o fenômeno da complexidade, tais como “manutenção de redundâncias” e “ornamentação retórica”, características que a princípio parecem ser apenas secundárias da língua, mas que com o tempo vão se tornando indispensáveis para seu bom funcionamento.

Junto a isso, as línguas parecem possuir sua própria organização, ou seja, uma língua tende para uma dada ordem interna, para que lhe seja garantida informatividade. Por exemplo, uma língua SVO tende a possuir “preposição”, enquanto línguas SOV, “posposição”, assim, embora exerçam funções análogas, a ordem interna em que se organizam essas línguas pode apresentar diferentes estratégias comunicativas. Na medida em que as ordens vão se assentando a novos paradigmas (a perda de redundância nos SN plurais, a perda de paradigmas verbais, por exemplo), a informatividade linguística precisa ser garantida por, possivelmente, novos e talvez mais complexos ornamentos linguísticos. Por exemplo, se eventualmente uma língua perde marcação morfológica de tempo no verbo, ela precisa se reorganizar para marcar essa propriedade por outros meios, possivelmente por adjuntos verbais, para que a informação que os falantes de tal língua queiram transmitir seja garantida. É por essa razão, talvez, que muitos desses traços ornamentais não sejam vistos em pidgins e mesmo em línguas crioulas, como atesta McWhorter (2005), justamente porque línguas crioulas são as mais jovens das línguas (ou as menos maturadas, nos termos de Dahl (2004)).

Ainda é importante que se diga que a noção de complexidade não é antônima de simplicidade, mas, como já vimos, complexidade é um traço inerente à língua, portanto noções como “simplicidade”, “quantidade”, “qualidade” não são pertinentes para tais estudos. Para Dahl (2004), por exemplo, complexidade significa “custo”. Dessa forma, para eficiência comunicacional, “custo” não significa vantagem, conforme já vimos. Já para Deutscher (2009; 2014), há, durante anos, acúmulo de significado no léxico, o que pode diminuir o vocabulário de uma língua, e, ao mesmo

tempo, sobrecarregar semanticamente suas unidades lexicais. Talvez um grande exemplo de simplificação aparente seja a perda de paradigma verbal em línguas como o inglês. Dizemos simplificação aparente porque i) o sistema pronominal do inglês tornou-se mais claro com a perda de paradigma verbal, e ii) a língua definitivamente não perdeu sua força expressiva ao se tornar, ao menos em seu sistema verbal, mais simplificada (i.e., com menos distinções morfológicas) quando comparada com estágios anteriores. O mesmo se pode notar com relação à perda de caso *versus* posição sintática mais rígida na passagem do latim para as línguas românicas.

A relevância da idade de uma língua para seu grau de complexidade depende da manutenção de sua transmissão linguística regular, pois com essa transmissão a passagem dos ornamentos de uma geração a outra é garantida, bem como seu acúmulo. Assim, pidgins e línguas crioulas são (i) língua jovens e (ii) resultam de transmissão linguística interrompida ou irregular⁴⁶ (grande massa de adultos aloglotas aprendem uma ou mais línguas), fazendo com que (i) a complexidade linguística das línguas de base se perca e que (ii) não haja ainda tempo suficiente para que pidgins e crioulos, sendo línguas jovens, desenvolvessem seus próprios ornamentos à época de uma dada análise. É importante notar que, com o passar do tempo e com garantia de transmissão regular geracional, línguas crioulas vão ganhar todos os ornamentos das línguas mais velhas. Desse modo, a depender de quando for analisada uma dada língua crioula ela não será notada como tal, mas já terá as características de uma língua velha, com seu desenvolvimento e complexidade previstas devidamente desenvolvidas. Se isso, estiver correto, conforme argumenta McWhorter (2005), ser ou não uma língua crioula é uma característica linguística que depende da idade da língua a ser analisada e de seu mecanismo de transmissão.

Como dissemos anteriormente, as definições mais contemporâneas de complexidade linguística têm a forma de sua época, e atualmente são livres de preconceitos raciais, “paixões idealistas” e políticas explicitamente colonizadoras. Não obstante, ainda falta consenso sobre a melhor forma de entender esse fenômeno, e talvez uma maneira de chegar a esse consenso, esclarecendo ao mesmo tempo a natureza da complexidade linguística, seja buscar maneiras de medi-la – se tivermos métricas precisas e objetivas de complexidade linguística, mais ou menos automaticamente tal conceito será iluminado e poderemos debatê-lo nesses termos. Passemos, então, no próximo capítulo, depois de uma breve conclusão do que vimos até aqui, a apresentar algumas das estratégias já propostas de como medir a complexidade linguística.

⁴⁶ Transmissão linguística irregular e interrompida não são a mesma coisa. Grosso modo, na transmissão linguística irregular ocorre simplificação gramatical quando duas línguas entram em contato em dado contexto o das Grandes Navegações, por exemplo. Já transmissão linguística interrompida, como o nome já sugere, é uma transmissão linguística que não teve seu ciclo concluído, mas não nos deteremos mais nesta diferença aqui. Cf., McWhorter (2007) e Martins (2013).

2.4 Conclusão

De modo geral, mesmo que não estabelecidos com clareza, os sistemas complexos influenciam sobremaneira os estudos sobre complexidade linguística, pois herdamos desses modelos seus principais conceitos, o que parece ser legítimo e esperado. Há de se levar em conta, obviamente, a tradição linguística, cujas descrições são fiáveis a ponto mesmo de questionar truísmos, como é o caso da complexidade linguística. É por essa razão que parecem tão descolados os chamados modelos da ciência da complexidade com os estudos da complexidade linguística. Ou seja, visto que a linguística tem seus próprios modelos, e a ciência da complexidade conflita com alguns desses modelos em alguns de seus preceitos, podemos concluir preliminarmente que da ciência da Complexidade, somente conceitos gerais e epistemológicos têm sido adotados por alguns linguistas, outros simplesmente os ignoram.

Demonstraremos, no capítulo 4, através de um fenômeno singular do português brasileiro, de que forma pode-se aplicar um modelo hoje disponível nos estudos sobre complexidade linguística. Entre este capítulo e o que demonstraremos o fenômeno mencionado, no entanto, apresentaremos alguns modelos vigentes de como medir complexidade linguística.

Capítulo 3 – Alguns modelos para se medir complexidade Linguística

3.1 Introdução

Após expostas algumas noções filosóficas e linguísticas do que seria complexidade, procuraremos neste capítulo mostrar algumas das propostas encontradas na literatura especializada para medir complexidade linguística. Obviamente, não detalharemos as minúcias de todas as propostas para se medir complexidade que é possível encontrar na literatura especializada, e a razão para tanto, para além de impossibilidade prática de tal feita, é que este trabalho se propõe panorâmico. Ainda vale lembrar que este capítulo, se tratando de uma pesquisa bibliográfica, procura expor as ideias de diversos linguistas inclinados à causa da métrica de complexidade linguística do modo mais imparcial possível, ainda que saibamos que a imparcialidade absoluta é inalcançável.

Isso posto, segue neste capítulo algumas das mais difundidas propostas para se medir complexidade linguística. Preferimos não dividir as seções a cada nível da língua, porque um autor ou uma proposta pode apresentar em seu modelo, mais de um nível, ou mesmo porque alguns níveis são mais privilegiados (morfologia) do que outros (semântica), dada a própria natureza linguística desses níveis. Inicialmente proporemos uma divisão de como se pode medir complexidade, mais especificamente sob três pontos de vista. Após isso, passaremos a explicar panoramicamente sobre as principais propostas de como se medir complexidade.

Antes de começarmos, é importante reforçar que, medir complexidade linguística, é uma forma de dar objetividade empírica aos estudos sobre complexidade, numa tentativa de atender aos critérios de cientificidade ora vigentes em linguística descritiva. Se uma métrica de complexidade linguística, o próprio conceito, além de perder seu poder heurístico, pois não teríamos como falar em estágios de maior ou menor complexidade, ficaria, a nosso ver, encerrado em discussões filosóficas, sem contrapartida empírica clara. E, como esperamos mostrar, a complexidade linguística, atualmente, faz parte de pesquisas com lastro empírico.

3.2 Propostas para Medir Complexidade Linguística

Segundo a literatura vigente, tanto sobre complexidade linguística, quanto, mais especificamente, sobre métricas de complexidade, parece haver, entre as propostas de medidas de complexidade, três pontos de vista principais, que culminam em medir complexidade linguística: (i)

pelo viés da **estrutura** da língua, (ii) pelo viés do **uso** da língua, ou (iii) pelo viés do **falante**. Ainda que efetivamente se use como critério de medida mais de um desses traços, um deles sempre receberá maior destaque em uma dada proposta⁴⁷. Por esse motivo, mesmo com uma ressalva, fazemos nossa apresentação sobre as propostas hoje vigentes sobre as métricas de complexidade linguística baseados nessa distinção.

3.2.2 Pelo viés da estrutura

Para Culicover (2013), a observação atenta à estrutura gramatical das línguas dará a resposta adequada sobre a complexidade linguística e sobre sua medida.

Para esse linguista, a tensão entre competência e desempenho pressupõe duas forças que o autor acredita estarem agindo na gramática: (i) a pressão para reduzir a complexidade formal (i.e., morfológica, fonológica e sintática) da gramática, e (ii) a pressão para evitar a complexidade na aplicação da gramática para expressões linguísticas reais, isto é, a complexidade de processamento. Para que isso aconteça, segundo o autor, é necessário responder a duas questões: o que é gramática e como medir a complexidade de uma gramática.

A dificuldade em lidar com essas questões, como vimos, reside no fato de que, para nenhuma delas, haverá respostas conclusivas, em grande parte porque há uma série de propostas em voga sobre o que constitui uma gramática, e sobre o que constitui a complexidade (tanto a complexidade formal quanto a de processamento), e que são, via de regra, incompatíveis entre si. Desse modo, um primeiro caminho para medir complexidade seria determinar o que é a complexidade formal numa gramática e, posteriormente, a complexidade de processamento dessa mesma gramática. Nesse ínterim, vale a pena lembrar que, segundo Culicover, há na gramática uma pressão para que se simplifique, de modo geral, seus sistemas internos.⁴⁸

Assim como para Culicover (2013), para a linguista Johanna Nichols, a complexidade deve ser medida a partir da observação da gramática das línguas. Por conta disso, ela estabelece alguns postulados que devem ser considerados: (i) o número de subsistemas de uma gramática, por exemplo,

⁴⁷ Benedickt (2011) acredita na seguinte divisão possível para se medir complexidade: “1) Medida de complexidade Absoluto-quantitativo; 2) Medida “Barroca” e 3) Irregularidade e Opacidade” (Benedickt 2011: 2). Não entraremos em maiores detalhes, mas podemos adiantar que 1) se refere a métricas ancoradas em dados precisos, 2) se refere àquilo que Dahl (2004) chama de ornamentos, e 3) se refere àquilo que Culicover (2013) chama de idiosincrasias. Preferimos a nossa classificação, por ser mais ampla, mais factível ante toda literatura corrente e por não privilegiar a visão de um só linguista.

⁴⁸ Culicover (2013) acredita em uma “força” ou “forças” que atuam dentro dos sistemas gramaticais para que haja simplificação dos sistemas linguísticos. Tais forças estariam opostas aos traços linguísticos provenientes da comunicação, estes sim, estariam forçando as línguas em direção a uma maior complexidade. Esta noção se aproxima, de certo modo, à noção de “ornamentos” de Dahl (2004).

o número de fonemas consonantais, tons, gêneros, casos, tempos, alinhamentos, vozes, ordens de palavras, estratégias de relativização, meios derivativos para formar verbos denominais, maneiras de formar cláusulas de tempo subordinadas, estratégias de complementação com verbos de cognição, bem como o número de processos derivativos que se aplicam em domínios semânticos particulares; (ii) o número de variantes paradigmáticas, ou graus de liberdade de cada um desses elementos ou conjunto de elementos do sistema: alofones, alomorfes, classes de declinação ou conjugação; (iii) fenômenos sintagmáticos: dependências sintagmáticas entre elementos como concordância, seleção de alomorfes não adjacentes, morfemas não adjacentes ou seleção de categorias e valências; e (iv) restrições dos elementos, aloformas e dependências sintagmáticas, incluindo restrições em sua combinação.

Para Nichols (2009), a questão da aquisição não deve ser levada em consideração para se medir complexidade – posição, aliás, diametralmente oposta à de outros linguistas, como Kusters. Além disso, a linguista frisa que: “[...] para as línguas, o que provavelmente é relevante não é apenas o número de elementos, etc., mas a quantidade de informação necessária para descrevê-los [...]” (Nichols 2009, *in* Sampson *et al.*: 111)⁴⁹. Dessa forma, Nichols toma como relevante os seguintes traços das línguas para sua métrica: estrutura silábica e sistemas tonais na fonologia; número de diferentes categorias flexionais marcadas ou não, marcas de plural e marcas de dual na morfologia; classificação de números ou numerais; diferentes estratégias de concordância entre argumento e predicado e argumentos pronominais na sintaxe, entre outros traços. Através desses parâmetros, a autora consegue computar dados que podem mostrar de que forma diferentes línguas ou diferentes variedades de línguas têm diferentes níveis de complexidade. É interessante observar que os resultados alcançados por esse método, organizados em gráficos pela autora, são muito próximos, ou seja, o nível de complexidade dentre línguas envolvidas, considerando o que foi tabulado é muito próximo. A autora ainda nota que a sintaxe foi o subsistema que apresentou menor variação.

Newmeyer & Preston (2014) organizaram um denso volume no qual vários linguistas debatem possíveis medidas de complexidade. Embora muitos dos linguistas que escrevem nessa coletânea tenham orientação teórica gerativista, eles apresentam propostas de medir complexidade na gramática de maneira um pouco diferente dos estudos propostos no âmbito da gramática gerativa, como também faz Culicover, ou seja, se valem de alguma metalinguagem, mas não de todo o aparato conceitual da gramática gerativa. Segundo esses autores, se assumirmos que as línguas são biologicamente determinadas, teremos que assumir que são todas igualmente complexas, ou que são igualmente simples, o que implicaria em dizer que não há nenhuma diferença substancial entre elas no que tange à complexidade. Nesse sentido, todos os esforços sobre complexidade estariam fadados ao fracasso.

A proposta para se medir complexidade a partir de um modelo gerativo só é possível se se assumir que os princípios, apresentados na teoria dos princípios e parâmetros, podem se manifestar de maneira mais complexas ou menos complexas nas línguas⁵⁰. A própria noção de universais linguísticos, aliás, tem sido também revisada pelos estudos sobre complexidade linguística, conforme veremos adiante.

Inicialmente, para Hawkins (2014 *in* Newmeyer & Preston 2014), há uma associação obrigatória entre complexidade e eficiência, a qual eles definem da seguinte forma: “A comunicação é eficiente quando a mensagem pretendida por S é entregue a H em tempo rápido e com o mínimo de esforço de processamento que pode atingir esse objetivo comunicativo”⁵¹ (Hawkins 2014, *in* Newmeyer & Preston 2014, p.31). Vejamos, ainda que brevemente, algumas das ideias presentes nessa obra para termos um panorama dos problemas e estratégias envolvidos na ideia de medir complexidade linguística pelo viés da estrutura.

Na mesma obra, Gil (2014 *in* Newmeyer & Preston 2014) lembra que as estruturas mais complexas de uma dada língua convivem com as mais simples. Analogamente se numa mesma língua é possível observar estruturas mais cristalizadas (tidas como mais complexas) e outras mais novas (tidas como menos complexas), por que não poderíamos também, enxergar sincronicamente, uma mesma língua que resguarde estruturas complexas e outras mais simples, nos mesmos parâmetros (ou próximos disso) que as mudanças históricas?

Em outro capítulo de Newmeyer & Preston (2014), Sheehan (2014), que trabalha sob a perspectiva do Minimalismo, defende as alegações de que pequenas cláusulas raiz são sintaticamente mais simples do que as cláusulas transitivas que resultam no que chamamos na tradição gramatical de orações subordinadas. Esses resultados sugerem a possibilidade de que a sintaxe evoluiu lentamente, ao invés de repentinamente, com fenômenos complexos surgindo a partir da fundação garantida por estruturas menos complexas (Newmeyer & Preston 2014: 26). A autora, para obter tais resultados, se baseia em evidências neurolinguísticas e até genéticas para afirmar que sintaticamente, ao menos, as línguas não são equiparáveis em suas complexidades.

Em seu capítulo, Theresa Biberauer (2014) defende a ideia de que a complexidade linguística estaria encrustada nos parâmetros, e não nos princípios. É por essa razão que ferramentas tão eficientes provindas da análise gerativa não teriam percebido as nuances de complexidade entre as

⁵⁰ Não é possível não notar aqui uma espécie de manobra *ad hoc* para lidar com o fato de que, mesmo indo em direção contrária aos postulados da teoria que advogam, os linguistas defendem que há, de fato, diferenças de complexidade entre as línguas. Para lidar com isso, ao invés de reverem a teoria, a dotam de “epiciclos” (com o perdão do anacronismo), ou seja, não há somente princípios e parâmetros, mas há manifestações de Princípios mais ou menos complexas. A nosso ver, trata-se de uma manobra que preserva a teoria e que tem como objetivo justamente contornar um problema específico.

⁵¹ No original: “Communication is efficient when the message intended by S is delivered to H in rapid time and with the most minimal processing effort that can achieve this communicative goal” (Hawkins 2014, *in* Newmeyer & Preston 2014, p.31). É interessante notar todas as noções vagas presentes aqui: “eficiente” (de que modo?), “tempo rápido” (quanto?), e “mínimo de esforço e de processamento” (quanto?). Obviamente, uma métrica de complexidade linguística mais objetiva não pode conter tanta vagueza.

línguas. Essa noção casa com discussões acerca da complexidade linguística que a aproximam mais do caráter social das línguas, do que do caráter estrutural.

Já Chasei e Moro (2014), que são neurolinguistas, procuram mostrar a relação entre complexidade formal, tal como a definida por Chomsky em sua hierarquia; e a atividade cerebral.

Eles argumentam que a Hierarquia de Chomsky facilita medidas de complexidade de processamento através dos autômatos que modelam os vários níveis de gramática na hierarquia. A complexidade pode ser precisamente quantificada como uma função do tempo (o número de travessias de estado necessárias para processar interações hierárquicas) e espaço (a quantidade de memória necessária para acompanhar as dependências) (NEWMAYER, PRESTON 2014: 12)⁵².

No limite, os linguistas procuram responder uma questão, no mínimo, incômoda ao modelo de Chomsky. Contrariamente a outros formalistas, e a outros gerativistas, eles acreditam que a resposta para se medir complexidade linguística se encontra dentro do programa gerativista.

Já Menn e Duffiel (2014, *apud* Newmeyer & Preston, 2014, p. 12) defendem o ponto de vista de que se possa medir complexidade linguística apenas quando se mede a atividade neurológica correspondente à linguagem. O que eles chamam de trabalho com língua em tempo real, com um índice batizado de “frequência de predicabilidade”, por meio do qual se mede comparativamente entre línguas o quanto a mesma informação pode ser transmitida em X números de sentenças ou mesmo de unidades menores, como morfemas e fonemas.

Ainda segundo os mesmos Newmeyer & Preston (2014), o objetivo geral do livro é procurar mostrar que a corrente formalista, ao contrário do que poderia ser pensado inicialmente, está apta a responder questões em torno da natureza da complexidade gramatical⁵³. Apostam nisso ao também se ampararem nas ciências cognitivas: psicolinguística e neurociências. Dessa forma, cruzam dados advindos da psicolinguística e da neurolinguística, ou seja, o que é habitualmente chamado de “processamento on-line” da língua para se determinar tanto o que é complexidade, quanto as medidas ideais. Neste sentido, a nosso ver, uma das maiores contribuições na obra é de Hawkins (2014), que justamente levanta alguns questionamentos, tais como: de que maneira o formalismo pode ajudar a

⁵² No original: “They argue that the Chomsky Hierarchy facilitates measures of processing complexity via the automata that model the various levels of grammar on the hierarchy. Complexity can be precisely quantified as a function of time (the number of state traversals required to process hierarchical embeddings) and space (the amount of memory needed to keep track of dependencies)”. (Newmeyer & Preston 2014: 12)

⁵³ A necessidade de pontuar tal fato se dá provavelmente porque boa parte do trabalho sobre o impacto de questões sócio-histórica na estrutura linguística, bem como de mudança linguística e gramaticalizações, se deu em sua maior parte como resultado de trabalho de cunho menos formalista, e certamente não gerativista. Apesar de haver trabalhos gerativistas sobre questões de linguística histórica e mudança linguística, não seria incorreto dizer que esses trabalhos são relativamente recentes.

medir complexidade? Há níveis de complexidade? O formalismo seria mesmo capaz de tampar buracos deixados pelo funcionalismo? (Newmeyer & Preston 2014: 27).

Hawkins assume teoricamente que existem primitivos sintáticos e que a partir desses primitivos sentenças cada vez mais complexas seriam derivadas. O mesmo acontece, segundo ele, com as inversões de ordens. Ou seja, algumas inversões autorizadas por algumas línguas, estariam ligadas à maior complexidade dessas línguas. “Quando as línguas permitem escolhas entre ordens de palavras diferentes, as alternativas podem ser descritas em termos dos tamanhos relativos de seus ‘Domínios de Combinação de Frases’.”⁵⁴ (Hawkins 2014, p. 15 *in* Newmeyer & Preston). Em outras palavras, a partir dessa premissa, Hawkins (2014), baseado nos modelos minimalistas, aponta alternativas para se medir com alguma precisão o que seria a complexidade sintática. Para tanto, ele estabelece alguns princípios gerais:

- Minimizar domínios. Quanto maior o domínio para uma tarefa de processamento, maior a complexidade.
- Minimizar formalizações. Quanto mais complexidade há numa descrição, mais complexo é o processamento gramatical.
- Maximizar o processamento on-line. Quanto menos propriedades que podem ser atribuídas a cada item X conforme X é processado, há mais complexidade na língua.

Além disso, Hawkins admite que fatores sociais e históricos afetam a complexidade linguística, e que uma longa tradição sustenta que diferentes tipos de contato de idioma e diferentes tipos de identidade de línguas afetarão a complexidade do idioma. Mas não há nenhum consenso sobre exatamente como. (Hawkins 2014, *in* Newmeyer & Preston 2014).

De modo prático, Hawkins (2014) usa um critério bem interessante para medir complexidade: observando pronomes resumptivos e, alternativamente, em seu lugar, sentenças sem os mesmos pronomes, ou seja, com um zero. Dado isso, percebe se há ou não maior ou menor processamento de conteúdo, a partir de um índice que, segundo ele, foi proposto por Keenan e Comrie (1977), a

[...] Hierarquia de Acessibilidade (AH) para descrever padrões na produtividade da formação de cláusulas relativas através de línguas e na seleção de diferentes tipos de cláusulas relativas
[...] As lacunas favoreceram as posições mais altas da AH, pronomes restritivos as posições

⁵⁴ No original: “When languages permit choices between different word orders, the alternatives can be described in terms of the relative sizes of their ‘Phrasal Combination Domains’” (Hawkins 2014, p. 15 *apud* Newmeyer & Preston).

mais baixas nos dados de Keenan e Comrie em idiomas que apresentam ambos os casos (NEWMeyer, PRESTON 2014: 20)⁵⁵.

Os linguistas Yoan Chabot e Christophe Nicolle (2015) são dos únicos nomes que se inclinam a medir complexidade semântica. A princípio eles trabalham com o léxico, talvez uma das partes mais salientes e mais rapidamente acessíveis da estrutura semântica das línguas naturais. Chabot & Nicolle se valem dos quantificadores, que eles julgam ser universais, comum a todas as línguas naturais, para procurar nuances de complexidade linguística entre os elementos dessa classe. Os quantificadores podem sugerir que uma língua pode sim ser semanticamente mais ou menos complexa do que outra. Mais especificamente, quantificadores ambíguos tendem a complexificar as línguas naturais.

Inicialmente, há já mais de duas décadas, criaram-se parâmetros para medir complexidade semântica com a finalidade de implementação em máquinas⁵⁶, ou seja, para fins de leituras computacionais, e o tratamento adequado da semântica das línguas naturais ainda hoje representa uma enorme barreira para a inteligência artificial (IA). Os parâmetros propostos para este fim limitam-se a distinguir parentescos semânticos, similaridades ou mesmo distâncias entre termos. Além disso, a tensão entre conhecimento linguístico e conhecimento de mundo sempre foi outro desafio a implementações da IA, aliás, é um grande desafio para boa parte dos estudos semânticos, que recorrem, ao longo de sua história, a modelos lógicos cada vez mais complexos e sofisticados para lidar com mundos possíveis, estados epistêmicos, gosto pessoal etc. Mesmo assim, a comparação de semelhança semântica tem sido implementada com certo sucesso em âmbito computacional, notadamente em sumarizadores textuais automáticos. Ainda que o fenômeno da ambiguidade, uma das características mais peculiares da semântica das línguas naturais, tenha dificultado bastante tal trabalho.

Com a ascensão dos estudos sobre a complexidade linguística, alguns desses parâmetros aqui vistos passaram a servir também aos estudos sobre complexidade semântica. Embora ainda hoje a finalidade principal de mensurar semântica seja, sobretudo, a IA, medir complexidade semântica das línguas é campo de estudos também reclamado pela complexidade linguística. É justamente neste campo de estudos que Chabot & Nicolle procuram mostrar alguns avanços nos estudos semânticos. Empiricamente, no entanto, há um problema: sabe-se relativamente pouco sobre como as línguas variam semanticamente. Além disso, se há variação semântica, não se sabe com precisão de que forma ela se dá, nem como medir tais variações. Por exemplo, estudos sobre mudança semântica, de um

⁵⁵ No original: “[...] Accessibility Hierarchy (AH) for describing patterns in the productivity of relative clause [...] Gaps favored the higher positions of the AH, resumptive pronouns the lower positions in Keenan and Comrie’s data in languages that had both.” (Newmeyer & Preston 2014: 20).

⁵⁶ Alguns desses parâmetros semânticos que linguistas sentem mais dificuldade em implementar computacionalmente seriam os fenômenos de ambiguidade, por exemplo, bem como noções sintáticas que acarretariam mudanças semânticas, como é o caso da mudança de “voz” nos verbos (Othero e Menuzzi 2005).

ponto de vista estrutural/formal, estão ainda em sua infância (Cf. Ilari e Basso, 2020). A exemplo de medições de complexidades sintáticas, haveria sim a possibilidade de variação semântica, bem como de quantificá-la. Outro problema sobre medir a semântica das línguas é de que não há consenso se realmente existem universais semânticos nas línguas nem sobre como identificá-los. Por isso mesmo uma das primeiras propostas no campo de complexidade linguística é observar os quantificadores que, aparentemente, são comuns em todas as línguas naturais, ainda que seu funcionamento varie bastante de uma língua para outra. O que, na verdade, pode ser o começo de um parâmetro para que se possa observar os diferentes graus de complexidade semântica.

Os modelos para se medir complexidade semântica utilizados por Chabot & Nicolle, vão desde métricas baseadas em gráficos que medem parentescos lexicais, até sofisticados softwares outrora utilizados em prognósticos de saúde. Aliados a tais softwares, os raros estudos sobre como medir complexidade semântica estão atrelados à tipologia linguística, bem como a algumas ferramentas utilizadas pela semântica de valor de verdade, que se mostra há muito tempo como uma boa estratégia para detectar as diferenças entre quantificadores. Dessa forma, Chabot & Nicolle (2015) procuram mostrar o que se tem desenvolvido nesse campo, e de que forma os linguistas têm criado métodos e noções dentro da tipologia e da semântica formal para medir complexidade semântica das línguas e, talvez o objetivo mais desafiador, tentar mostrar que há pequenas variações dessa complexidade entre as línguas ou mesmo entre variedades de uma mesma língua.

Outro volume bastante importante para a questão da métrica de complexidade, especificamente sobre morfologia, foi organizado por Baerman, Brown e Corbett (2015). Nele, os autores argumentam que a morfologia é o nível mais saliente para que se possa observar as diferentes complexidades nas línguas. Mesmo assim, a morfologia pode variar bastante entre as línguas, de modo que um dado morfema, aparentemente com a mesma função em duas diferentes línguas, pode não apresentar correspondência clara. Além disso, a abundância morfológica de uma língua não prova necessariamente que esta língua é mais complexa do que uma língua que apresente um inventário morfológico reduzido. Talvez por essas razões é que seja tão limitada a contagem computacional em se tratando de descrição morfológica automatizada. De todo modo, no volume citado alguns linguistas mostram algumas alternativas quantificadas por computadores. Lembram os organizadores que em morfologia, o grau de entropia pode se aproximar de zero, condição diferente da sintaxe. Isso porque, diferentemente da sintaxe, em morfologia, eventuais inversões são, quase sempre, desastrosas do ponto de vista do processamento de língua. Os mesmos organizadores mostram que o tema mais debatido na obra é a flexão, pois parece ser a característica mais eficiente para se comparar complexidade entre as línguas. Casos comparativos como a adição de morfema em português para marcar plural (“livro”/“livros”), parecem ser menos complexos do que em casos em que, além da adição de morfemas, há também mudança fonética, como no alemão (“buch”/“bücher”).

Na obra organizada por Newmeyer & Preston (2014), Sagot & Walther (2014), em seu capítulo, propõem uma métrica que prevê que estruturas morfológicas reduplicadas tendem a ser menos complexas, isso porque são mais previsíveis. Ao mesmo tempo em que, em praticamente todas as línguas verificadas, as reduplicações apresentam função semântica muito bem demarcada, que é, em geral, reforçar o sentido do morfema duplicado em questão.

Stump e Finkel (*in* Newmeyer & Preston, 2014) mostram um número de diferentes métricas implementadas computacionalmente para medir complexidade nas flexões. Sua grande contribuição é a de contrastar a saída e a entrada da informação, via fonologia. Ou seja, eles avaliam se tais e tais morfemas têm o mesmo valor tanto para o falante quanto para o ouvinte. Dessa forma, os linguistas conseguem resultados mais seguros para medir flexão.

A obra de Baerman, Brown e Corbett (2015) também enfoca a organização interna do paradigma, em particular no mapeamento indireto entre os valores morfossintáticos e a forma morfológica que caracterizam um sincretismo. Assim, Milizia (2015 *on* Baerman, Brown e Corbett 2015, p. 8) invoca a teoria da informação para mostrar de que maneira os morfemas são semanticamente úteis, bem como que forças contribuem para a formação de sincretismos, “[...] sugerindo que a carga de informação inerente a um sistema com classes de flexão pode ser uma motivação para a fusão de células paradigmáticas”⁵⁷ (Baerman, Brown e Corbett 2015: 9).

Num outro capítulo da mesma obra, os linguistas Bank e Trommer (2015) desenvolvem um método automatizado de segmentação morfológica que “[...] permite uma avaliação quantitativa de diferentes estratégias de segmentação, equilibrando as complexidades da representação morfossintática e da representação exponencial da morfologia”⁵⁸ (Baerman, Brown e Corbett 2015: 9). Através desta pesquisa, os autores percebem que a morfologia é montada numa base bastante elegante e econômica do ponto de vista do usuário. Deixam em aberto, no entanto, a questão do porquê, mesmo com todas as vantagens, haver línguas que não possuem morfologia. Caso das línguas isolantes, como o mandarim, também conhecidas como línguas monomorfêmicas, em que não há fragmentação de unidades menores, sejam gramaticais ou léxicas. Em suma, não possuem morfemas entre as palavras.

Num texto seminal para se debater métricas de complexidade linguística, Miestamo (2009) se vale de um universal linguístico de Greenberg (1966) para mostrar como seria possível medir complexidade entre línguas naturais.

⁵⁷ No original: “[...] suggesting that the information load inherent in a system with inflection classes may be a motivation for the conflation of paradigmatic cells”. (Baerman, Brown e Corbett 2015: 9).

⁵⁸ No original: “[...] allows a quantitative assessment of different segmentation strategies, balancing the complexities of morphosyntactic representation and morphological exponence” (Baerman, Brown e Corbett 2015: 9).

Greenberg formulou seu universal 34 da seguinte forma: “Nenhuma língua tem um número trial a menos que tenha um dual. Nenhuma língua tem um dual a menos que tenha plural”⁵⁹ (Miestamo 2009: 80). Assim, para Greenberg, haveria a seguinte hierarquia: “(1) hierarquia de número: (singular) < plural < dual < trial” (Greenberg 1963, *apud* Miestamo 2009: 80)⁶⁰. Nesse sentido, Miestamo procura provar aquilo que McWhorter (2004 *apud* Miestamo 2009) já dizia: “[...] quanto mais “números” há na marcação, mais a língua fica complexa [...] quanto maior uma língua sobe em uma dada hierarquia, mais complexa é sua gramática a esse respeito”⁶¹ (Miestamo 2009: 80).

Embora a conexão entre complexidade e hierarquia não seja tão evidente, Miestamo cita Keenan e Comrie (1977), para quem há hierarquias de ordem evidentes, em que dadas ordens prevalecem, enquanto outras se rareiam nas línguas. Seja como for, Miestamo (2009) procura associar os universais de Greenberg e sua tipologia com os parâmetros da complexidade linguística. O volume que ele ajuda a organizar (Miestamo *et al.* 2008) mostra com detalhes de que forma a tipologia linguística contribui com as medidas de complexidade.

Ao estudar três diferentes variedades do quéchua em sua tese de doutoramento, Kusters (2003) observou aspectos sobre a natureza verbal dessas variedades: um dos sistemas (o mais antigo) com o paradigma verbal pleno, e duas das variedades com o paradigma verbal mais simplificado, entre elas está o quéchua usado como língua franca, que tem o paradigma verbal ainda mais simplificado. Diante desses fatos, Kusters (2003) conclui que há três critérios conjugados que poderiam determinar diferentes graus de complexidade das variantes dessa língua: 1) *Critério de Economia*, ou seja, o número de categorias expressas pelo verbo (quanto menos paradigmas um verbo apresentar, menos complexo ele é); 2) *Critério de Transparência*: ao expressar, por exemplo, a noção de pessoa num morfema, o verbo é mais transparente do que aquele que não faz isso; e 3) *Critério de Isomorfia*: diz respeito à cristalização de certas palavras e expressões que a língua possui.

McWhorter (2005; 2007) apresenta pelo menos dois trabalhos importantes sobre línguas crioulas, pidgins e línguas de contato que envolvem estudos sobre complexidade linguística. McWhorter tem mostrado convincentemente como os estudos sobre complexidade linguística podem ser aplicados de maneira bastante produtiva aos estudos sobre línguas resultantes de contatos e línguas novas (que seriam justamente línguas que tiveram, no tempo histórico recente, origem em pidgins). Dessa forma, suas propostas para medir complexidade entre as línguas têm obtido algum êxito e notoriedade, e a razão para tanto é que seus apontamentos nesse sentido vêm mostrando não apenas

⁵⁹ No original: “No language has a trial number unless it has a dual. No language has a dual unless it has a plural” (Greenberg 1963, *apud* Miestamo 2009: 80).

⁶⁰ No original: “(1) The number hierarchy/(singular) < plural < dual < trial” (Miestamo, 2009, p. 80).

⁶¹ No original: “[...] the higher a language climbs on a given hierarchy, the more complex its grammar is in that respect” (Miestamo, 2009, p. 80).

que línguas jovens e línguas de contato são mais simples (ou relativamente menos complexas), mas também, de que forma as línguas mais antigas ganham complexidade.

Seus critérios pautam-se em (i) superespecificação (a marcação aberta e obrigatória das distinções semânticas), (ii) elaboração estrutural (o número de regras em morfologia, fonologia, sintaxe etc.) ou o tamanho dos inventários (categorias funcionais, fonemas etc.), e (iii) irregularidade.

Assim, tais parâmetros não servem apenas para “medir” a complexidade em línguas emergentes, mas também a complexidade de toda e qualquer língua. Embora tenha causado furor nas pesquisas sobre línguas crioulas, o método de McWhorter tem sido largamente comentado e mesmo utilizado há mais de dez anos, obviamente não sem críticas. De todo modo, McWhorter (2001) tem o mérito de tentar explicar a emergência de línguas jovens em termos estritamente linguísticos e sincrônicos. Seu modelo foi um dos primeiros e até agora uma das principais contribuições para os estudos da complexidade linguística. Além disso, o trabalho de McWhorter tem um viés tipológico que merece ser mencionado. Sua teoria, ao fim e ao cabo, reconhece três tipos de línguas. A primeiro tipo seria aquele das “línguas velhas”, ou seja, línguas cuja transmissão geracional se deu normal e ininterruptamente, de modo que todos os ornamentos cabíveis puderem se desenvolver e se fixar, fazendo com que a língua tenha uma complexidade compatível àquela das “línguas velhas”. Os dois outros tipos de língua correspondem a “línguas jovens”, e, portanto, com menor complexidade justamente porque não houve tempo suficiente para que os ornamentos se desenvolvessem e se fixassem. A primeira dessas línguas jovens é formada pelas línguas crioulas que, segundo a teoria de McWhorter, se formam sempre a partir de um pidgin. Assim sendo, como já mencionamos, essas línguas seriam menos complexas porque têm menos idade, e resultam de uma organização de um pidgin, muito longe de uma transmissão linguística geracional regular. Um terceiro tipo de língua jovem é o que podemos chamar de “língua interrompida”. Nesse caso, temos a situação em que uma dada língua velha deve ser aprendida por falantes, em geral, adultos, num grande número, em pouco tempo, o que configuraria um processo de transmissão linguística irregular (ou interrompido), cujo resultado é uma versão mais simplificada da língua velha em questão. Há muitos pontos polêmicos aqui, mas há também uma heurística positiva bastante interessante, e vemos nessa teoria uso produtivo de ideias que mobilizam a complexidade linguística.

Nessa seção, vimos de que forma as estruturas das línguas podem nos fornecer dados para que se meça complexidade linguística. As propostas são as mais variadas e englobam tanto linguistas da vertente funcionalista, como linguistas da vertente formalista. Seja como for, todos atualmente parecem concordar que, por mais que haja modificações de complexidade das línguas em suas estruturas, não são nelas que a complexidade eclode. Pelo contrário, Culicover (2013) afirma que há uma força nas estruturas das línguas que tendem a simplificá-las.

3.2.2 Pelo viés do uso

Östen Dahl apresenta uma proposta bastante debatida e criticada, tanto para medir complexidade, quanto para a aplicação das noções advindas desses estudos. Como vimos, ele aposta na idade da língua e em sua conseqüente maturação para explicar certos aspectos da complexidade linguística. Dahl (2004) procura apontar indícios de que à medida que as línguas envelhecem, e são regular e ininterruptamente transmitidas de geração a geração, elas adquirem certos “ornamentos”⁶² que, ao olhar menos acurado, podem parecer meros caprichos obsoletos produzidos pelo tempo. No entanto, é justamente a partir desses ornamentos que Dahl aponta saídas para medir os níveis de complexidade entre os sistemas de diferentes línguas. Por sua vez, línguas mais jovens não teriam tais (ou tantos) ornamentos. A esses “meros ornamentos”, Dahl chama de *superexposição*, que na verdade são traços cumulativos que as línguas naturais adquirem somente com o tempo e parecem ter a função de tornar a transmissão de conteúdo mais clara ao interlocutor. Assim, Dahl acredita que a língua é, em dada medida, herança social e não apenas inata, ou em grande parte inata.

Na verdade, Dahl se baseia na noção de *meme* de Richard Dawkins (1976), para quem a língua é um perfeito exemplo do que seja um meme: uma “instituição” cujas premissas inatas não são suficientes para desenvolvê-la, ou seja, é preciso muito “input” da sociedade. Mais do que isso, esse “input” social seria herdado desde há muitas gerações. Note-se que esta noção adotada por Dahl, a partir de Dawkins (1976), é também bem próxima do que acredita Gell-Mann (1994), para quem a língua é um tipo de “sistema adaptativo complexo”, que só seria concretizado com os dois fatores: inato e social. Visto que esta discussão já foi devidamente desenvolvida no capítulo anterior, a trazemos aqui mais como um ponto de convergência para mostrar em maior detalhe as ideias apregoadas por Dahl (2004).

Assim como outros linguistas – talvez os principais nomes da complexidade linguística –, Dahl se fia na tipologia para mostrar que as línguas naturais podem diferir em sua complexidade:

É bom lembrar que o dispositivo inato que supostamente nos permite adquirir nossa língua nativa tem figurado sob pelo menos dois nomes diferentes: UG (“Universal Grammar”) e LAD (“Language Acquisition Device”). Esses nomes podem ser entendidos como representando duas formas bem diferentes de olhar para o papel desse componente, onde UG

⁶² Dahl (2004) entende como “ornamentos”, traços que as línguas adquirem durante o tempo para, dentre outras funções, garantir a qualidade da informação a ser transmitida. Um bom exemplo de ornamento, como vimos anteriormente, seria a marcação redundante de número plural numa cadeia nominal em português.

representaria o conhecimento estático e LAD para a dinâmica de aquisição⁶³ (DAHL 2004: 59).

Assim, Dahl estaria apontando para outras possibilidades além desta que ele propõe. Ou seja, se a língua é vista sob o prisma da tipologia ou se é vista sob o prisma da aquisição, pode-se medir complexidade linguística a partir dessas diferentes perspectivas, logo, pode-se inferir que “complexidade linguística” talvez não seja um termo que se refira a algo, num certo sentido, estático, que se possa medir de apenas um ponto de vista.

Dahl (2004) faz uma explanação de um modelo matemático que deverá auxiliar em algumas de suas análises e premissas sobre complexidade linguística. Como exemplo, ele nos fornece três números (3, 4 e 7) entre três posições com possibilidades de 0-9, ou como ele chama: três probabilidades. Probabilisticamente temos então 1.000 possibilidades (= $10 \times 10 \times 10$). Em seguida, nos fornece quatro números e igualmente quatro possibilidades (ou probabilidades), e da mesma forma, as chances de se obter números diferentes são $10 \times 10 \times 10 \times 10$. E assim por diante. No entanto, inclui letras em lugar de alguns números e constata que as probabilidades agora aumentam, visto que na média da escrita das línguas, o número de letras num alfabeto é geralmente 26. Se assim for, para quatro possibilidades teremos $26 \times 26 \times 26 \times 26$ diferentes combinações para se formar, por exemplo, uma palavra de quatro letras (é importante notar que o recurso à escrita é apenas uma simplificação). Obviamente, essa regra não anula as regras próprias de cada língua em suas possibilidades de formar palavras que tenham sentido. Para todos os exemplos, sobretudo para este último, ele se utiliza da metáfora de um cadeado, em que algumas ou só uma das possibilidades o abrirá. Este exemplo ilustra um passo adiante na sua explicação, qual seja: dadas as limitações impostas por um dado sistema, digamos o sistema que rege a formação de palavras, as chances de se descobrir ou “adivinhar” um dado em cada local, se reduz:

No caso mais simples, representado por travas de combinação do tipo descrito acima, este número corresponde ao tamanho do espaço de estado do sistema e também determina uma de suas propriedades fundamentais - o que é frequentemente referido como "incerteza", mas poderia talvez melhor ser chamado de "dificuldade de adivinhação" - isto é, a probabilidade de que uma estimativa aleatória corresponda a um determinado membro do conjunto (DAHL 2004, p. 6)⁶⁴

⁶³ No original: “Remember that the innate device that supposedly enables us to acquire our native language has figured under at least two different names: UG (“Universal Grammar”) and LAD (“Language Acquisition Device”). These names can be understood as representing two rather different ways of looking at the role of this component, where UG would stand for the static knowledge and LAD for the dynamics of acquisition” (Dahl 2004, p. 59).

⁶⁴ No original: “In the simplest case, represented by combination locks of the kind described above, this number corresponds to the size of the state space of the system and also determines one of its fundamental properties—what is often referred to as “uncertainty,” but could perhaps better be called “guessing difficulty”—that is, the probability that a random guess will match a given member of the set” (DAHL 2004, p. 6)

Assim, Dahl introduz algumas noções importantes como as de “incerteza” e de “adivinhar dificuldade”, que nada mais são do que a probabilidade em desvendar um determinado membro do conjunto, em princípio, aleatório. Nesses termos, caberia um modelo matemático conhecido como **código binário**, o qual as chances se reduzem à medida que se vai “adivinhandando” os algoritmos de cada lugar adequado, e sempre se reduzem a duas possibilidades para cada caso/lugar. Dentro dessa possibilidade, propõe outro exemplo ilustrativo:

Assim, para identificar um número entre 1 e 8, basta sempre fazer três perguntas ao longo do seguinte padrão: “É maior que 4?”, “É maior que 6?”, “É 7 ou 8?”. Em vez de declarar o número de combinações, podemos simplesmente fornecer o número de escolhas binárias (perguntas sim-não) necessárias para identificar um membro árbitro do conjunto (DAHL 2004, p. 7)⁶⁵.

Dessa forma, Dahl apresenta um modelo de código binário que a princípio poderia ser utilizado para medir complexidade. No entanto, ele argumenta que essa metalinguagem, de certa forma, só serve para “valores informacionais”, assim não se presta necessariamente à semântica ou à pragmática, mas basicamente aos “bons usos” da língua, com sentenças bem formadas, gramaticais. Nesse sentido, suscita Shannon 1947 (*apud* Dahl 2004), que, basicamente, estuda a maneira e as boas condições que uma informação é transmitida com sucesso. Algo parecido com que Roman Jakobson (1979) chamava de **canal**. Aqui o objetivo é tentar definir quão fácil ou quão difícil é transmitida uma informação: “Quanto menos provável for um sinal, mais informações serão transmitidas por ele. Isso às vezes é chamado de “informação sintática”, em oposição a “informação semântica”, que seria “informação sobre” algo” (DAHL 2004, p. 7)⁶⁶.

Já para Shannon, então, “informação sintática” quer dizer “predicabilidade”. E predicabilidade quer dizer: informação entre dois termos, ou melhor, entre duas espécies/tipos de informação, isso para efeitos de formalização em uso de código binário, ou, no exemplo que Dahl está usando aqui: “Uma escolha binária – sob o nome de bit – é de fato a unidade básica em que a informação é medida na teoria da informação (DAHL 2004, p. 7)⁶⁷. Por fim, Dahl diz que o código binário não é totalmente adequado para aplicar no tipo de problema a que se almeja, a complexidade linguística, mas é um exemplo de uma tal estratégia.

⁶⁵ No original: “Thus, to identify a number between 1 and 8, it is always enough to ask three questions along the following pattern: “Is it larger than 4?”, “Is it larger than 6?”, “Is it 7 or 8?”. Instead of stating the number of combinations, we can then simply give the number of binary choices (yes-no questions) necessary to identify an arbitrary member of the set” (DAHL 2004, p. 7).

⁶⁶ No original: “The less probable a signal is, the more information is transmitted by it. This is sometimes called “syntactic information”, as opposed to “semantic information”, which would be “information about” something” (p. 7)

⁶⁷ No original: “One binary choice — under the name of bit — is in fact the basic unit in which information is measured in information theory” (DAHL 2004, p. 7).

Dahl opta por outro modelo matemático já bem estabelecido para medir complexidade linguística, trata-se do modelo de Kolmogorov. Esse modelo tem como objetivo medir a quantidade de informação, por isso é também conhecido mais amplamente como teoria da informação de Kolmogorov, ou seja, um dos vários modelos de teoria da informação encontrados em ampla literatura. Neste, a informação de uma dada sentença é vertida em valores que podem ajudar a calcular o custo de processamento que tem determinada informação em uma língua. Apresenta, assim, um modelo de cálculo para analisar a língua, mais propriamente, um modelo de cálculo que visa diminuir o ruído entre duas ou mais informações num dado ambiente de comunicação. O problema é que, por mais complexo que pareça o cálculo, mesmo que em sua forma abreviada, em código binário, a língua real é muito mais complexa, não se resume em, por exemplo, informações do tipo “sim-não”. Além disso, Dahl diz que na língua real há inevitáveis ruídos⁶⁸ e que estes são mesmo até necessários à língua. São importantes para a produção da língua natural. Mesmo assim, seu modelo, possivelmente para fins didáticos, elimina de seu cálculo tais ruídos.

Talvez um dos problemas constatado pelo modelo oferecido por Dahl seja a questão da expressividade⁶⁹. Para o autor, um desses ornamentos, a redundância das marcações de plural em línguas românicas, por exemplo, não seria um mero detalhe adquirido pelo capricho histórico das línguas, mas sim um traço que ajudaria a fortalecer o potencial semiótico, expressivo ou mesmo comunicativo da língua. Nesse sentido, afirmar que línguas que apresentam esse traço são mais expressivas do que línguas que não o apresentam seria uma afirmação perigosa. Não se observa perda de expressividade, justamente, em línguas românicas que vêm perdendo marcas redundantes de plural nos seus sintagmas nominais, caso já muito bem sabido do francês, por exemplo.

Por outro lado, línguas que não apresentam tais ornamentos, possuiriam menos complexidade nesses sistemas, mas também diminuiriam seus “custos” de processamento computacional. Dito de outra forma: os ornamentos tornam as línguas mais complexas, ao passo que lhes geram mais “custos” no processamento computacional. Assim, se por um lado, segundo o mesmo Dahl (2004), as línguas ganham ornamentos que garantem a qualidade de sua informação, por outro, essas informações,

⁶⁸ Chamamos aqui de ruído o que Jubran (*in* Jubran e Koch, 2006, pp. 33-35) reconhece como fenômenos naturais da fala, tais como as *hesitações* e as *interrupções*.

⁶⁹ Estamos usando “expressividade” de um modo um tanto intuitivo, e não é uma tarefa nada simples definir precisamente esse conceito. Por exemplo, para Searle, expressividade é um *ato de fala* que tem relação com atitude do falante frente a uma determinada situação (Brown e Miller, 2013, p. 163). Para os estruturalistas e para a tradição gramatical, expressividade tem a ver com a tradução de certos pensamentos, mesmo que diluídos em morfemas. Já para McCready (2014), *expressividade* tem relação também com *tradutibilidade* de alguma expressão linguística. Assim, expressividade é parte de um modo de significar, que pode diferir entre línguas, “[...] o que pode ser expresso difere entre as línguas em termos do que é transmitido pela língua, não apenas pela estrutura comunicativa” (McCready, 2014, p. 54). Para nossos propósitos, adotaremos uma noção ampla, segunda a qual expressividade de uma língua tem a ver com tudo aquilo que pode ser dito por uma dada língua. Em princípio, todas as línguas, independentemente de terem ou não graus diferentes de complexidade, têm a mesma expressividade. Ou seja, tudo pode ser dito em qualquer língua, ainda que com recursos estruturais diferentes. Ao assegurar que todas as línguas são igualmente expressivas – algo que adotaremos, sem, contudo, defender explicitamente – podemos falar sobre diferenças de complexidade sem cair em diferenças de expressividade ou “poder expressivo”.

digamos morfológicas, adicionais, acarretariam num maior custo de processamento dessas línguas. Em muitos casos, afirma Dahl, tais ornamentos garantem a veiculação da informatividade entre os falantes desta língua – a redundância oferece mais garantia de que uma determinada informação seja veiculada –, porém exigem, ao mesmo tempo, mais processamento, e as línguas encontrariam sempre um equilíbrio entre a redundância e a economia.

Para Szmrecsanyi & Kortmann (2012), não há evidências nem biológicas nem comunicativas de que há diferenças de complexidade nas estruturas das línguas. Portanto, a complexidade linguística só pode ser diferença social, e, sendo assim, para o linguista, a estrutura linguística adquire sua complexidade através da junção de fatores sociais e sócio-históricos:

- 1) Contato linguístico: simplificação de adulto no/a SLA (Second Language Acquisition): simplificação (Lupyan & Dale 2009);
- 2) Bilinguismo infantil: complexificação (Trudgill 2011: 42);
- 3) Isolamento: complexificação (Wray & Grace 2007, Nichols 2013).

Dessa forma, Szmrecsanyi & Kortmann (2012) estão aparentemente de acordo com Culicover (2013), para quem a complexidade de uma língua é indissociável de suas influências externas: somente a partir delas é que a estrutura de uma dada língua se tornaria mais ou menos complexa.

Em um artigo publicado em 2007, o matemático polonês Relja Vulcanović propõe uma medida de complexidade que também leva em conta a relação entre estrutura linguística e a informação por ela transmitida. Segundo o autor, sua “fórmula permite uma comparação da complexidade das estruturas em relação à informação linguística por elas transmitidas”⁷⁰ (VULANOVIC, 2007: 399). Desse modo, pode-se vislumbrar tanto a estrutura gramatical das línguas estudadas, bem como seus traços semânticos e mesmo pragmáticos, para os quais estão previstas as devidas variáveis computáveis no seu modelo. Para tanto, o autor aplica um modelo de análise sintática funcionalista para medir complexidade apenas de sentenças simples. Ainda que em formalizações, a teoria empregada aqui é flagrantemente inspirada em Dik (2000, *apud* Vulcanovic 2007), pois prioriza uma dada medida que usa como parâmetro alguma organização sintático-semântica em contraste com a quantidade de informação viabilizada por este montante de signos nelas empregadas (Vulanovic, 2007: 399). Assim como Dahl, o autor opta por trabalhar com parte da teoria que tomo como metodologia teórica, notadamente uma teoria cuja informação é de suma importância: “Simplificando, uma gramática é menos complexa, e ao mesmo tempo mais eficiente, se transmite

⁷⁰ No original: “The formula enables a comparison of the complexity of these structures relative to the linguistic information they convey” (Vulanovic, 2007: 399).

mais informações (mais significado) com menos formas e regras⁷¹.” (VULANOVIC, 2007: 399). E assim como Dahl, o autor acredita que menos processamento redunde em ganhos para transmissão de informação na língua, tornando-a, via de regra, menos complexa. Curiosamente, no modelo utilizado pelo matemático, na representação sintática não há propriamente hierarquias, redundado, então numa mudança de parâmetros métricos, com relação a todos os outros linguistas aqui observados. Assim, quanto mais simples a gramática, maior sua eficiência. É por isso que se define as medidas de complexidade gramatical e eficiência como recíprocas entre si. Este conceito de eficiência gramatical, portanto, remonta a Vulanovic (1991 *apud* Vulanovic, 2007: 400)), o qual é inspirado pela eficiência da máquina.

A eficiência da máquina pode ser definida como a medida da saída útil dividida pela medida da entrada. É energia ou trabalho que é medido em física e engenharia, mas se as gramáticas também são consideradas “máquinas”, então a entrada e a saída linguísticas devem ser medidas. A informação transmitida pela gramática é vista como saída da máquina e as formas e regras como entrada da máquina⁷² (VULANOVIC 2007, p. 400).

Nesse sentido, o autor parece se utilizar de uma metáfora clássica para o problema da complexidade linguística: se se tem um dado hardware que trabalha em dada velocidade e com grande capacidade de armazenamento (ainda que limitado), bem como um bom ganho em “softwares” já disponíveis de série (aparato biológico) para processar toda e qualquer língua natural, logo a complexidade pode variar, mas não o “hardware”, ou seja, ainda que haja nuances de complexidade entre as línguas, seus usuários (ou “hardwares”) são rigorosamente iguais. Uma prova disso seria a rápida aquisição de L1. Por outro lado, há tempos intuitivamente se percebe diferentes dificuldades para se aprender uma outra língua, seja o vernáculo, seja uma língua estrangeira. Nesse sentido, um tipo de aprendizado de L2 em adultos seria razoavelmente defensável para tentar medir complexidade, ou mesmo, em outros parâmetros, aquisição de falantes L1.

Vulanovic (2007) acusa Dahl (2004) de somente observar um lado bem pequeno do que seja a língua, i.e., a informação. Ao passo que também critica McWhorter (2001), e trabalhos no formato que só veem a complexidade nas estruturas, argumentando que suas descrições são falhas e também que não contemplam a língua em sua inteireza. Ao contrário, ele propõe medir complexidade linguística, baseada numa proposta formalizada por Dik, a qual engloba, ao mesmo tempo, a estrutura das línguas em questão, bem como sua semântica e, até mesmo, sua pragmática. O modelo dikiano,

⁷¹ No original: “Simply put, a grammar is less complex, and at the same time more efficient, if it conveys more information (more meaning) with fewer forms and rules” (Vulanovic, 2007: 399).

⁷² No original: “Machine efficiency can be defined as the measure of the useful output divided by the measure of the input. It is energy or work that is measured in physics and engineer-ing, but if grammars are also considered “machines,” then linguistic input and output have to be measured. The information conveyed by the grammar is viewed as machine output and the forms and rules as machine input” (Vulanovic 2007, p. 400).

já conhecido, toma as partes de uma sentença e atribui valores a partes desta sentença, incluindo neste cálculo valores para operações semânticas e pragmáticas, mesmo considerando ambiguidades. Dessa forma, tais valores tendem a uma variabilidade maior do que os valores puramente estruturais.

Nesta seção pudemos ver as principais propostas de se medir complexidade pelo viés do uso, que aliás é um aspecto linguístico muito explorado para se medir e observar a complexidade linguística. Embora isso, há entre linguistas e teóricos diferentes pontos de vista de como a complexidade pode ser medida, alguns desses pontos de vista, inclusive, são bastante divergentes.

3.2.3 Pelo viés da aquisição

É sabido desde os anos 1980 sobre como se dá a aquisição de uma língua, que algumas gramáticas (ou partes de gramáticas) levam mais tempo para se adquirir do que outras (Hawkins 2014).

Ao adquirir uma L1, um “erro” desses “falantes emergentes” pode tender a uma normatização natural. Ou seja, os “erros” que crianças cometem ao adquirir a língua (L1) podem funcionar como um fator de mudança linguística nessa transmissão. Daí, estes “erros” podem se espalhar pela sociedade, dado o movimento de “normatização” espontânea. As crianças, com seus “erros” (somente aqueles que são transmitidos, obviamente) tornam a língua mais ou menos complexa.

As mudanças cumulativas na língua parecem simples, mas, talvez dado seu efeito cascata, não resultem numa mudança simples, nem mesmo numa mudança ordenada. A exemplo disso, pensemos na redução de paradigmas morfológicos verbais no PB que têm acarretado diversas mudanças na estrutura dessa língua. Nesse sentido, é mais fácil enxergar regras em fenômenos ordenados. Como disse jocosamente Deutscher (2009): foi uma espécie de preguiça ou relaxo dos gramáticos ao não estudarem suficientemente as exceções às regras gramaticais. O mesmo pensa Hawkins (2014) ao observar gramáticos e mesmo linguistas que se detiveram a problemas descritivos outrora vistos apenas pelos prismas das regularidades:

Dan Slobin (1982) comparou as crianças que adquiram inglês, italiano, servo-croata e turco em 4 grupos etários. Basicamente ele descobriu que quanto mais iconicidade de significado de forma, mais rápida é a aquisição. Assim, as crianças turcas aprenderam a morfologia rapidamente, mas os aspectos da sintaxe (por exemplo, as cláusulas relativas) são relativamente atrasados⁷³ (HAWKINS 2014: 39).

⁷³ No original: “Dan Slobin (1982) compared children who acquire English, Italian, Serbo-Croatian and Turkish in 4 age groups. Basically he found that the more iconicity of form meaning, the faster the acquisition. Thus, Turkish children

Na sua opinião, a língua inglesa é mais simples do que outras línguas germânicas devido à aquisição inglesa L2 massiva por escandinavos no período do inglês antigo. Como resultado, o inglês perdeu: i) marcação de gênero nos artigos; ii) a maioria da sua morfologia de caso (na verdade só sobrou o um resquício de genitivo) (Hawkins 2014).

A questão da aquisição de segunda língua revela muito do que há de complexidade entre as línguas, tanto que, para vários autores, o papel da aquisição de L2 é um parâmetro importante para medir complexidade. Já em 1997, Rod Ellis (*apud* Kusters, 2013) notou que falantes de L2 tendem a “importar” as dificuldades de línguas estrangeiras a aspectos de sua própria língua, como, por exemplo, tempos verbais, marcações pessoais nos pronomes etc. Esses “erros” cometidos por estrangeiros ao tentar se comunicar em uma L2 são dados importantes que podem ser usados para medir o nível de complexidade entre as diferentes línguas envolvidas num dado processo desse tipo. O que esses estrangeiros fazem, de fato, é tentar “traduzir” as dificuldades apresentadas em suas próprias línguas ante as dificuldades encontradas na L2. Dessa forma, para os estudos da complexidade, um “falante ideal” de L2 que almeja apenas uma boa comunicação numa língua estrangeira seria um bom avaliador de complexidade entre duas ou mais línguas. Além disso, sabemos que as transmissões linguísticas – tanto no aprendizado de L2, como mesmo na aquisição de L1 – são fatores importantes para mudanças linguísticas e, em consequência disso, para a complexidade, seja em ganho ou perda.

Aliás, nos estudos de aquisição, desde pelo menos os anos 1990, há parâmetros e ferramentas para se medir complexidade. Lea Nieminen (2009), por exemplo, propõe uma métrica de complexidade justamente se valendo de duas ferramentas advindas dos estudos de aquisição de língua, que vimos brevemente antes. São elas o MLU, comprimento médio do enunciado (sigla inglesa de *Mean Length of Utterance*) e IPSyn, índice de produtividade sintática (sigla inglesa de *Index of Productive Syntaxis*)

Segundo Niemenen: “Uma coisa que os estudos de aquisição apresentam claramente é que a complexidade é uma propriedade mensurável da linguagem, e diferentes escalas de medida são usadas para executar essa tarefa.”⁷⁴ (2009: 176). A premissa básica da MLU é muito simples: o número de morfemas produtivos produzidos em 100 enunciados inteligíveis consecutivos é calculado e dividido pelo número de enunciados para determinar o número médio de morfemas por enunciado (Nieminen, 2009: 176). Nesse sentido, segundo a autora, aumentando o número de morfemas, aumenta-se também a quantidade de informação semântica do enunciado. De qualquer forma, o MLU só

learned morphology quickly, but the syntax aspects (for example, the relative clauses) are relatively backward” (Hawkins 2009: 39).

⁷⁴ No original: “One thing that acquisition studies present clearly is that complexity is a measurable property of language, and different scales of measurement are used to execute this” (Niemenen 2009: 176).

descreveria a semântica, e, por esse motivo, tal ferramenta deve ser usada associada à IPSyn, que é justamente a ferramenta para descrição sintática dos enunciados das línguas naturais. Aplicados os testes a falantes tanto de L1, quanto de L2, a autora argumenta que os testes revelam que crianças aprendem tardiamente estruturas mais complexas. No entanto, testes para medir o tempo que cada criança demora para adquirir tais e tais estruturas complexas de cada língua, ainda não foram sistematicamente realizados. Além disso, devemos ter em mente que falantes de L2 podem ser agentes de simplificação das línguas, bem como crianças em período de aquisição podem ser fatores de complexificação das línguas.

Segundo Szmrecsanyi & Kortmann (2012), o trabalho comparativo mais recente sobre a complexidade da linguagem assume uma perspectiva funcional-tipológica, ou sociolinguística, da linguística de contato, e adota metodologias empíricas. Ainda segundo esses autores, as perguntas mais relevantes hoje em dia para o campo da complexidade linguística são: de onde vem a complexidade linguística? Como os falantes a adquirem? O que a complexidade linguística pode nos dizer sobre a natureza do conhecimento linguístico? Os autores defendem que boa parte da medida de complexidade linguística se deve também a fatores sociais, principalmente quando levamos em conta a aquisição de L2. Para comprovar sua hipótese, os autores desenvolveram experimentos que mostraram que, de fato, aprendizes de L2 tendem a simplificar a língua alvo, regularizando algumas formas e eliminando algumas outras. Essa dinâmica, contudo, é balanceada pelas motivações dos aprendizes de L2 e também por fatores sociais mais amplos que podem levar a uma maior ou menor simplificação do sistema linguístico sendo aprendido.

3.2.3.1 Método CAF e a Complexidade a serviço da SLA

Dado que um dos principais objetivos dos estudos sobre aprendizagem⁷⁵ de segunda língua, ao menos no domínio da linguística aplicada, é a proficiência do falante estrangeiro, alguns parâmetros para atingir tal objetivo são necessários. Desde os anos de 1970 há um parâmetro que vem sendo desenvolvido, a que se passou a chamar de CAF (*Complexity, Accuracy and Fluency*). Justamente, na SLA (*Second Language Acquisition*) tal parâmetro serve para “medir” a proficiência de um falante frente ao desafio de aprender – ou adquirir – uma segunda língua. Dizemos que tal

⁷⁵ Chama-se de aquisição apenas o aprendizado em determinada fase da infância de um falante, quando ele começa efetivamente a usar a língua. É natural, serve principalmente para interagir com outros e com o ambiente. Aquisição se dá de forma inconsciente e informal, ou seja, uma assimilação natural. O resultado disso é o desenvolvimento de uma habilidade, em parte inata, do ser humano. Ao contrário, a aprendizagem é feita depois de adquirida a primeira língua ou em muitos casos, quando não adquirida completamente, mas requer instrução formal, como gramáticas escolares, e esforço do aluno. Não é natural. Geralmente na aprendizagem o falante se vale de sua própria língua para, por exemplo, fazer traduções do idioma estrangeiro. (Krashen (1981), *apud* Eckert e Frosi 2015).

“medida” vem se desenvolvendo nos últimos anos precisamente porque vem se aliando aos modernos estudos da complexidade linguística (Szmrecsanyi & Kortmann, 2012 e Housen, Kuiken & Vedder, 2012).

Housen, Kuiken & Vedder (2012) propõem as seguintes questões: o que faz com que um falante de L2 seja proficiente numa língua estrangeira? De que maneira, para um falante de L2, se atinge proficiência de modo mais eficaz? Dessa forma, novamente, percebe-se que a “proficiência” em L2 é essencial para a SLA. Nesse sentido, para tentar sanar as questões acima propostas, chegou-se à conclusão de que a proficiência não depende apenas de um aspecto linguístico – qual seja, o ensino ou o aprendizado individual do falante de L2 – mas que tal característica é de natureza multicomponencial. Os mesmos autores acreditam que tal empresa seja melhor aproveitada em observância a três aspectos inerentes à língua em seu estado de aquisição: *complexidade*, *precisão* e *fluência*, ou CAF na sigla inglesa.

Dito de outra forma, o método CAF presta-se a conferir se houve aquisição completa, a proficiência de L2, e isso se torna possível “medindo” de algum modo sua complexidade:

Nos últimos anos, a tríade CAF surgiu como um complemento notável a outros modelos de proficiência estabelecidos, tais como o modelo de quatro habilidades tradicionais e os modelos sociolinguísticos e cognitivos de proficiência de L2 (e.g. Bachman 1990; Bialystok 1994; Canale & Swain 1980). (HOUSEN, KUIKEN & VEDDER 2012: 3)⁷⁶

Ainda, o método CAF procura suas premissas em dados qualitativos, na medida em que sua pretensão é o ensino e aprendizagem de L2 de modo pleno e satisfatório, e por mais que tal método tenha sido bastante disseminado, há críticas a sua validade, sobretudo a sua validade científica.

3.2.3.2 Outros métodos disponíveis

Segundo Szmrecsanyi & Kortmann (2012), mesmo dentro da SLA, há várias formas de se enxergar e mesmo de medir complexidade linguística, as quais sumariamente apresentamos em seguida:

→ *Complexidade Fonológica* (Nichols, 2009; Shosted, 2006 *apud* Szmrecsanyi & Kortmann, 2012): mede-se aí o tamanho do inventário fonológico, a incidência de fonemas marcados, a distinção total

⁷⁶ No original: “In recent years the CAF triad has emerged as a notable complement to other established proficiency models such as the traditional four-skills model and sociolinguistic and cognitive models of L2 proficiency (e.g. Bachman 1990; Bialystok 1994; Canale & Swain 1980).” (HOUSEN, KUIKEN & VEDDER 2012: 3).

de fonemas suprasegmentais, as restrições fonotáticas e, finalmente, a complexidade máxima de encontros consonantais.

→ *Complexidade Morfológica* (Dammel & Küschner, 2008 *apud* Szmrecsanyi & Kortmann, 2012; Kusters, 2003): segundo os autores, deve-se observar aqui: o escopo da morfologia flexional em uma dada língua ou variedade. Especificamente, o que é alomorfa e quais são os processos morfofonêmicos induzidos por “fatores de perturbação” (Braunmüller 1990 *apud* Szmrecsanyi & Kortmann, 2012), e não apenas na morfologia flexional.

→ *Complexidade sintática* (Givón 2009, Karlsson 2009): quantas regras existem na sintaxe de uma língua (note: quanto mais, mais complexa, parafraseando McWorther, 2007), e o que e quais são os níveis de cláusulas subordinadas e encaixadas (recursividade) permitidos pela língua em questão.

→ *Complexidade semântica e lexical*: Complexidade lexical métrica, ou para avaliar inventário léxico adquirido pela SLA.

→ *Complexidade pragmática*: Segundo Bisang (2009 *apud* Szmrecsanyi & Kortmann, 2012), “complexidade escondida”. Qual a interferência da pragmática de uma língua por parte dos aprendizes desta para adquiri-la?

Em suma, nenhuma dessas técnicas se faz totalmente eficiente se usadas completamente isolada em apenas um aspecto da língua. Da mesma forma, não há viabilidade ou proveito em se medir complexidade segundo esses critérios em apenas uma língua ou mesmo em algumas poucas línguas. Dessa forma, seriam bem aplicados tais critérios de medição de complexidade se comparados a várias e diversas línguas:

Shosted (2006), por exemplo, investiga complexidade morfológica e fonológica em uma amostra de 34 línguas e descobre que não existe uma relação significativa, e não existe troca entre complexidade fonológica e morfológica; Nichols (2009) também explora as complexidades locais numa ampla variedade de línguas e não consegue apontar uma troca. Fenk-Oczlon & Fenk (2008) diagnosticam “efeitos de equilíbrio” entre complexidades locais, mas não na medida em que o *dogma equicomplexional*⁷⁷ poderia ser assumido em todas as realidades e em todas as circunstâncias. Gil (2008) argumenta que isolar as línguas não

⁷⁷ “Dogma equicomplexional” é um termo utilizado por Gil (2008) para esclarecer a antiga noção, ponto pacífico entre quase todos os linguistas, de que há compensações nos sistemas linguísticos, fazendo com que a complexidade linguística se equipare. No entanto, isso é um mito, como estamos tentando mostrar aqui.

mostra necessariamente que há uma simplificação morfológica com maior complexidade em outros domínios. (SZEMRECSANIY & KORTMANN, 2012, p. 3)⁷⁸

Na introdução do livro que esses autores organizaram (Szemrecsaniy & Kortmann, 2012), em sua revisão bibliográfica, citam vários teóricos que lidam com complexidade. Entre estes, citam vários estudos feitos sobre o assunto e nesses, dados impressionantes comprovam, por exemplo, que algumas línguas são menos complexas que outras, e nesse sentido, não há compensações (Shosted, 2006; Gil, 2008; Nichols, 2009). Tais estudos ainda atestam preliminarmente que línguas de comunidades mais isoladas tendem a maior complexidade estrutural, e menor complexidade vocabular (cf. Deutscher, 2014).

Esses autores ainda citam mais duas noções que são importantes para SLA, a *Complexidade de redundância induzida*, que revela por exemplo os graus de dificuldade de um falante de L2 ao se deparar com uma língua que marca de maneira redundante um sintagma nominal, por exemplo, caso do português padrão na marcação do número em nomes. Nesse sentido, Dahl (2008) diz que a redundância garante a fidelidade da informação a ser veiculada pela língua. A *Complexidade de irregularidade reduzida*, utilizada por McWhorter (2007) atesta que línguas que possuem o verbo *ser* com paradigma flexional irregular tendem a ser mais complexas, portanto mais difíceis de se adquirir, a princípio.

Szemrecsaniy & Kortmann (2012) ainda concordam com McWhorter (2007) e com Dahl (2004) quando afirmam que uma língua jovem não pode ser mais complexa do que uma língua antiga, pois não tem os ornamentos de uma língua de idade mais avançada. Ainda, concordam com Trudgill (2001) quando este afirma que um adulto aprendendo uma L2 obviamente fará algum tipo de simplificação linguística, isso porque ele tende a suprimir redundâncias e irregularidades, e tende também a aumentar a transparência da língua⁷⁹. Nesse sentido, Dahl (2004) é incisivo: redundâncias e irregularidades são ornamentos linguísticos, características que apenas surgem com a “maturação” de uma língua. Assim, uma língua jovem – digamos, uma língua crioula ou uma língua de contato, leia-se L2 – não tem tais ornamentos. Note-se que isso afeta diretamente a aquisição de L2. Dito de

⁷⁸ No original: “Shosted (2006), for example, investigates morphological and phonological complexity in a sample of 34 languages and finds that there is no significant correlation, and no trade-off, between morphological and phonological complexity scores; Nichols (2009) likewise explores local complexities in a wide range of languages and fails to obtain a trade-off. Fenk-Oczlon and Fenk (2008: 63) do diagnose “balancing effects” between local complexities, but not to the extent that the equicomplexity dogma could be assumed to hold true under all circumstances. Gil (2008) argues that isolating languages do not necessarily compensate for simpler morphology with more complexity in other domains.” (Szemrecsaniy & Kortmann 2012: 3).

⁷⁹ Isso pode significar tanto o recurso a formas mais analíticas do que sintéticas (ou seja, evitar o uso de, por exemplo, morfemas e recorrer a itens independentes para a expressão de uma certa nuance gramatical, como usar “tinha pensado” por “pensara”) e/ou aversão a morfemas que trazem mais de uma informação sintático-semântica.

outra forma, nesse cenário, é o falante de L2 que promove tal mudança, ou melhor, que deixa de lado tais ornamentos em nome da simplificação exigida pela mera comunicabilidade em língua estrangeira.

Esses autores também mencionam um método de avaliação de complexidade para SLA chamado de *método Freiburg*. Tal método procura aliar noções metodológicas e interpretação de dados de parentescos entre várias línguas. Este método, bastante utilizado por Kortmann (2004 *apud* Szmrecsanyi & Kortmann, 2012) visa: i) elucidar padrões de complexidade em larga escala entre as línguas do mundo; ii) desenvolver uma métrica necessária para julgar os diferentes graus de complexidade. Dois dos primeiros trabalhos utilizando esta técnica (Kortmann & Szmrecsanyi, 2009 e Szmrecsanyi & Kortmann, 2009c, *apud* Szmrecsanyi & Kortmann, 2012) revelaram que existem 46 variedades de inglês, dentre elas, 11 como L2, 15 de crioulos emergentes. Dentro deste método, os autores pontuam a *regra de complexidade ornamental* que consiste em avaliar a incidência relativa no estoque de uma variedade de características que adicionam contrastes, distinções ou assimetrias sem fornecer um bônus comunicativo ou funcional à língua (cf. McWhorter, 2007). Um exemplo disso é que uma língua pode ou não atribuir o par de pronome pessoal ‘ele’/‘ela’ a entidades inanimadas.

Há ainda a chamada *regra da simplicidade*. Ela explica como é possível simplificar a língua em um contato entre falantes. Por exemplo, pode acontecer a perda de pronome reflexivo. Note-se que isso vem ocorrendo no Português Brasileiro atualmente, particularmente em algumas variedades, e é uma tendência bem forte no PB como um todo. Nessa esteira, surge o conceito de *simplicidade na aquisição de L2*: “Como uma métrica de complexidade relativa, esta medida estabelece a incidência relativa no inventário de características de uma variedade que a pesquisa de SLA mostrou recorrer em variedades de interlíngua.” (Szmrecsanyi & Kortmann, 2012, p. 9)⁸⁰, por exemplo: falantes de L2 tendem a não usar marcadores de tempo verbal ou a fazer a regularização de verbos irregulares. Dadas tais distinções teóricas, a pesquisa acima mencionada nas variedades do inglês atestou que o inglês como L1 tende a ser mais complexa, ou seja, tem mais “ornamentos” do que suas cognatas: crioulos do inglês e, em menos incidência, de inglês como L2, que são variedades atestadamente menos complexas.

Outro importante estudo na área de complexidade aplicada à SLA é o trabalho de Hong Han & Man Lew (2012). Nesta obra, (*apud* Szmrecsanyi & Kortmann, 2012), os autores estão procurando parâmetros para definir o que eles chamam de “complexidade aquisicional”⁸¹, procurando elucidar a complexidade no domínio da SLA. Dizem ainda que *complexidade* é um conceito *relativo*. Eles

⁸⁰ No original: “[...] as a relative complexity metric, this measure establishes the relative incidence in a variety’s inventory of features that SLA research has shown to recur in interlanguage varieties [...]” (SZMRECSANIY & KORTMANN, 2012, p. 9).

⁸¹ A noção de *complexidade aquisicional* envolve uma série de valores em aquisição e aprendizado de L1 e L2, desde medidas sintáticas até parâmetros mais finos para se atestar proficiência em uma língua. Alguns dos parâmetros estão aqui expostos, para maiores detalhes, ver Hong Haw & Man Lew (2012).

abordam um conceito bastante intrigante, a *complexidade de desenvolvimento*. Tal noção serve de critério para medir a proficiência de falantes L2 dada sua complexidade: i) tamanho médio de palavras utilizadas por frase/sentença; ii) quantidade de sentenças coordenadas; iii) quantidade de sentenças subordinadas. Para Ortega (2003) (*apud* Hong Han & Man Lew, 2012, p. 194), quanto mais complexa, mais desenvolvida é a *complexidade de desenvolvimento*. No entanto, não necessariamente melhor. Para a mesma Ortega (2003), o desenvolvimento sintático para falantes de L2 significa que as formas linguísticas estão mais apropriadas para a comunicação, mesmo assim não suficientes para medir proficiência. Dessa forma, por si só, medidas de complexidade sintática não são suficientes para provar aquisição plena. Mas os autores vão além, para eles, nenhuma medida (cognitiva, psicológica, linguística etc.), por si só, são totalmente suficientes para medir complexidade ou desenvolvimento de maneira satisfatória.

Os autores ainda mencionam outras noções bastante importantes para os estudos da complexidade em L2, tal como o da *complexidade aquisicional*, que segundo eles consistem nos seis seguintes aspectos:

- i) O que é adquirível (de um modo complexo) por um falante ou grupo de falantes, pode não ser para outro. Complexidade de aquisição não pode ser medida nos mesmos parâmetros que a complexidade linguística, quais sejam: fatores externos, universais etc.;
- ii) A Complexidade de aquisição é estática, ao passo que a complexidade de desenvolvimento, não;
- iii) Complexidade de aquisição, ou o mesmo que aquisição de forma e função;
- iv) Complexidade aquisicional é uma função da interação estabilizada de contingências exógenas e endógenas. Dessa forma, um adequado estudo dela só ocorre se ela estiver ligada com os elementos, tais como o relacionamento entre L1 e língua-alvo;
- v) Evidências da complexidade do desenvolvimento (*developmental complexity*) podem, a princípio, ser coletadas em qualquer estágio do desenvolvimento. Mas o ponto de partida para localizar incidência de complexidade aquisicional só pode ser coletada na amostragem de alunos que foram aprendendo na fase exógena e em condições endogenamente favoráveis, como a experiência ampliada e sustentada com a metalinguagem (livros didáticos, por exemplo). Evidências de qualquer dificuldade na aquisição da L2 podem ser usadas para fazer inferências sobre a complexidade aquisicional;
- vi) Complexidade de aquisição pode ser medida ao longo do tempo, embora isso não precise ser necessariamente feito (pp. 196-197).

Em seguida, os autores relacionam *complexidade de aquisição* com *dificuldade de aprendizado*, ressaltando que são noções distintas, amparados por Miestamo *et al.* (2008), Dahl (2004), Collins

(2009) e Krashen (1982 *apud* Hong Han & Man Lew 2012). Para eles, é mais produtivo na SLA avaliar *dificuldade de aprendizado*, na qual destacam-se as seguintes observações pertinentes:

- a) A *frequência* com que um traço ocorre no input;
- b) *Saliência*: na medida em que a funcionalidade é perceptível no input;
- c) *Valor funcional*: se o recurso é mapeado sobre um significado claro e distinto; e
- d) *Processabilidade* de um traço (Ellis 2006, 2009 *apud* Hong Han & Man Lew 2012).

Diferentemente dos outros autores aqui citados, Hong Han & Man Lew (2012) apresentam um estudo de caso. Trata-se de um professor chinês que há 15 anos lecionava inglês, supostamente proficiente, portanto. Nas aulas – ou seja, num discurso falado, controlado – o professor sempre utilizava uma estrutura nas sentenças inglesas, que em português corresponde a algo como “Não só... mas também”⁸². Os autores apontam que tal estrutura também existe no chinês, utilizada de uma forma bem parecida ao inglês. Nesse sentido, seja talvez por esta razão que o professor chinês utilizava sempre tal estrutura. Dito de outra forma, o professor utilizava sempre essa estrutura porque fazia uma ponte da língua estrangeira que ele está utilizando como sua língua mãe. Poderíamos citar um exemplo sentencial mais próximo. Informalmente, ao ver um diálogo entre dois falantes, um nativo (L1) em língua inglesa e outro brasileiro (L2) falando inglês, constatamos a seguinte sentença: “He brokes the legs”⁸³. Ao ser proferida, o falante de inglês não entendeu plenamente o que o brasileiro falou. Analogamente ao caso do professor chinês, parece que falantes de L2 tendem a “parametrizar” ou aproximar parâmetros de sua língua nativa, “traduzindo-os” à L2.

Os autores comentam ainda a noção de *fossilização*, que é um conceito chave para SLA. Baseado em Long & Crookes (1992) (*apud* Hong Han & Man Lew, 2012, p. 200), os autores adotam as seguintes premissas:

- 1) Unidades linguísticas não são unidades de aquisição, i.e.: unidades que são realmente processadas e adquiridas;
- 2) Complexidade de aquisição é diferente de dificuldade, complexidade é externa, e dificuldade é interna. Complexidade de aquisição pode não ser difícil para certos falantes (Robinson (1996), *apud* Hong Han & Man Lew, 2012, p. 201);
- 3) Estruturas fossilizadas são complexas;

⁸² O exemplo utilizado pelos autores para ilustrar a sentença utilizada pelo professor é “*Although* there are multiple types of complexity, **but** we are going to focus on one type” (HONG HAN & MAN LEW, 2012, p. 200), cuja melhor correspondência encontramos em português é “Não só... mas também”, embora a tradução literal, evidentemente, não seja exatamente a mesma.

⁸³ “He broke his legs” seria a sentença gramaticalmente aceita.

4) Somente em pesquisas longitudinais pode-se provar o que é complexidade de aquisição.

Por fim, os autores definem o que seja aquisição segundo Larsen-Freeman (2001) (*apud* Hong Han & Man Lew, 2012, p. 201), uma relação entre forma, função e sentido, dada nesta ordem: forma > forma-sentido > forma-sentido-função ganhando, desse modo, e nessa medida, mais complexidade. Para Sorace (2005 *apud* Hong Han & Man Lew, 2012), a complexidade de aquisição, ainda que não seja plena, exige do falante não apenas conhecimento sintático, mas a habilidade de coordenar conhecimento sintático com outros conhecimentos.

Nessa seção, vimos de que maneira os linguistas se utilizam do fenômeno da aquisição, tanto de primeira língua, quanto de língua estrangeira, para estabelecerem um parâmetro do que seja medir complexidade linguística. Aliás, a aquisição tem se mostrado um dos vieses mais eficientes para se medir complexidade, que inclusive já é proposto pela linguística aplicada desde os anos 1970.

3.3 Conclusão

Vimos sucintamente neste capítulo algumas propostas para medir complexidade linguística. Embora tenhamos divididas as possibilidades de métricas em três, há no capítulo, diversas e diferentes ideias de como medir complexidade. Também expusemos visões de linguistas de diferentes linhas teóricas, tanto funcionalistas quanto formalistas, que investigam questões sobre métricas de complexidade linguística.

Acerca das contradições sobre as diferenças de complexidade entre as línguas, sobretudo no que tange à sua expressividade, ou seja, se é fato que tudo pode ser dito em qualquer língua, e que, nesse sentido, como diria Jakobson (1979), as línguas se diferem não no que permitem dizer (porque todas permitem dizer tudo), mas sim no que obrigam seus falantes a dizer, também é fato que a defesa de que todas as línguas são igualmente complexas com base em argumentos científicos e evidências verificáveis ainda está para ser feito. Um primeiro passo nesse sentido é certamente nos livrarmos dos preconceitos históricos que sempre estiveram associados a essa ideia, e investigá-la de modo objetivo. Esse é o saldo positivo das pesquisas feitas no final do século passado e começo deste século – vemos crescer rapidamente uma discussão rica, que vem das mais diversas áreas dos estudos linguísticos, na direção de uma definição razoável e explícita de complexidade gramatical.

O próximo passo, segundo argumentamos, é o estabelecimento de uma métrica de complexidade gramatical, que, como vimos acima, também pode vir de diversas áreas de investigação dos estudos linguísticos e mesmo combinar *insights* de algumas delas. Com uma métrica estabelecida teremos condições que permitem afirmar, segundo os critérios de cientificidade do nosso tempo, se,

de fato, as línguas têm ou não todas a mesma complexidade linguística, sem atrelar preconceitos injustificáveis a essa reflexão. Poderemos também dotar a complexidade linguística de poder heurísticos, como mencionamos, pois poderemos detectar se uma dada mudança, com uma dada motivação, levou ou poderá levar a uma maior ou menor complexidade da língua ou do fenômeno linguístico investigado. Seja como for, o caminho que nos levará a alguma conclusão é, por si só, um ganho inestimável aos estudos linguísticos.

Postas as devidas definições do que seja a complexidade e a complexidade linguística, bem como as diversas propostas para medi-la, passamos ao próximo capítulo, no qual procuraremos demonstrar através de dois tipos de organização sintática do português brasileiro, como pode haver diferentes graus de complexidade linguística.

Capítulo 4 – Construções de Tópico no Português Brasileiro medido sob o viés de um modelo de complexidade linguística: um exercício de análise

4.1 Introdução

Expostas as bases do que podemos chamar genericamente de “estudos sobre complexidade linguística”, juntamente com um breve apanhado historiográfico e, ainda, com algumas reflexões de cunho científico-filosófico, procederemos agora a um exercício de demonstração prática do potencial heurístico e explicativo das teorias de complexidade linguística, usando dados do português brasileiro. Para tal empresa, escolhemos um fenômeno particularmente saliente na literatura linguística brasileira, as chamadas construções de tópico (CT). Essas construções, no âmbito do português brasileiro (PB), foram e são extensivamente investigadas por linguistas formalistas e funcionalistas, com os objetivos de descrevê-las, explicá-las, determinar sua origem e contrapô-las à estrutura sintática canônica do português europeu (PE). Contudo, do que pudemos averiguar, as construções de tópico do PB ainda não foram alvo de estudos pautados por questões sobre complexidade linguística.

Sendo assim, discutiremos sobre construções de tópicos encontrados no português brasileiro, sob a ótica dos estudos sobre complexidade linguística (CL). Como adiantamos, as construções de tópico (CT) do PB, tais como as atestadas em Pontes (1987):

(1) A Belina o Hélio levou pro concerto

, ou aquelas atestadas por Callou *et al* (2002):

(2) Feijão eu não gosto

, dentre outras, foram já bastante discutidas e mesmo descritas em PB, sob diversas perspectivas teóricas. Por outro lado, as CT foram também bastante estudadas e descritas em várias outras línguas (Hajicova *et al* 1998; Neeleman & Vermeulen 2012, Roberts 2012, entre outros), sendo mesmo fruto de discussões tipológicas, a ponto de alguns pesquisadores elevarem tais construções a tipos linguísticos definidores⁸⁴ (Li & Thompson 1976; Ross 1967 *apud* Pontes 1987). Embora devidamente estudadas, as CT parecem ainda carecer de mais investigações, visto que não se

⁸⁴ Chamo aqui de “Tipos Linguísticos definidores” traços tão particulares e universais de algumas línguas que as definem e as diferenciam. Li & Thompson (1976) dizem mesmo que as línguas de proeminência de tópico apresentam esse traço de modo tão marcante, que os diferenciam das chamadas línguas de sujeito-predicado.

encontra, entre os linguistas, consenso sobre vários aspectos descritivos e teóricos relevantes, sobretudo quando consideramos os dados do PB, caso para o qual a literatura acerca do assunto é extensa.

Pontes (1987), por exemplo, toma como parâmetro a classificação tipológica de Li & Thompson (1976), que ainda hoje é referência em se tratando de construções de tópico. Na verdade, Li & Thompson (1976), baseadas em Ross (1967, *apud* Pontes 1987), tomam as construções de tópico como classificação definidora de tipo de língua. Assim, há em Li & Thompson (1976) línguas que se organizam entre sujeito-predicado em oposição a línguas que se organizam em tópicos.

Assim, a escolha por utilizarmos as construções de tópico em PB para nosso exercício se justifica por (i) abundância do fenômeno em PB, (ii) abundância de estudos sobre tal fenômeno e (iii) uma suposta visibilidade que teremos ao confrontar construções de tópico a construções do tipo sujeito-predicado (CSP), ou seja, para nossa intenção aqui – procurar mostrar que uma dada construção pode ser mais complexa do que outra – o fenômeno de CT é exemplar, no sentido de se apresentar como um fenômeno que evidencia uma operação sintática marcada, com relação à construção canônica de sujeito-predicado. Além disso, sentenças como

(3) A Sarinha está nascendo dentes (Pontes 1987)

não são detectadas no PE, o que leva a linguistas como Galves (1988) e Araújo (2006) a defenderem que o PB se distingue do PE com relação às construções de tópico.

Ainda que encontremos construções de tópico em PE, já nos séculos XVIII e XIX, tais como as atestadas por Araújo (2006):

(4) “O meu D. João VI mandei-o encadernar na Imprensa Nacional

(5) Provas tens tu lá, e eu é que as reclamo revistas

(6) Respeito a casas, sei de duas que neste momento estão devolutas” (ARAÚJO, 2006, p. 19)

–, a literatura pertinente, desde Epifânio Dias (1918) trata tais exemplos de CT simplesmente como inversões, ou como “anacoluto”, uma figura de sintaxe, ou seja, um fenômeno idiossincrático. Diferentemente, no PB, atualmente, linguistas têm tratado as CT como próprias da língua e não como apenas inversões estilísticas.

Na verdade, e nossa intenção é que fique claro na primeira parte deste capítulo, o que genericamente chamamos de “tópico”, “construções de tópico”, “topicalizações” e cognatos, fazem parte de uma gama grande de fenômenos linguísticos que têm em comum o fato de se colocar, geralmente, a informação dada em posição de destaque do enunciado, que é fronteando-a com relação

ao restante de um dado proferimento. Dessa forma, este capítulo está organizado da seguinte maneira: além desta primeira parte introdutória, na segunda parte buscaremos elaborar uma definição operacional para os objetivos desta tese, ainda que não definitiva ou consensual, do que representam na linguística as chamadas construções de tópico. Na terceira parte, procederemos aos exercícios segundo alguns parâmetros postos nos estudos sobre complexidade linguística, e encerraremos este capítulo com uma breve conclusão. Antes, discutiremos descrições sobre CT expostas na literatura linguística proposta por vários autores e suas mais diversas orientações teóricas, sem o intuito de sermos exaustivos.

4.2. Definição do que sejam as CT

4.2.1 Articulação Tema-Rema, Ruptores e outras noções

Definido nosso objeto, é lícito agora que expliquemos a) o que exatamente vamos entender por construção de tópico (CT) e b) de que maneira os estudos sobre complexidade linguística podem contribuir para as discussões sobre esse tipo de construção em PB. Sobre a primeira questão, nos valeremos de parte da vasta literatura em PB, que vem sendo desenvolvida desde os anos 1980 no Brasil, cuja pioneira, Eunice Pontes (1987), nos fornecerá boa parte da definição e algumas sentenças que comporão os exemplos a serem utilizados nos posteriores exercícios. Sobre a segunda questão, ela ficará mais clara quando procedermos aos exercícios de análise propriamente. Aliás, é importante que se registre aqui que, para fins de *corpus*, todas as sentenças aqui utilizadas foram retiradas das leituras que estão devidamente catalogadas. Assim como Pontes (1987), fiamo-nos na já clássica noção tipológica de Li & Thompson (1976), que será exposta na sequência. Embora aparentemente datada, tal definição ainda não foi propriamente contestada.

Procuraremos agora estabelecer uma noção, ao menos provisória para este trabalho, do que entendemos como construções de tópico. Obviamente esta não é uma noção definitiva. Ao contrário, pretende-se aqui, mais detidamente, delimitar um conjunto de fenômenos para fins práticos, próprio do fazer científico e, para nosso caso particular, apto a ser investigado por uma teoria de complexidade linguística, do que propriamente polemizar as classificações e noções outrora definidas na já vasta literatura sobre o assunto.

Antes de falarmos das construções de tópico propriamente, convém falarmos sobre a questão da ordem nas línguas naturais, sobretudo no português, nosso objeto de estudo. Primeiramente vamos discutir sobre a articulação tema-remata, depois falaremos brevemente sobre os chamados “ruptores”.

Segundo Ilari (1992), baseado nos estudos funcionalistas da Escola de Praga, a ordem dos constituintes de uma sentença não deve ser apenas pensada a partir do ponto de vista de seus sujeito e predicado ou subpartes similares. Para Ilari, e já para as primeiras elucubrações gramaticais dos gregos, a questão se colocava segundo a ordem das informações pertinentes no enunciado. Dessa forma, tem-se a classificação de ordem na sentença segundo a informação, se dada ou nova. A partir disso, fala-se em articulação tema-rema (ATR), em que *tema* é a informação dada, ou velha; e *rema* é a informação nova, dentro da dinâmica funcional e informacional das línguas naturais. Também pode-se pensar na ordem gramatical em oposição a uma suposta ordem psicológica:

Herman Paul teria sido primeiro a constatar que a sintaxe nem é a única maneira possível de classificar sentenças. Sua noção era mais voltada para o “psicológico”. Assim, para Paul, havia o Sujeito e o Predicado, mas também o Sujeito psicológico e o Predicado psicológico, que não eram propriamente os mesmos elementos a que a tradição apontava. Portanto, para ele, parece que havia, então, a oposição: a análise “gramatical” e a análise “psicológica”. (ILARI, 1992, pp. 12-13).

Lembra Ilari (1992) também, que o que ajuda a definir a relevância da informação na dinâmica comunicacional é a prosódia. Mais adiante veremos que, em alguns casos, em última instância, o que define se um constituinte é tópico ou sujeito são justamente critérios prosódicos mais do que a ordem linear superficial. Aliás, em mandarim, língua de tópico, além de marca morfológica própria, o tom é fundamental para determinar as CT.

Além disso, o que nos aponta Ilari (1992) é que há uma estrutura subjacente à ordem gramatical canônica, o Tema-Rema – que alguns linguistas chamam também de Tópico-Comentário – que aponta para a referencialidade informacional dos constituintes de uma sentença, se dado (Tema ou Tópico), ou se novo (Rema ou Comentário). A partir disso, pode-se pensar que tanto uma estrutura sujeito-predicado, quanto uma estrutura de tópico carregam em si a informação tema-rema em alguma configuração sintática possível.

Já em Tarallo, Kato *et al.* (1991) levanta-se a hipótese de que haja em PB construções próprias desta língua, e que tornam, portanto, o PB diferente do PE, as quais apresentam comportamento controverso pois não seriam comuns nem ao PE nem à história da sintaxe portuguesa. Isso porque há algumas construções, sintagmas ou mesmo itens lexicais, que não se comportam nem bem como complementos de uma sentença, nem bem como adjuntos. Esses elementos, por romper a ordem canônica da sentença – isto é: sujeito, verbo, objeto (SVO) – são por isso chamados de “ruptores”. Um desses casos de ruptores, justamente, são as chamadas construções de tópico, porque ficam alojadas antes do sujeito, mas não o compõe. Como ilustração, se considerarmos o modelo teórica

conhecido como “princípios e parâmetros”, os tópicos ficariam entre os níveis IP e CP, justamente onde o dispositivo discursivo se encontra.

Tarallo, Kato *et al.* (1991) propõem, então, que as CT são uma espécie de “ruptor”, cuja relação sintática com sua sentença é diferenciada. Precisamente, o que eles propõem é que tais construções não mantêm qualquer relação sintática direta com suas sentenças comentário, notadamente o verbo. Mas mantêm sim relação semântica e discursivas. Num quadro teórico mais robusto, pode-se perceber que existem algumas relações mais profundas entre complementos e seus núcleos, entre adjuntos e seus núcleos, e até mesmo constituintes sem núcleos bem definidos.

Dito isso, retomamos as construções de tópico propriamente. Roberts (2012) chama a atenção para a ambiguidade dos termos “tópico”, “topicalização” e cognatos. Há inúmeros usos para tais termos, tanto do mais simples exercício de escrita, passando pelos termos utilizados pela sintaxe tradicional ou pelo entendimento do que sejam tópicos pelo viés funcionalista, até aos mais sofisticados termos que a linguística formal, notadamente a gramática gerativa em suas versões mais recentes, fazem das noções. Assim, tópico pode ser tanto o assunto de um texto, dito no senso comum ou numa aula de ensino médio sobre a prática de escrita, quando uma noção digna de núcleo de seus especificadores e complementos nos termos de uma sintaxe cartográfica. Seja de uma forma ou de outra, o consenso dita que “tópico”, num contexto mais sintático, é sempre aquele elemento ou termo que se coloca em destaque num discurso ou numa sentença, geralmente na primeira posição desta. Do ponto de vista funcional-discursivo seria propriamente “(...) tópico discursivo (...), categoria abstrata e analítica, com a qual se opera na descrição da organização tópica de um texto” (Jubran 2006, p. 91). Ou seja, é o assunto pelo qual se inicia um turno do enunciado:

O tópico decorre [...] de um processo que envolve colaborativamente os participantes do ato interacional na construção da conversação, assentada em um complexo de fatores contextuais, entre os quais as circunstâncias em que ocorre o intercâmbio verbal, o grau de conhecimento recíproco dos interlocutores, os conhecimentos partilhados entre eles, sua visão de mundo, o *background* de um em relação ao que falam (JUBRAN 2006, p. 90).

Pioneira no assunto em PB, Eunice Pontes (1987), toma como base a também pioneira classificação de Li & Thompson (1976). Segundo esses linguistas, há quatro formas de classificar as línguas naturais segundo critérios que estariam entre o sujeito-predicado e tópico-comentário. Dessa forma, considerando os dois primeiros tipos, tem-se línguas a) cuja proeminência é da sujeito-predicado, b) línguas cuja proeminência é de tópico, dessas, as primeiras teriam seus parâmetros concentrados na estrutura gramatical, enquanto as línguas de tópico teriam seus parâmetros concentradas no discurso. Kato (1989) e Galves (1987 *apud* Tarallo, Kato *et al* 1991) preferem dizer que estas línguas, cujas construções proeminentes são de tópico, têm sua orientação voltada a características semântico-

discursivas, isso quer dizer que, segundo as autoras, uma ferramenta teórica que privilegia a estrutura sintática, por exemplo, não teria respostas conclusivas para explicar o que é uma construção de tópico, visto que entre o SN Tópico e a sentença comentário a qual ela é ligada – tanto pela informação, quanto pela sua semântica – não encontraremos relações sintáticas propriamente. Ainda haveria, segundo os mesmos Li & Thompson (1976), dois outros tipos de línguas, c) aquelas cuja proeminência é tanto de sujeito-predicado, quanto de tópico e d) aqueles sem proeminência alguma. Dessa forma, Pontes (1987) já afirmava que o PB, diferentemente de sua cognata europeia (PE), estaria modificando seu status tipológico segundo os parâmetros de Li & Thompson (1976): de uma língua de proeminência sujeito-predicado (tipo a) para uma língua cujas ambas estruturas são proeminentes (tipo c).

Nesse sentido, o PB, ao apresentar tais estruturas, ou estaria mudando seu curso tipológico, ou simplesmente estaria seguindo seu curso histórico, como outras línguas indo-europeias, tais como o italiano e o francês. O problema dessa última afirmação é que o que se chama de “tópico” nessas duas línguas está quase sempre associado com sentenças encaixadas, sobretudo às relativas, o que não acontece necessariamente com o PB (Callou *et al.* 2002): “(...) o pronome e a categoria vazia encontram-se muitas vezes, não em distribuição complementar como no italiano, mas em variação livre no mesmo contexto [...] Aqueles livros da coleção SS, eu comprei um deles por US\$ 40 [...] Aqueles livros da coleção SS, eu comprei um por US\$ 40” (p. 316). Assim, dada sua característica pouco comum entre as demais línguas românicas, foi a saída mais razoável dizer que as CT, embora sejam semanticamente pertinentes à sentença a que lhe é associada, não apresenta relações estruturais com esta, ou seja, é uma estrutura não-argumental (Galves, 1995 *apud* Callou *et al.* 2002). Novamente, toda essa operação de estrutura pertinente, mas não argumental à sentença, não é verificada em outras línguas próximas, conforme as já citadas línguas francesa e italiana.

Modesto (2004), por exemplo, tem uma proposta mais ousada no nível sintático. Ele mostra possíveis associações entre as CT do PB e estruturas próprias do chinês e do finlandês, o que nos faria crer numa saída tipológica. Assim, as propostas dos estudos sobre complexidade linguística, mesmo ainda não aplicadas a esse tipo de fenômeno, parecem uma abordagem promissora, pois ultrapassariam os limites dos estudos sintáticos, semânticos e mesmo prosódicos sobre o assunto. Ou seja, não se observou ainda se as CT tornam ou não a ordem de uma língua mais complexa. Vários estudos apontam que com a perda de desinência verbal e com a deterioração do sistema pronominal, especialmente os pronomes clíticos, o PB tende a se tornar uma língua de tópico. Assim, ao procurarmos medir a complexidade desta ordem, procuraremos também lançar luz a tal discussão já longa.

De maneira mais simplória, poder-se-ia afirmar que uma CT nada mais é do que um item deslocado à esquerda, um adjunto, por exemplo. Decorre daí o seguinte:

(7) Bate sol nessa casa

, que deslocando seria:

(8) Nessa casa bate sol.

Acontece que existem outras sentenças cujo significado aparentemente se aproxima a essa, como:

(9) Essa casa bate sol

, sem a presença da preposição. Ora, é natural que em um deslocamento, ao menos num primeiro olhar, haveria de vir acompanhada a preposição, o que não se observa neste e em inúmeros casos análogos.

Estudos como os de Pontes (1987) sugerem que não há em “Essa casa bate sol” propriamente um deslocamento, mas sim que o sintagma “essa casa” seja tópico da oração comentário “bate sol”, cuja valência, nessa configuração, não exige um sujeito. Por outro lado, estudos mais recentes, com base no modelo de princípios e parâmetros (Galves, 1998), sugerem que há sim deslocamento do sintagma “nessa casa”, mas a preposição desaparece:

A ausência ou presença de uma preposição não está condicionada ao constituinte à esquerda estar ou não vinculado a um papel temático do verbo, uma vez que só se verifica com o complementarizador de um nome, não estando, pois, ligado a um papel temático atribuído ao verbo. Igualmente os adjuntos, com ou sem cabeça, não têm também seu papel temático atribuído pelo verbo, por serem eles não-argumentais [...] Os tópicos sem cabeça podem ter ou não o seu papel temático atribuído pelo verbo” (LEITE *et al.* 2002, pp. 325-326)

Assim, as preposições, outrora nucleares, perdem seu sentido, dando espaço para que um nome seja o núcleo do sintagma que prefigura, agora movido, como CT da sentença comentário. A mesma relação de não-tematicidade entre CT e sentença comentário, afeta a preposição, agora um sintagma nominal ao ser movido⁸⁵.

Essa e outras construções já foram notadas desde, pelo menos, fins do século XIX, e não apenas no PB. A gramática tradicional tratava-as como desvios da norma ou, quanto muito, figuras

⁸⁵ Passos e Fonseca (2011) chamam o fenômeno de “Topicalização Selvagem” ou de “construção de tópico com função oblíqua sem preposição”, como veremos na classificação adiante, tal classificação se faz, justamente, pelo comportamento anômalo da preposição, ou melhor dizendo pela perda da preposição nesses casos.

de linguagem, aliás como sempre faz quando se depara com fenômenos que não se enquadram na tradição prescritiva (Leal, 2014). É o caso do chamado anacoluto: “[...] pôr no princípio de uma cláusula (ou membro da cláusula) sem ligação gramatical, a designação do objeto, a respeito do qual vem depois um asserto” (DIAS, 1959, p. 359 *apud* Pontes 1987, p. 52), ou o que Said Ali chamara de pleonasma:

Colocando-se no princípio da oração um complemento expresso por substantivo, ou palavra substantivada, e pronuncia-se este complemento com ênfase seguida de pequena pausa, é costume repeti-lo junto ao verbo da oração. (SAID ALI, 1964, p. 219).

Da citação de Said Ali é interessante notar que ele consegue descrever com clareza até mesmo a prosódia das CT, muito antes de estudos linguísticos mais especializados virem a lume.

Dadas tantas explicações para se precisar as CT, muitos linguistas acreditam que apenas com critérios sintáticos ou prosódicos não seja possível descrever de maneira satisfatória as CT. Desde Pontes (1987) já se vem apelando para critérios da ordem discursiva para explicar as construções de tópico. Mesmo no modelo da cartografia sintática há um nível, o TopP, que se aloja entre os níveis IP e CP (Negrão, 2017, *in* Fiorin, 2017, pp. 71 e ss.). Há linguistas que tratam as CT como, por exemplo, sujeito do discurso, em oposição ao sujeito gramatical (Berlink, Duate e Oliveira, 2009 *apud* SILVA *et al.* 2017). Segundo Pontes (1987):

A relação entre SN inicial e a sentença seguinte é uma relação de discurso, uma vez que a sentença seguinte é sempre um comentário sobre o SN inicial que é o tópico. É a justaposição do SN com a sentença que cria o elo semântico entre elas. (PONTES 1987, p. 98).

Na verdade, a autora acredita que há um *continuum* entre sintaxe e discurso (Pontes, 1987, p. 102), o que anos mais tarde viria a ser hipotetizado como níveis na sentença, conforme comentamos acima.

Costuma-se atribuir a mudanças próprias do português brasileiro a eclosão das CT. Entre os vários candidatos que podem estar por trás do surgimento das CT tal como elas funcionam no PB, é sempre lembrada a erosão da morfologia do PB e também o enfraquecimento de uso de pronomes clíticos; esses fenômenos levaram a uma nova estrutura que preencheria os papéis gramaticais antes expressos pela morfologia e por pronomes clíticos. Como essas mudanças não ocorrem no PE, é muito possivelmente por essas razões não se encontram CT como as do PB no PE.

Como mencionamos, além da queda dos clíticos, é atrelado ao aumento de CT em PB a erosão do sistema pronominal da nossa língua:

As construções de tópico-sujeito, encontradas no PB, merecem comentário à parte. Uma hipótese para a sua penetração no sistema é a simplificação dos paradigmas flexionais, favorecendo a presença de um sujeito pleno, sempre acompanhando um verbo. Assim, além das construções de [Tópico Sujeito] com referentes definidos ((6) *você*_i; eu *te*_i pego. (inq. 20 / década de 90)), começam a proliferar sujeitos não referenciais preenchidos com elementos referenciais nominais (locativos) ou pronominais ((18) aquilo *você* ficava um monte de gente conversando (inq. 11 / década de 90) (ORSINI, 2003, p. 8).

Dito isso, exporemos alguns traços a que Pontes (1987) atribui para que sejam condições de existência de CT em PB:

- a) O tópico sempre é definido;
- b) O tópico não precisa ter relações seletivas com o verbo;
- c) O verbo determina o sujeito, mas não o tópico;
- d) O tópico determina o papel funcional da sentença, o sujeito, nem sempre;
- e) Não há concordância do tópico com o verbo da sentença;
- f) A posição inicial na sentença, está ligada à sua função discursiva;
- g) Tópico não tem dependência gramatical com a sentença.

(Pontes, 1987, p. 86, adaptado).

Aliada a esta informação, Pontes (1987), bem como Modesto (2003), Tarallo e Katto *et al.* (1991), entre outros linguistas levantaram a hipótese de que o PB estaria mudando sua tipologia para uma língua de tópico e sujeito. A fim de verificarmos tal informação, exporemos uma série de traços das línguas de tópico, adaptados de Li & Thompson, para que possamos pensar o caso do PB:

- a) As passivas são marginais;
- b) Apresenta sujeitos vazios;
- c) Duplo sujeito – sendo um gramatical, outro discursivos;
- d) Controla co-referência, e não o sujeito que o faz;
- e) Não restrição sobre qual constituinte pode ser tópico;
- f) São sentenças básicas;
- g) Dificuldade de determinar o sujeito da sentença.

(Li & Thompson, 1976).

Dadas as noções de Li & Thompson (1976), nota-se que o PB atualmente possui algumas dessas características. Dessas, destacaríamos o duplo sujeito, já há muito descritas em gramáticas

normativas, conforme mostramos aqui; a característica de a CT se construir sob a sentença básica de ordem SVO; e a dificuldade de determinar o sujeito da sentença, justamente por todas as características que temos aqui apontados sobre as CT.

Feito esse reconhecimento geral do que seja uma construção de tópico em PB, procuraremos, na próxima seção, esmiuçar uma classificação baseada na literatura sobre o assunto.

4.2.2. Tipos de Tópicos correntes na literatura – uma classificação

É interessante se notar, mais uma vez, que o fenômeno a que chamamos genericamente de tópico, é descrito na literatura de diversas formas: deslocamento à esquerda, duplo sujeito, sujeito do discurso, tópico discursivo, ruptores, anacoluto, adjuntos especiais, dentre outros. Ao contrário dessa nomenclatura, para fins descritivos neste trabalho, tomaremos como Tópico ou Construções de Tópico todo constituinte, geralmente um SN⁸⁶, à esquerda de uma sentença, cuja relação gramatical (de concordância com o verbo, por exemplo) não se faça necessariamente obrigatória, mas que ainda assim mantenha relação semântica com a sentença-comentário.

Uma das primeiras distinções que se encontra é a diferença entre topicalização e deslocamento à esquerda, que na verdade, se distinguem por um detalhe: a presença ou não de um pronome-cópia. Pontes (1987) evoca a distinção entre topicalização e deslocamento à esquerda conforme formulada por Ross (1967, *apud* Pontes, 1987, pp. 51-53). Segundo este linguista, ambas operações se relacionam com o fato de haver deslocamento à esquerda de um sintagma, seja ele um SN ou um SP. Desse modo, um deslocamento que não deixa vestígio é tratado por Ross (1967) como Topicalização (“[Feijão] eu não como \emptyset ”) e Deslocamento à Esquerda (DE) se refere àquele constituinte que, ao ser movido à esquerda da sentença, deixa um traço (“[O João]₁ eu vi ele₁ ontem”). Tais distinções já foram revisadas por Callou *et al.* (2002) como indistinguíveis sintaticamente, sobretudo no caso português. Dizem os autores que só é possível distinguir essas possibilidades através de marcas prosódicas. Seja como for, o que nos interessa aqui é o fenômeno genérico de topicalização, principalmente de sintagmas nominais, à periferia esquerda da sentença.

Baseado em Passos e Fonseca (2011), Tarallo e Kato *et al.* (1991) e Orsini (2013), estabelecemos uma breve classificação das CT observadas em PB:

1. Topicalização do OD: “A bola ele chutou”.

⁸⁶ Em muitos casos este sintagma nominal (SN), antes de ser movido, era um sintagma preposicional (SP). Ao ser alçado à primeira posição, justamente dita de tópico, essa preposição é deixada de lado, muito possivelmente por sua irrelevância na nova construção de tópico.

2. **Deslocamento do clítico à esquerda:** “*Esse jornal eu não o faria*” (Araújo, 2006, p. 115).
3. **Tópico Sujeito:** “*A Sarinha tá nascendo dente*” (Pontes, 1987, p. 11).
4. **Tópico pendente:** “*De novela só sei uma coisa: que cada trama é mais boba que outra*” (Pontes, 1987, p.12)
5. **Tópico pendente com retomada:** “[...] existe uma relação sintática entre o tópico e o comentário, pois o tópico é retomado por um elemento dentro da oração, que pode ser um pronome forte ou clítico, uma expressão genérica, **uma categoria vazia**, um pronome demonstrativo, um numeral, dentre outros” (Passos e Fonseca, 2011, p. 12 – **negrito meu**): “*As crianças, os pais precisam ser mais firmes com elas*” (Passos e Fonseca, 2011, p. 12)
6. **Tópico cópia:** “*Vôlei, Márcio não gosta de jogar vôlei*” (Passos e Fonseca, 2011, p. 12).
7. **Duplo Sujeito:** “*Os homens, eles acham que podem tudo*” (Passos e Fonseca, 2011, p. 12)
8. **Topicalização Selvagem:** “*Outro tipo de vestido, Suzana não gosta*” (Passos e Fonseca, 2011, p.12).
9. **Locativo:** “*Daquela casa, o muro é alto*” (Passos e Fonseca, 2011, p. 13).

Em outros termos, o que notamos em todos casos de CT é que o constituinte que se presta ao papel de tópico está relacionado de algum modo, semântica, discursiva ou pragmaticamente, com o sentido global da sentença analisada, também chamada de sentença-comentário, por fazer alusão ao comentário do SN tópico, presente em qualquer um dos exemplos acima. Pode parecer deveras ampla tal classificação, mas nosso interesse aqui é, primeiramente, a descrição do fenômeno de um modo amplo, de forma que possamos “medir” tal fenômeno no modelo de complexidade aqui apresentado.

No entanto, com o que até agora temos, parece não ser possível explicar as CT sem o auxílio de um viés discursivo. Seja como for, Pontes (1987) aponta para uma espécie de *continuum* entre sintaxe e discurso. Justamente é aí que talvez resida a chave para explicar as CT em PB. Mesmo em modelos como os da Cartografia Sintática (Negrão 2017) já se prevê um nível discursivo para as construções de tópico, i.e.: TopP. Tal nível fica acima do nível CP, cuja tarefa é reconhecer camadas discursivas nas sentenças.

Como mencionamos, nunca é demais salientar que as CT têm um aparato prosódico muito peculiar (O Pedro / ele não vem mais aqui), que pouco ou nada se parece com sentenças organizadas em construções do tipo sujeito-predicado (CSP). A questão é obviamente muita instigante, mas ultrapassa os limites do nosso trabalho. Por essa razão, as questões prosódicas sobre as CT, embora importantes, serão aqui pouco consideradas.

Exposta a classificação das CT em PB, passemos ao exercício de analisá-las segundo uma perspectiva de complexidade linguística. Antes disso, no entanto, uma brevíssima recapitulação da teoria que usaremos, a saber Dahl (2004).

4.3 Dahl (2004) aplicado às CT

Já vimos em capítulos anteriores como se dá o pensamento de Dahl (2004) acerca do que seja complexidade linguística. Neste capítulo, nos valeremos de alguns preceitos do mesmo Dahl (2004) para investigarmos se as construções de tópico em PB são mais ou menos complexas do que as construções do tipo sujeito-predicado. A princípio, a intuição indica que as CT são mais complexas do que as CSP, no entanto, veremos à luz dos preceitos aqui expostos se tal conclusão pode ser sustentada. Nesta seção testaremos as CT sob a ótica de um modelo que chamamos aqui de *informacional*, pois seu autor, Östen Dahl, em sua obra *The Growth and Maintenance of Language Complexity* de 2004, se refere à língua como um sistema informacional, e a partir das teorias da informação ele constrói seu mecanismo teórico-metodológico, do qual nos valeremos aqui, em partes, para apontar algumas propriedades das CT.

Dahl (2004) enxerga a língua, principalmente, como meio de transmitir informações. Nesse sentido, sua visão é de que a língua se presta principalmente à comunicação, tais como os preceitos de vertentes funcionalistas de estudos linguísticos. Desse modo, Dahl (2004) acredita que a cada informação realizada, há um “custo” de processamento verbal dessa informação. Tal noção também deverá ficar mais clara ao longo dessas nossas considerações acerca das ideias deste linguista que em muito contribui para os estudos da complexidade nas línguas naturais. Vejamos a seguir como se dão as noções de Dahl (2004) e como elas nos serão úteis aqui.

Primeiramente, talvez possamos associar as estruturas de tópico ao privilégio da informação-alvo. Ou seja, num primeiro olhar, parece que a CT se dá, justamente, para privilegiar a informação alvo. Parece que a estrutura de tópico se dá, justamente, para privilegiar a informação que Ilari (1992) chamou de “dada”, em oposição à informação “nova” dentro da dinâmica informacional-funcional que estamos trabalhando aqui. Assim como em Dahl (2004), ao admitirmos o que foi dito por Culicover (2013), a estrutura de tópico, ao privilegiar a informação-alvo (um SN geralmente), traz à língua maior custo de processamento sintático. Veremos como isso se dá e se aplica às CT, para tal, carece uma digressão acerca das noções postas em Dahl (2004), a começar pelas noções de *recursos* e *regulamentos* (*resources* e *regulations*).

Segundo Dahl (2004), a língua é uma entidade cujas realizações podem ser flexibilizadas em escolhas do falante juntamente com as regras impostas por essa língua. Ou seja, à língua ele atribui, inicialmente, dois valores indissociáveis: i) os *recursos* (*resources*), que são as maneiras pelas quais os falantes podem “operar” a língua em seus mais diversos modos de organização, seja sintática, estilística etc. Assim, o falante é mais ou menos livre para executar uma sentença, como as abaixo:

(10) Ele ama a Maria

(11) A Maria é amada por ele

, sem maiores prejuízos no que concerne à informação transmitida.

Por sua vez, ii) os *regulamentos* (*regulations*) são todas as regras ou leis impostas pela língua em sua gramática, e assim o falante pode escolher entre (8) e (9), mas jamais (10):

(12) Ele ama Maria

(13) A Maria ele ama

(14) *A ele ama Maria

, pois isso violaria uma regra sintática e, portanto, comprometeria a integridade da informação a ser transmitida.

Se pensarmos em termos de CT, podemos notar que o falante pode escolher realizar a CT ou não. Isso porque, conforme vimos na primeira seção deste capítulo, segundo a tipologia de Li & Thompson (1976), o PB pode ser caracterizado como uma língua de construção de sujeito-predicado e de tópico, concomitantemente, visto que ambas as construções são lícitas na língua e geram sentenças gramaticais. Nesse sentido, pode-se dizer que realizações de CT em PB são *regulamentos* que podem ou não serem usadas pelos falantes. Ou seja, o falante pode tanto se utilizar de (11) quanto pode optar por proferir (12):

(15) João ama Maria

(16) O João ele ama a Maria⁸⁷.

De todo modo, mesmo com a relativa liberdade trazida pelos *regulamentos*, a língua precisa preservar o trânsito da informação entre falante e receptor. Tal trânsito, como já dissemos, tem um “custo”. Dahl (2004) entende que “custo” é a energia demandada para se executar determinada tarefa (Dahl, 2004, p. 8), diferentemente de complexidade, que engloba, além de custo, dificuldade e outras medidas, tais como largura de banda (no caso das teorias da informação), cumprimento da descrição sintática ou morfológica, nível de ruído etc. Assim, uma sentença como o popular trava-língua “o rato roeu a roupa do rei de Roma” tem um custo lexical e fonológico relativamente maiores do que “Espero que ela venha”. No entanto, esta última sentença – composta de duas orações – parece mais complexa com relação ao trava-língua, dada natureza de sua descrição sintática. Podemos também relativizar o que se entende

⁸⁷ Nesta sentença, temos que considerar a prosódia para que a CT de fato se realize, caso contrário, poderia soar agramatical. Assim, em termos prosódicos, lê-se: “O João / ele ama a Maria” com uma pequena pausa entre a CT “O João” e a sentença comentário “ele ama a Maria” indicada pela barra inclinada.

por “custo”, mas o “custo” a que nos referiremos aqui é o que Dahl (2004) admite ser aquela demanda de energia dispensada para uma dada tarefa verbal. Assim, para simplificarmos, poderíamos dizer que uma sentença como “O João ele ama a Maria” teria mais custo de processamento do que “João ama Maria”. Admitimos aqui que “custo de processamento” é aquele tipo de energia demandada pelo cérebro para executar determinada ação verbal. Seja como for, “O João ele ama a Maria” teria mais custo de processamento do que “João ama Maria” não apenas porque contém mais palavras, mas porque sua descrição sintática é maior. Veremos isso mais adiante, a questão que fica aqui, no entanto, é de saber por que os falantes optam por sentenças com maior custo?

Custo pode ser traduzido tanto pelo esforço em pronunciar sentenças, quanto pelo número de fonemas dispensados para transmitir tal informação. Tal custo também pode ser medido pelo processamento despendido pelo falante, digamos, em seus processos mentais que são “traduzidos” pela sintaxe. Ocorre que muitas vezes a transmissão não é passada de forma íntegra, sendo desviada por ruídos, que podem ser desde vacilações do falante, até outros distúrbios sonoros externos. Internamente, os ruídos resultam em sentenças não tão bem formadas ao receptor. Então, para que haja bom trânsito de informação, a língua dispõe, muitas vezes, de redundâncias que, ainda que despendam maior custo ao falante, garantem a integridade da informação transmitida (ver capítulo 3).

É interessante notar que Dahl (2004) julga importante ao bom trânsito da informação nas línguas tanto os ruídos, quanto, mais ainda, as redundâncias (ornamentos, discutidos no capítulo 3), que são justamente mecanismos cuja função é evitar que os ruídos deturpem a mensagem. Como já citado, falávamos que a marcação de plural redundante nos SNs do português podem ser um exemplo dessa noção que Dahl (2004) tanto estima, a de ornamento, que as línguas só adquirem com o tempo, com sua maturação ao longo de sua história. Ora, do mesmo modo, podemos dizer que são também redundantes as informações contidas nas CT. Como vimos na segunda seção deste capítulo, a informação contida em uma CT é geralmente uma informação dada, a que falante e ouvinte compartilham em comum. Ela geralmente é repetida como destaque, pois se encontra na primeira posição da construção. Logo, podemos dizer que o PB ganha, sobretudo no último século (Araújo 2006) este ornamento, graças a erosão do sistema de clíticos pronominais (Pontes 1987, Tarallo e Kato *et al* 1991). Essa noção está em consonância ao que diz Deutscher (2014) ao afirmar que as mesmas forças de destruição da língua são as que também constroem outras e novas estruturas. Seja como for, ao desaparecerem os clíticos (possivelmente por razões fonéticas), a informação que eles transmitiam não pode se perder. Por esse motivo é que, possivelmente, a informação reforçadora, tais como pronomes cópia, das CT eclode.

Às CT propriamente, há em Dahl (2004) uma noção que talvez seja mais adequada. A noção de “mecanismos de aumento de redundância” (*redundancy-increasing mechanisms*). Além das

redundâncias próprias do sistema, a língua se utilizaria ainda de outras estratégias para garantir e até mesmo evidenciar certas informações que no discurso merecem mais importância. Dito de outra forma, para que não haja vacilos ou ruídos na transmissão de uma informação, as redundâncias garantiriam a integridade dessa informação, diminuindo assim o que Dahl (2004) chama de “redução de incerteza”. Ele diz que tal noção ocorre porque a língua é uma entidade ordenada, e que justamente tende a tal redução. Lembrando, como vimos no capítulo 3, que “ordem” é uma noção que depende bastante do ponto de vista. Pode ser ordenada a fila de carros num congestionamento, visto de um avião, mas não será ordenada, certamente, para o motorista de um carro que está no meio do congestionamento. Seja como for, a língua se vale da redução de incertezas para a garantia da informação, e o faz com mecanismos de redundância. Assim “os livros velhos” dão, em princípio, mais garantia de que a informação sobre a pluralidade desse SN chegue ao ouvinte do receptor do que “os livro velho”. Afora discussões do que está acontecendo nos plurais em PB, podemos confirmar que redundância é um tipo de ornamento cujo papel é auxiliar no bom trânsito da informação nas línguas naturais, apesar de as línguas poderem “funcionar” sem eles (cf. McWhorter 2001).

Mais ou menos da mesma forma ocorre com a comunicação verbal: um remetente pode manipular o nível de redundância, gastando mais ou menos energia ou tempo na mensagem que procura transmitir. Podemos assim concluir que a informação redundante contida numa CT é uma espécie de “proeminência” da informação dada, que é rememorada pelo falante ao ouvinte, situando-o no assunto-alvo.

Dahl afirma que a “redundância” – e seu exemplo são as marcações redundantes que acontecem em certos plurais do Inglês, ou seja, em português os temos também, até de maneira mais incisiva, conforma comentamos anteriormente – é até mesmo uma condição necessária para garantir a integridade da informação transmitida entre falantes. A redundância pode ser utilizada não apenas na transmissão da mensagem, mas em vários outros sistemas linguísticos. Quando um transmissor fala a seu receptor, leva-se em conta outros fatores além da mensagem, como os pressupostos do receptor, por exemplo: novamente, o que Ilari (1992) chama de “dado” ou informação dada.

Assim, adotamos o postulado de que as redundâncias, ornamentos próprios da língua obtidas pela sua maturação, são importantes ferramentas para o bom trânsito das informações contidas nessa língua. No caso das CT, ao admitirmos que são redundâncias, elas são, muitas vezes, necessárias para a organização da língua.

Outro aspecto linguístico interessante que Dahl (2004) usa para medir complexidade é o *tamanho da descrição*. Ora, já que para Dahl (2004) a língua é vista como informação, a CL é uma espécie de “medida de quantidade de informação necessária para reconstruir ou descrever tal língua”⁸⁸ (DAHL 2004, p. 2). Ou seja, assim como Culicover (2013), Dahl (2004) acredita que o que pode medir a complexidade de uma língua é, de alguma forma, o tamanho de sua descrição sintática. Então, o tamanho da descrição também produz uma informação que, por si só, também pode ser medida. Como Culicover

⁸⁸ No original: “a measure of the amount of information needed to describe or reconstruct it” (Dahl 2004, p. 2)

(2013), Dahl (2004) acredita que quanto maior a descrição de um fenômeno linguístico, mais complexo ele se torna. Ao tratar do assunto, ele primeiramente fala de complexidade no âmbito semântico, mas lembrando que a complexidade de um conteúdo deve ser diretamente relacionada ao comprimento de sua definição (DAHL 2004, p. 16), logo, da mesma forma, um significado complexo seria aquele que corresponde a um número maior de traços semânticos. O mesmo poderíamos atribuir à sintaxe, ou seja, o tamanho da descrição sintática de um enunciado deve servir para medir sua complexidade. Sintaticamente falando, nem sempre o tamanho da sentença e mesmo a quantidade suposta de informação contida nela garante que ela seja mais complexa sintaticamente do que sentenças menores ou com menos informação. Assim, ao representarmos “João ama Maria” e sua descrição,

[S [SV [SN V SN]]]

podemos perceber que sua versão como CT, “O João ele ama a Maria”:

[S [Top SN₁ [SV [SN₁ V SN₂]]]]

possui uma descrição maior. Segundo, então os parâmetros de Dahl (2004), a sentença com CT, tendo a descrição maior, é mais complexa. Obviamente, podemos falar aqui apenas em termos sintáticos.

Dentre as redundâncias tratadas por Dahl, gostaríamos de destacar uma noção que embora ele pouco mencione, será crucial para nossa análise. Trata-se da noção de “verbosidade”:

Outro desvio relativamente “benigno” da concatenação irrestrita é encontrado [...] onde a expressão se desvia da justaposição simples pela adição de algum elemento fixo. Por exemplo, em português, palavras quantificadoras e substantivos de massa podem ser unidos por justaposição simples, como em “muita neve” ou “pouca neve”. Em francês, por outro lado, uma palavra como *beaucoup* ‘muito’ não pode simplesmente ser justaposta ao substantivo *neige* ‘neve’; em vez disso, deve-se inserir uma preposição entre: *beaucoup de neige* ‘muita neve’. O desvio é benigno no sentido de que a saída ainda pode ser descrita em termos de uma operação de concatenação, incluindo os dois elementos de entrada e a preposição. No entanto, do ponto de vista do português, o elemento *de* parece redundante. Podemos dizer que as construções quantificadoras francesas são caracterizadas pela verbosidade, definida como tendo um peso fonético maior (contendo mais material) do que seria minimamente necessário (portanto, um caso especial de redundância)⁸⁹ (DAHL 2004, pp.: 52-53)

⁸⁹ Another relatively “benign” deviation from unrestricted concatenation is found [...] where the expression deviates from simple juxtaposition by the addition of some fixed element. For instance, in English, quantifier words and mass nouns can be joined by simple juxtaposition, as in much snow or little snow. In French, on the other hand, a word like *beaucoup*

Podemos definir verbosidade, então, como expressões linguísticas que se utilizam de mais elementos do que o necessário para sua realização informacional. Assim, podemos também nos utilizar da noção de verbosidade em CT, como em “O João ele ama a Maria”, ao compararmos com a CSP “João ama Maria”. Diríamos, portanto, que há mais verbosidade em CT do que em CSP, precisamente porque na CT temos, a mais, “O”, “ele” (que aliás é redundante) e “a”. No caso do português brasileiro, é importante notarmos que a presença dos artigos definidos é, simplificando as coisas, uma questão dialetal, mas a presença do pronome ‘ele’ é, de fato, uma diferença estrutural.

Argumentamos aqui, portanto, que há redundância nas CT, pois, como vimos, a informação exposta no SN que serve de tópico geralmente é dada, que é reforçada, na maioria dos casos, pela sua repetição, seja ela marcada por um pronome ou por uma categoria vazia, portanto uma informação redundante⁹⁰. Ora, se um dos princípios, para Dahl (2004) de complexificação das línguas é a redundância – a exemplo da marcação redundante de plurais numa cadeia de SN típica das línguas indo-europeias – e se parece que as CT configuram redundância, logo, temos um caso de complexificação do português brasileiro. Como vimos também que o ganho de redundância faz parte da maturação da língua – ou seja, em condições naturais, tais fenômenos acontecem nas línguas – o que vem ocorrendo em PB é aumento de casos de CT, logo podemos perceber que a língua, de alguma forma, vem se complexificando, apenas, obviamente, conforme constatamos aqui, no quesito ordem. Por outro lado, se as CT são redundâncias da língua, e que redundância está ligada a garantia de informatividade, logo são nesses termos que as CT serão aqui tratadas para fins de testagem. É já conhecida a relação entre a erosão do sistema pronominal clítico do PB e a eclosão de CT. Ora, novamente recorrendo à Dahl (2004), como já vimos, poderíamos dizer que a erosão dos pronomes pode ter dificultado a transmissão de informação (ou, melhor dizendo, a taxa de redundância/ruído), o que foi garantido por outra estratégia linguística, a realização de CT. Dessa forma, para maior segurança de que a informação seja efetivamente transmitida, como uma consequência da erosão do sistema pronominal e da informação semântica que dele provinha, surgiu uma nova estrutura linguística cuja função é a viabilidade dessa informação. Dito de outra forma, os pronomes clíticos, ou sua ausência, não seriam mais suficientes para garantia da informação transmitida pelo conteúdo das sentenças, por

‘much’ cannot simply be juxtaposed with the noun neige ‘snow’; rather, you must insert a preposition *de* in between: *beaucoup de neige* ‘much snow’. The deviation is benign in the sense that the output can still be described in terms of a concatenation operation including the two input elements and the preposition. From the point of view of English, however, the element *de* appears redundant. We may say that French quantifier constructions are characterized by verbosity, defined as having a larger phonetic weight (containing more material) than would be minimally necessary (thus a special case of redundancy) (Dahl 2004, pp. 52-53). Adapteí o exemplo mencionado para o português, que também funcionou perfeitamente.

⁹⁰ Leite *et al.* (2002) e Tarallo, Kato *et al.* (1991) diriam que há pronomes elípticos mesmo em casos como “Essa casa não bate sol”. Assim, teríamos: “Essa casa [ela] não bate sol” e “O carro [ele] quebrou o vidro”. Caso contrário, não seriam casos de construção de tópico, mas apenas inversão de objeto.

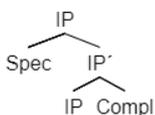
isso eclodem no sistema linguístico do PB, as CT. Concluímos, portanto, que a construção de tópico em PB é um *regulamento* escolhido pelos falantes entre duas *regras* impostas pela língua: construções de tópico versus construções do tipo sujeito-predicado.

Dada nossa breve apresentação, nos utilizaremos, então, de dois conceitos aqui apresentados. O conceito de **verbosidade** (ou seja, expressões linguísticas que se utilizam de mais elementos do que o necessário para sua realização informacional) e o conceito de **tamanho da descrição** (ou seja, conforme o tamanho da descrição de dado segmento linguístico, mais complexo ele é) para procurarmos medir um pouco mais objetivamente a complexidade das CT. Embora o resultado já possa ser vislumbrado, aplicaremos a algumas sentenças do *corpus* os testes dos dois conceitos mencionados por Dahl (2004) para identificarmos a maior complexidade das CT sobre as CSP. Procederemos aos testes com as sentenças do *corpus* sempre comparando intuitivamente as CT às CSP para fins demonstrativos, sem nos atermos a eventuais correspondências e suas respectivas discussões, feitas de maneira pontual na primeira seção deste capítulo.

Conforme veremos nos testes, a CT onera o processamento da língua, tornando, assim, a língua mais complexa no quesito ordem, em comparação a uma sentença equivalente em CSP. Aliás, para ilustrar bem essa ideia de equivalência, no exercício abaixo, para cada sentença com CT, acrescentaremos a ela uma sentença semanticamente equivalente em CSP. Em alguns casos o leitor poderá perceber que houve deslocamento de um sintagma à primeira posição da sentença, isso por conta do vestígio que é deixado na posição que o sintagma saiu, geralmente em forma de pronome ou de outro elemento de retomada. No entanto, em outras sentenças isso não fica tão evidente. Seja como for, todas as sentenças em CT se caracterizam por ter um nó acima de IP (*Inflectional Phrase* na terminologia da teoria *Government and Binding* (GB))⁹¹, ao contrário, conforme abaixo indicado em representações parentéticas, as sentenças em CSP não apresentam esse nó a mais. Procederemos, então, às sentenças que compõem nosso *corpus* acompanhada de análises de cada par de sentenças.

4.4 Exercício de interpretação conforme noções de Dahl (2004)

⁹¹ Segundo os parâmetros da GB, o nível IP (*Inflectional Phrase*), conforme o próprio nome já aponta, foi postulado para precisar o nível sintático onde acontecem as flexões dos verbos, acima de VP (*verbal phrase*). Embora não nos utilizemos inteiramente dela, sua estrutura padrão no modelo ([IP [Spec] [IP' [IP] [Compl]]]), em diagrama arbóreo é representado da seguinte forma:



Conforme dissemos acima, as sentenças com CT que mostramos aqui (sete no total) apresentam-se em pares semanticamente equivalentes com CSP, conforme se vê logo abaixo. Tomamos essas sete sentenças por conta da natureza do teste que aqui estamos realizando, pois elas podem ser colocadas de modo mais claro com relação a seus cumprimentos descritivos. As sete sentenças foram selecionadas, pois dão uma amostra razoável – veremos nas classificações – do que acontece em larga escala com as CT no PB, ou seja, são suficientes para uma boa amostragem do fenômeno atualmente em voga na língua. Em todo caso, os exemplos foram retirados dos textos que tomamos aqui como base deste capítulo, e conforme seus autores, extraídos de situações de falas naturais. As sentenças que analisaremos são as seguintes:

17a) A Maria eu vi ela ontem

17b) Eu via a Maria ontem.

18a) Meu cabelo dessa vez eu não gostei nem um pouco dele.

18b) Eu não gostei nem um pouco do meu cabelo.

19a) O Pedro ele não vem mais aqui.

19b) O Pedro não vem mais aqui.

20a) Verdura na parte da Argentina nós não vimos nada.

20b) Nós não vimos verdura na parte da Argentina.

21a) Aquela escola roubam muito ela.

21b) Roubam muito aquela escola.

22a) As cadeiras optativas, você precisa ter um bom conhecimento delas antes⁹².

22b) Você precisa ter um bom conhecimento das cadeiras optativas antes.

23a) O preconceito ele atinge a todos.

23b) O preconceito atinge a todos.

⁹² Como já dissemos, todos os exemplos do nosso *corpus* foram retirados das leituras que tiramos como base neste capítulo. Este, no entanto, é o único exemplo que veio com vírgula. Dessa forma, decidimos ser fiéis ao original e deixá-lo com vírgula, ainda que o exemplo pareça uma exceção ao ser comparados com todos os outros que não

Apresentados os pares de sentenças, as analisaremos a seguir, acrescentadas de suas divisões em blocos parentéticos que representam de forma ilustrativa como estão dispostos seus principais nós sintagmáticos.

Das sentenças que compõem este *corpus*, é importante que se mencione suas fontes. Dos exemplos (13) até o (16) as sentenças foram retiradas dos capítulos contidos na referência das “Gramáticas do Português Falado”, pontualmente do capítulo “Topicalização de Deslocamento à Esquerda: Sintaxe e Prosódia” do volume III (Callou *et al* 2002); e do capítulo “Tópicos e Adjuntos” do volume IV (Leite *et al* 2002). Já os exemplos de (17) até o (19), foram retirados de Eunice Pontes (1987). Ainda, é bom lembrar que os exemplos de Pontes (1987) foram também retirados de coletas informais da autora. Os exemplos das “Gramáticas do Português Falado” foram retirados do banco de dados no NURC, coletados de conversas informais ainda dos anos 1980. Os pares com CSP foram criados para este exercício.

Começemos com o exemplo em (13):

17a') [TopP A Maria] [IP eu vi ela ontem]

17b') [IP Eu via a Maria ontem]

Já neste primeiro par podemos observar que a sentença ordenada com CT tem um nó a mais, acima do nó IP. A literatura costuma observar que a sentença de sentido completo “eu vi ela ontem” é um comentário sobre o tópico “A Maria” (Tarallo, Kato *et al.* 1991, Galves 1998). Ora, se para Dahl (2004) é possível medir complexidade conforme o tamanho da descrição, podemos afirmar que aqui há aumento de complexidade, obviamente apenas no que tange à ordem sintática dos elementos da sentença. Também podemos considerar o movimento de um constituinte de dentro para fora do IP, conforme mostra nossa representação parentética. Ora, se considerarmos co-referentes o sintagma nominal, agora tópico, “A Maria”, e sua retomada no interior da sentença comentário com o pronome “ela”, fica ainda mais claro perceber o movimento do SN para fora do IP, tornando a construção de sujeito-predicado para uma construção de tópico-comentário, sendo o comentário toda a estrutura SVO “eu vi ela ontem” sobre o tópico “A Maria”. Assim, conforme nossa classificação acima, podemos dizer que (13a') é um caso de “duplo sujeito”, isso porque, ao que parece, o TopP “A Maria” nasceu como sujeito da sentença comentário e se pôs acima do nó IP, um típico caso de topicalização em que um SN é alçado à primeira posição da sentença, agora de estrutura de tópico, não mais de sujeito-predicado.

Em (13a') pode-se constatar também que há verbosidade em “ela”, que retoma a informação contida na CT “A Maria” ao referenciá-la. Desse modo, podemos dizer que “ela” é uma informação

redundante com relação à informação geral transmitida pela sentença. Em todo caso, a descrição de (13a´) é maior, como podemos ver acima, do que a descrição de (13b´), logo, segundo os parâmetros de Dahl aqui utilizados, podemos afirmar que a sentença (13a´) é mais complexa do que a sentença (13b´).

Vejamos agora um outro exemplo:

18a´) [TopP Meu cabelo] [IP dessa vez eu não gostei nem um pouco dele]

18b´) [IP Eu não gostei nem um do meu cabelo]

Num par de sentenças com um pouco mais de estrutura, como em (14), vemos que o sintagma “Meu cabelo” é o representante de uma CT, tendo como sentença-comentário o IP “dessa vez eu não gostei nem um pouco dele”. Aqui podemos classificar a sentença (14a´) como “topicalização selvagem”, pois é aquela que o sintagma preposicional é movido para a posição de tópico, isto é, acima de IP, e em que a preposição passa a não nuclear mais o sintagma. Dito de outra forma, o objeto indireto (OI) é alçado à posição de tópico da sentença comentário, cuja estrutura continua SVO, senão vejamos: “Meu cabelo” é tópico da sentença comentário “dessa vez eu não gostei nem um pouco dele”, cuja estrutura está intacta. Novamente vemos que embora haja relação semântica e discursiva da CT “Meu cabelo” com sua sentença comentário, eles não mantêm relações sintáticas.

No que tange à redundância, o pronome “dele”, no fim da sentença-comentário, pode ser classificado como verbosidade na sentença, pois é flagrante que sem ele poderíamos, talvez com um pouco mais de dificuldade, interpretar a sentença da mesma forma (“Meu cabelo dessa vez eu não gostei nem um pouco”). Ou seja, o que não foi “gostado” foi o “meu cabelo”. O que podemos inferir do “dele” também é que o constituinte é um vestígio do SP movido para frente da sentença. Ou seja, ao ser movido o OI “do meu cabelo”, ficou em seu lugar o pronome cópia “dele”.

A sentença equivalente em (14b´) mostra que de uma forma mais simplificada, dada a representação de sua estrutura que utilizamos, pode-se transmitir a mesma informação que em (14a´). De todo modo, novamente podemos observar que em relação a (14b´), (14a´) apresenta uma estrutura a mais quando temos a CT na frente de sua sentença comentário.

Mais um exemplo:

19a´) [TopP O Pedro] [IP ele não vem mais aqui]

19b´) [IP O Pedro não vem mais aqui]

Novamente, como no caso de (13a), a sentença em (15a´) possui sua descrição sintática maior do que em (15b´). Mais uma vez vemos o nó adicional TopP (“O Pedro”) em (15a´), assim como vemos o mesmo sintagma (“O Pedro”) em (15b´) não mais como tópico, mas simplesmente como sujeito da sentença. Assim, vemos que diferentemente de (15b´) em (15a´) temos um nível sintático a mais por conta da CT. Esse sintagma (um SN) possivelmente tenha se movida da posição de sujeito para um lugar acima de IP, portanto a descrição de (15a´) é maior do que a descrição de (15b´). Na classificação que estabelecemos acima, ela seria classificada como “duplo sujeito”, e embora haja coincidência entre CT e sujeito da sentença comentário, a CT não mantém relação sintática com o verbo da oração principal, ao menos não há relação direta.

Em (15a´) podemos ver que há informação redundante, naquilo que chamamos de verbosidade a retomada da CT “ele”, que aparece, então, como sujeito da sentença comentário.

O exemplo (16) traz mais uma estrutura:

20a´) [TopP Verdura na parte da Argentina] [IP nós não vimos nada]

20b´) [IP Nós não vimos verdura na parte da Argentina].

Em (16a´) vemos que a construção que serve de tópico é um pouco mais robusta (i.e., tem mais estrutura) do que as outras, contendo além de um sintagma nominal (“Verdura”), um sintagma preposicional (“na parte da Argentina”), cujo conteúdo ainda suporta, dentro dele, ainda outro sintagma preposicional (“na Argentina”), o qual contém outro sintagma nominal (“a Argentina”). Novamente vemos que sua contraparte com CSP (16b´) possui descrição sintática mais simples, novamente porque possui um processo a menos envolvido, ou seja, só apresenta o nó IP, enquanto que a (16a´) tem a mais, justamente, uma CT, que representamos sintaticamente aqui como nó TopP. Por outro lado, o nome “nada” no final da sentença, de certo modo, está retomando toda informação contida na CT, o que denotaria que toda construção de tópico está ali para reforçar (redundância) uma informação que está em consonância com a sentença-comentário “nós não vimos nada”. Dada tal redundância, podemos, talvez, afirmar que o elemento de retomada “nada” seja a verbosidade da sentença (16a´).

Quanto a uma classificação da sentença (16a´), poderíamos arriscar como “tópico pendente”, talvez devamos levar em consideração “nada” como retomada, embora não muito evidente. “Nada”, como retomada, talvez também seja o vestígio de algo que foi movido daquela posição, justamente o CT, representada pelo constituinte “Verdura na parte da Argentina”.

Ao se observar a sentença (16b´), em CSP, podemos inferir que todo constituinte “Verdura na parte da Argentina” foi movido para a posição de tópico, tornando a CSP, uma CT, cujo comentário

é toda estrutura SVO (“nós não vimos”). Desse modo, podemos observar que houve o deslocamento, nesta operação do OD à posição de tópico. Por esse motivo, inferimos também que “nada” é o vestígio deixado por esse OD agora topicalizado.

No par em (17), verificamos em (17a´) que o constituinte “Aquela escola” faz o papel de CT para a sentença-comentário em IP “roubam muito ela”. Novamente, assim como em (14) há retomada da CT em “ela”. Assim, podemos dizer que, com relação à (17b´), a sentença (17a´) tem a mais a verbosidade do pronome “ela”, não necessariamente essencial para o entendimento global da sentença. Vale lembrar que em (17a´) a sentença poderia ser pronunciada sem o pronome de retomada “ela”, sem prejuízo de significado.

Ao detalharmos ainda mais o par de sentenças do nosso exemplo, em relação a (17b´), nota-se que houve o deslocamento do constituinte “aquela escola”, originalmente um OD numa estrutura CSP, que se posiciona agora acima do nó IP, para uma posição de tópico. Dado isso, podemos inferir que, segundo nossa classificação de CT acima, se trata de um caso de “tópico pendente com retomada”. Em todo caso, o que vemos aqui é o deslocamento do OD “aquela escola” em posição de tópico, portanto acima de IP. Esse constituinte deslocado deixa um vestígio, sua retomada “ela”, que se refere a “aquela escola” e no qual podemos classificar como “verbosidade” da sentença (17a´), conforme vemos abaixo:

21a´) [TopP Aquela escola] [IP roubam muito ela].

21b´) [IP Roubam muito aquela escola].

Nosso penúltimo exemplo é:

22a´) [TopP As cadeiras optativas], [IP você precisa ter um bom conhecimento [delas antes]];

22b´) [IP Você precisa ter um bom conhecimento das cadeiras optativas antes].

, observamos que em (18a´) a CT “As cadeiras optativas” é o tópico da sentença-comentário “você precisa ter um bom conhecimento delas antes”. Observando, ainda (18b´) em CSP, notamos que houve o deslocamento de “das cadeiras optativas”, um OI, portanto, à posição de tópico, o que a tornou uma CT, por isso vemos um nível a mais em (18a´), o TopP. Conforme se verifica, a preposição “de” do constituinte movido não mais faz parte da, agora, CT, o que denotaria, segundo nossa classificação acima, como “topicalização selvagem”. Por ter uma estrutura acima de IP, (18a´) apresenta estrutura maior do que (18b´), portanto, segundo os critérios que adotamos de Dahl (2004),

(18a´) é mais complexa que (18b´) no que tange a ordem sintática. Ainda, observamos que a retomada “delas”, que faz menção ao constituinte supostamente deslocado “as cadeiras optativas”, gera mais verbosidade à sentença (18a´) sobre seu par, (18b´).

Finalmente, abaixo, apresentamos a última sentença que analisaremos neste exercício:

23a´) [TopP O preconceito] [IP ele atinge a todos]

23b´) [IP O preconceito atinge a todos]

Temos aqui em (19) estrutura semelhante ao do nosso primeiro exemplo, a qual a primeira sentença (19a´) apresenta estrutura TopP a mais do que a (19b´). Em (19a´) a CT “O preconceito” deixa a sentença mais complexa do que seu par em (19b´), isso porque agora não se trata mais de um caso de CSO. O que pode ter havido em (19a´) em relação a (19b´) foi o deslocamento do outrora sujeito “o preconceito” para uma posição de tópico, acima do nó IP, tornando (19a´) uma CT. Notamos, então, que (19b´), uma sentença em CSP, com ordem regular SVO, numa sentença do tipo tópico-comentário (19a´), sendo o comentário, toda estrutura SVO “ele atinge a todos”. Ao ser movido para posição de tópico, o sujeito “o preconceito” deixou um vestígio, representado pelo pronome de retomada, “ele”, que pode também ser contado como verbosidade, pois em (19b´), em que a sentença está em CSP, não há retomada, isso porque, agora, o sintagma nominal “O preconceito” não é mais uma CT, e sim parte integrante da sentença, como vemos na representação parentética.

Quanto à sua classificação, podemos dizer que esta CT em (19a´) é do tipo “duplo sujeito”, pois há coincidência entre CT “o preconceito” e o sujeito da sentença-comentário “ele”, que justamente retoma o CT.

Como vimos em todos os exemplos de pares de sentença, já nas representações parentéticas, as sentenças que contêm CT possuem sempre uma descrição maior do que seus pares, postas em CSP. Maiores não somente no que tange à descrição sintática, mas no que tange à possibilidade de recursos da própria língua, isso porque não temos mais apenas a ordem SVO, mas também um outro tipo de estrutura sintática. Lembrando Pontes (1987), que baseada na classificação de Li & Thompson (1976) afirma que o português brasileiro estaria passando de uma língua ordenada em sujeito e predicado para uma língua cuja organização tanto pode ser de sujeito-predicado, quanto de tópico-comentário, como acabamos de atestar aqui, diante dos pares semanticamente equivalentes de nosso pequeno *corpus*. Nesse sentido, segundo Dahl (2004), sentenças com CT parecem ser mais complexas do que as sentenças com CSP. Como acredita Tarallo, Kato *et al.* (1991), e como evidencia a prosódia, as CT são itens à parte da sentença, ou seja, se encontram em primeira posição na sentença, têm relação semântica com as sentenças-comentário, mas muitas vezes não mantêm nenhuma relação sintática

com essas. Ora, já que a estrutura básica de uma língua que se utiliza das CT (Li & Thompson 1976) prevê uma construção acima do nó IP, podemos ver, nos exemplos dados, que todas as sentenças escolhidas apresentam tal características. Os testes deste exercício mostram que as sentenças que contêm CT, apresentam estruturas maiores e mais complexas. Além disso, quase todas apresentam aquilo que Dahl (2004) chamou de verbosidade.

Para ilustrar melhor nossos resultados, apresentamos um quadro que resume os dois tipos de construções aqui envolvidos, bem como suas representações, tanto parentética, quanto uma representação mais tradicional:

	CSP	CT
Representação Parentética	IP [SN...]	TopP [IP [SN...]]
Representação Normativa	S.V.O.	Top-S.V.O.

Através do quadro ilustrativo percebemos novamente que as construções de tópico parecem ser mais complexas, pois envolvem um processo a mais. Percebemos isso em ambas as representações.

Ainda, é interessante se notar que mesmo que identifiquemos, aos termos de Li & Thompson (1976), o PB como língua de proeminência de tópico e de sujeito-predicado, a construção de tópico necessariamente eclode da velha ordem SVO. Prova disso são os constantes pronomes lembretes que quase sempre são encontrados em tais construções, e mesmo quando ausentes, na maioria dos casos, assume-se ali uma categoria vazia co-referente da informação presente nas CT. A CT, além de apresentar, flagrantemente, um processo a mais na cadeia sintagmática de suas sentenças, parece apresentar também mais interação do que na CSP, ou seja, em quase todos os casos verificamos retomada, seja com o sujeito, seja com o objeto da sentença-comentário. É visível que tais processos tornam a língua sintaticamente mais complexa, mas talvez, do ponto de vista do falante, pareça mais simples, já que o processo é, de alguma forma, redundante. Isto é, ao mesmo tempo que aumenta o processamento da língua, também garante a maior informatividade, notadamente nos casos em que há abertamente realizado um pronome.

Assim, mostramos uma descrição do ponto de vista da complexidade, obedecendo parâmetros descritos por Dahl (2004) de uma forma de medir complexidade linguística, qual seja, medindo sua descrição. Ou seja, a partir de uma pequena descrição sintática, tomamos a noção de Dahl (2004) para procurar mostrar que na CT há um processo a mais do que a mesma sentença em sua versão CSP. Além disso, podemos ver também que há um processo maior de interação do ponto de vista informacional. A descrição de complexidade exigiu uma pequena descrição sintática para mostrarmos

de que forma as CT são mais complexas, do ponto de vista sintático e informacional, do que as CSP.

Como vimos anteriormente, para Dahl (2004), um dos traços que torna a língua mais complexa são suas redundâncias. Ora, vimos nesta exposição que as CT podem ser consideradas redundâncias da nossa língua que, numa dinâmica informacional, servem como lembretes ao interlocutor sobre o assunto da informação dada. Novamente, através da redundância gerada pelas construções de tópico, podemos dizer que este traço tem tornado o PB uma língua, ao menos no que tange à ordem sintática, mais complexa do que sua variedade europeia, por exemplo.

De todo modo, o que procuramos mostrar aqui, para além das representações sintáticas, é que as descrições e as possibilidades de recursos das sentenças ampliam o repertório em PB. Tanto o tamanho das descrições sintáticas, quando a possibilidade de mais recursos mostram que o PB, ao menos nesse quesito, vem se tornando uma língua mais complexa.

4.5 Conclusão

Vimos, no início deste capítulo, um dos principais modos possíveis de classificar as construções de tópico: primeiramente elas aparecem geralmente como informação dada, em oposição à informação nova, geralmente atribuída ao sujeito de uma oração. Tema-remã ou informação nova e informação dada, ou ainda, como preferem alguns linguistas, tópico-comentário são termos análogos que dizem respeito à informação transmitida por uma sentença. Também vimos que, segundo Li & Thompson (1976), as CT são definidoras de tipos linguísticos, tanto que Pontes (1987) e outros linguistas afirmam que o PB estaria transitando, em sua tipologia, de uma língua de sujeito-predicado para uma língua de proeminência tópico-comentário e sujeito-predicado. Nesse sentido, as CT não mais fariam parte da sintaxe canônica SVO, mas de algum modo, delas eclode. Não fazendo parte da ordem canônica SVO, as CT, então, não mantêm relações sintáticas evidentes com suas sentenças-comentário. Dito de outra forma, uma CT não está em relação de concordância como o verbo, da mesma forma que um sujeito está. Portanto, construções de tópico, ao eclodirem no PB, estariam mudando a tipologia desta língua. De todo modo, não fazemos aqui uma classificação exaustiva, isso porque a literatura a respeito do assunto é vasta e, muitas vezes, controversa. Ao contrário, o que fazemos aqui é uma definição operacional de nosso objeto de estudo, as construções de tópico que, posteriormente foram submetidas a um teste métrico de complexidade linguística.

Antes dos testes, em forma de exercícios, propriamente, procuramos mostrar como pensa Dahl (2004) em relação a uma possível medida de complexidade linguística. Vimos primeiramente que Dahl (2004) trata a língua como ferramenta de informação, por conta disso, sua visão do que sejam

as línguas naturais compactua com alguns dos preceitos funcionalistas. Primeiramente expomos duas importantes noções trazidas de Dahl (2004), a de “custo” – especialmente custo de informação transmitida – e a noção, que já havíamos debatido no capítulo 3, de redundâncias. A noção de “custo” nos foi útil quando Dahl (2004) explica que uma medida possível de complexidade se dá pelo comprimento das descrições de dado fenômeno linguístico: no nosso caso, sentenças. O comprimento descritivo da sintaxe de uma dada sentença também denota quanto custo foi demandada na execução desta sentença. Além disso, Dahl (2004) explica que o que também gera maior custo a uma sentença é a redundância, tão necessária à comunicação linguística. Apesar de insistirmos no exemplo da marcação redundante de plurais em PB, conseguimos mostrar de que modo as construções de tópico são também informação redundante no PB. Novamente, Dahl (2004) insiste que as redundâncias tornam as línguas mais complexas, e que elas só eclodem em línguas com mais maturidade, ou seja, não se apresentam em línguas novas, conforme também já discutimos no capítulo 3. Um pouco antes, ainda falávamos das noções de *recursos* e *regulamentos* que também nos ajudaram a explicar o fenômeno das CT em PB. Além do tamanho da descrição, expomos outra noção para que usemos na nossa métrica nas CT, a noção de verbosidade, que é tão somente uma outra forma de redundância, de ornamento que as línguas adquirem com o tempo.

Feitos os exercícios, tendo como base as noções acima citadas de Dahl (2004), pudemos constatar que a descrição das construções de tópico, comparadas a sentenças equivalentes, construídas com sujeito-predicado, tendem a ser mais complexas. Isso por conta não só do tamanho de sua descrição, como também pelo número de operações sintáticas a mais, conforme representações parentéticas. Além disso, pudemos também observar nos pares de sentenças, de que forma age a verbosidade nas sentenças com CT, outro traço que também torna o PB uma língua mais complexa do que, digamos, o PE (segundo essa métrica, e com relação a esse fenômeno em particular), que parece não apresentar, estatisticamente falando, o mesmo número de sentenças com CT como o faz o PB. Além disso, é bom lembrar que a prosódia das CT com relação às CSP parecem ser mais sofisticadas, ou mesmo, mais complexas. Enquanto na CSP temos uma prosódia mais “regular”, na CT, para ser compreendida como tal, é preciso que se tenha, necessariamente, duas forças tonais ali agindo: “O João / ele ama a Maria”.

Feitos nossos exercícios para procurar medir a diferença de complexidade linguística em duas variedades do português, encerraremos, no próximo capítulo, nossa tese com algumas considerações finais.

Capítulo 5 - Considerações Finais

Pudemos perceber, ao longo desta tese, a importância dos estudos sobre complexidade linguística. Tanto no passado, quando se associava complexidade a algum tipo de valoração da língua; quanto nos estudos atuais, que lidam com a questão de um ponto de vista completamente diferente daquele do passado, ou seja, elevam a complexidade, até mesmo, a um novo paradigma nas ciências, ou a tomam como noção indissociável de vários sistemas naturais, inclusive a linguagem humana. Nas últimas décadas tem havido um esforço conjunto de diversos estudiosos – filósofos e cientistas – para procurar determinar da melhor maneira possível o que seja a complexidade e, principalmente, objeto desta tese, o que seja complexidade linguística. Inicialmente as questões que nos propusemos aqui a responder são: i) o que é complexidade; ii) o que é complexidade linguística; iii) como medir complexidade linguística. As questões i) e ii) estão postas nos capítulos 1 e 2, à questão iii) procuramos mostrar algumas métricas atualmente em voga na literatura (capítulo 3), e depois procuramos demonstrar, através de um exercício que teve como tema um fenômeno do português brasileiro (capítulo 4), como uma dessas métricas pode ser aplicada numa proposta de medir complexidade linguística.

Logo na introdução, ao apresentarmos nosso objetivo nesta tese, procuramos mostrar de que maneira os estudos sobre as línguas viam a complexidade linguística, ou seja, sempre associando complexidade a critérios de valor. Também lá já deixamos claro que esta tese não pretendia polemizar as questões éticas, culturais ou ideológicas em torno da complexidade linguística. Já na mesma introdução procuramos também definir alguns parâmetros, filosóficos e linguísticos, do que seja complexidade. Por exemplo, já procurávamos, ainda que timidamente, mostrar que há diferentes tipos de complexidade linguística, e também já procurávamos mostrar que medir complexidade global é bastante difícil. Aliás, este tema vem sendo desenvolvido, aos poucos, ao longo de toda tese. Mesmo que saibamos que a língua é composta de diferentes níveis interligados, medir cada um deles é uma dificuldade a parte, não somente pela natureza própria de cada nível, mas pelas controversas de onde começam ou terminam cada um deles, bem como a natureza de sua interrelação.

No capítulo 1, com uma abordagem historiográfica, primeiramente procuramos estabelecer razões para que a complexidade linguística (CL) deixasse de ser objeto de estudos da linguística. Vimos que, pelo menos, desde o século XIX estudiosos associavam CL com algum critério qualitativo de seus falantes. Antes disso, já na antiguidade, acreditava-se que línguas de paradigmas flexionais ricos, como o latim e o grego, possuíam mais complexidade do que outras línguas. Na verdade, parece ser no século XIX que surgem as primeiras elucubrações sistemáticas sobre a complexidade das línguas. No entanto, dada a influência da filosofia romântica, a complexidade das línguas geralmente

era associada à capacidade mental de seus falantes. Ora, já que desde há muito enaltecia-se a riqueza morfológica das línguas europeias, estabeleceu-se, a partir de então, que línguas com maior riqueza morfológica são mais complexas, logo tornando outras línguas sem esse traço, inferiores, bem como seus falantes e suas sociedades.

Apesar disso, o século XIX também trouxe questões importantes acerca da CL. Duas importantes questões que vieram do século XIX e que até hoje nos desafia são: i) de onde vem a complexidade das línguas? e ii) por que as línguas estão deixando de ser complexas (ao menos do ponto de vista de sua morfologia)? Sobre a ii), Deutscher (2014) diz-nos que é outro truísmo da linguística, pois na mesma medida que as línguas ficam menos simples em dados níveis, ficam mais complexa em outros e isso sem necessariamente haver uma lei de equicomplexidade, a qual Gil (2009 *apud* Sampson *et al* 2009) tanto critica. É bom lembrar que no século XIX havia a associação entre complexidade e riqueza morfológica, ou seja, ao passo que o latim perdia desinências verbais, resultando nas línguas neolatinas, acreditava-se piamente numa simplificação linguística. Ainda assim, as línguas continuavam complexas: isso era o que mais intrigava os linguistas da época, principalmente August Schleicher (1821-1868). O mesmo ocorria com as chamadas irregularidades. Pensava-se que línguas ancestrais, como o hipotético protoindo-europeu, não apresentasse qualquer irregularidade, e que elas surgissem devido justamente a alguma deformação causada pela perda de complexidade. Hoje em dia, ao contrário, sabe-se que as irregularidades não somente são necessárias, como tornam a língua mais complexa (Culicover 2013, Dahl 2004). A ideia geral era de que o tempo “corroía” a complexidade das línguas. De todo modo, outra questão que intrigava os pesquisadores era o fato de que as sociedades estavam se tornando mais complexas, na mesma medida que as línguas iam se degradando (novamente, considerando quase que exclusivamente, como era praxe então, sua morfologia). Por conta disso, Schleicher (Deutscher 2014) postula uma teoria em que acreditava que a língua se torna mais simples à medida que o tempo passa e que as “nações” tornam-se mais complexas, num estranho equilíbrio entre manter a complexidade da língua ou investigar na complexidade da sociedade, mas nunca haver energia para ambas as atividades. Seja como for, as ideias concebidas nessa época abriram caminho para o chamado “relativismo linguístico” que, *grosso modo*, acreditava que fatores externos, tais como a cultura e organização social, influenciavam na forma da língua. Tal corrente, embora vista hoje de outra forma, também corroborou para leituras enviesadas sobre a CL. E foi talvez o mesmo “relativismo linguístico” que tenha gerado aversão aos linguistas das próximas gerações aos estudos sobre CL.

Novamente, a tradição historiográfica na linguística aponta o século XX como aquele que firmaria a linguística no campo genuíno das ciências. Diferentemente das gerações anteriores, e por conta da influência do positivismo, da antropologia e dos estudos sobre as línguas ameríndias, agora fala-se em CL com mais cautela. Na verdade, constatado o grau de complexidade nas línguas

ameríndias, bem como seus sofisticados costumes e cultura, inspirados sobretudo nos estudos de Franz Boas, a linguística, agora, trata a questão da CL de forma completamente diferente daquela outrora observada pelos estudiosos no século XIX. Mais precisamente, inspirada na antropologia, a linguística agora admite a possibilidade – que depois se tornaria uma certeza e um truísmo – de que todas as línguas são igualmente complexas. A ideia geral era de que as culturas americanas, chamadas primitivas, mantinham em sua cultura e em sua língua estruturas tão complexas quanto qualquer língua europeia, logo, não haveria razão para acreditar que as línguas desses ditos povos primitivos fosse menos complexa do que as línguas europeias. O mesmo diziam os antropólogos sobre seu aparato cultural. Na linguística propriamente, o mais destacado estudioso a versar sobre a equicomplexidade talvez tenha sido Charles Hockett, que afirmara categoricamente que as línguas não poderiam ter outro destino a não ser a equidade de sua complexidade. Hockett, tendo dito em 1958, porém de forma mais especulativa do que comprobatória de fato, cria um truísmo na linguística moderna, repetido exaustivamente até hoje inclusive em manuais e aulas introdutórias de linguística. Altamente justificável, a nosso ver, mas ainda assim um truísmo que cabe a nós ajudar a desfazer.

A partir da segunda metade do século XX, com o advento do chamado pós-estruturalismo, a situação da CL não mudaria muito. Há, em certa medida, predominância da escola gerativista que divide a língua em princípios e parâmetros, ou seja, aquilo que toda língua tem comum versus as particularidades de cada língua. Seguindo um modelo cartesiano, o gerativismo se importará apenas com aspectos que são comuns em todas as línguas. Ora, sendo a busca desse programa científico princípios comuns a todas as línguas, não faz sentido falar em diferenças, inclusive de complexidade entre as línguas naturais. Os chamados “Universais Linguísticos” também foram objeto de importante ensaio do linguista Joseph Greenberg (1963). Diferentemente do trabalho de Chomsky (1965), no entanto, a tipologia moderna encetada por Greenberg, hoje, em seu atual estado da arte, auxilia a novas descobertas sobre as diferentes complexidades das línguas. No entanto, foi no âmbito da sociolinguística e dos estudos sobre aquisição de segunda língua que começaram a ressurgir, através de sofisticadas ferramentas, a velha indagação de que as línguas poderiam diferir entre suas complexidades. Mas somente no início do século XXI (McWhorter 2001) começa-se a pensar em complexidade linguística, desvinculada da capacidade de seus falantes e com vistas ao estabelecimento de critérios objetivos para sua investigação.

Além de procurar erradicar o truísmo de que todas as línguas tenham a mesma complexidade, os estudos atuais sobre CL encaram outros desafios, tais como: quais métricas são mais seguras e eficientes para se precisar complexidade; relacionar complexidade linguística e expressividade, ou melhor dizendo, procurar mostrar que o menor grau de complexidade de uma língua não está diretamente associado ao seu poder expressivo; e, sobretudo, mostrar que complexidade linguística não tem relação com processamento mental ou capacidade de seus falantes, dito de outro modo: se

eventualmente uma dada língua A apresenta paradigma flexional menos complexo que dada língua B, isso não implica em dizer que a língua A, em sua globalidade, é menos complexa do que a língua B. Como acredita McWhorter (2001): parece que todas as línguas naturais têm um grau de complexidade elevado em todos os seus níveis, umas, no entanto, apresentam menor ou maior grau de complexidade em alguns níveis do que outros, sem haver, necessariamente, compensação entre esses níveis, e em nenhum caso com algum comprometimento do poder expressivo da língua.

No capítulo seguinte procuramos definir o que seja complexidade e, depois, o que seja complexidade linguística. Primeiramente procuramos definir complexidade ante várias áreas do conhecimento, mas, sob as lentes da filosofia da ciência, principalmente baseados em dois pensadores que se inclinaram a investigar o termo: Gell-Mann (1995, 1996) e Edmonds (1999). De modo geral, a complexidade tem se mostrado uma noção profícua nos últimos anos, a ponto de alguns estudiosos encararem complexidade, até mesmo, como um novo paradigma nas ciências. Há várias noções envolvidas que compõem o entendimento do que seja complexidade. Uma das mais importantes talvez seja a noção de *saber*. Saber, na verdade, é uma espécie de avaliação e métrica sobre algum objeto observável. O saber também pode determinar se algo é regrado ou completamente aleatório. Para Gell-Mann (1995), ambos os extremos caracterizam a total ausência de complexidade. Ou seja, se algo no mundo é perfeitamente regrado, não se pode dizer que ele seja complexo, o mesmo vale para algo completamente aleatório. Assim, a complexidade se definiria como algo nem regrado, nem aleatório, mas situado entre esses dois polos. Para Edmonds (1999), por exemplo, complexidade – que pode ser aplicável apenas a modelos e não diretamente a dados da natureza – é o oposto de linearidade e determinismo. Ou seja, o contrário do que apregoa o paradigma newtoniano na ciência.

Outro aspecto que eclode nas últimas décadas é a de “sistemas adaptativos complexos”, do qual, segundo cientistas como Gell-Mann (1995), a língua seria o melhor exemplo. Trata-se de sistemas biologicamente determinados, mas que também sofrem influências dos meios sociais a que estão atrelados. Ou seja, são sistemas naturais e inatos, mas que devem sofrer alguma influência externa para sua eclosão. Embora naturais e inatos, não surgem sem um *input* do meio social em que estão inseridos. Por esse motivo esses sistemas se adaptam e se modificam com maior velocidade do que outros sistemas naturais. Dada essa definição, percebemos também, e mais uma vez, que a complexidade é característica inerente e indissociável das línguas naturais.

Após essa exposição, finalmente passamos a definir o que seja complexidade linguística. Primeiramente, há já um quase consenso entre linguistas de diferentes orientações de que o que torna a língua mais complexa são características externas à língua (Culicover 2013). Além do auxílio vindo de noções trazidas de pensadores, tais como Gell-Mann (1995 e 1996) e Edmonds (1999), na segunda parte do capítulo 2 procuramos mostrar, principalmente através de Kusters (2003), Dahl (2004), Culicover (2013) e Newmeyer & Preston (2014), o que se entende por complexidade linguística.

Esses linguistas estão entre os principais pesquisadores que não apenas trabalham com definições de CL, mas também com seu emprego ao lidar e (tentar) solucionar questões linguísticas empíricas.

Para Kusters (2003), a complexidade linguística tem relação direta com o julgamento dos falantes. Sobretudo o falante estrangeiro, aprendiz de uma segunda língua (L2). Desse modo, mede-se complexidade linguística através da proficiência dos falantes estrangeiros de L2. Naquilo que o falante estrangeiro apresenta mais dificuldade de aprendizagem é onde estariam os itens mais complexos das línguas. O sistema de tons é um bom exemplo disso. Obviamente isso depende da língua materna do falante estrangeiro.

Para Dahl (2004) o conceito de CL está na tensão entre “custo” e “economia”. Quanto mais custo de processamento tem uma língua, mais complexa ela é. Ao contrário, ao se preservarem os custos, a economia garantiria simplicidade à língua, isso posto de modo muito simplificado, é importante dizer, pois Dahl (2004) está observando em separado os diversos níveis da língua. Ou seja, não há, para Dahl (2004), maneira de se medir complexidade global de uma língua, seja ela qual for.

Para linguistas de correntes formalistas – Culicover (2013) e Newmeyer & Preston (2014) são os maiores expoentes nesta tese – naturalmente a estrutura é o local onde habita a complexidade das línguas. Assim como Dahl (2004), a estrutura das línguas, ou mais propriamente sua descrição, é a chave para se entender onde reside a CL. Seja como for, nenhum estudo constata, em nenhuma língua, um grau zero de complexidade. É também ponto pacífico de que em nenhuma língua, e principalmente de que nenhum falante, tenha menos capacidade computacional do que outro. Seja como for, há já uma espécie de consenso de que a complexidade global não é viável, nem desejável, de ser medida.

Outra espécie de consenso que há entre linguistas que se dedicam a estudar a CL é a diferença entre complexidade absoluta e complexidade relativa. Embora os termos sejam usualmente comuns, suas noções nem sempre estão acordadas entre esses linguistas. De modo geral, a complexidade absoluta é impossível de ser medida por si. Assim, não é viável que olhemos para a inteireza da língua e, a partir daí, determinemos se ela é ou não mais ou menos complexa do que outra ou do que outros de seus estágios temporais. O mais viável, no entanto, seria medir complexidade relativa, ou seja, observar as línguas em seus vários níveis, ou em suas interfaces, para daí sim observar se a língua é mais ou menos complexa nesses níveis específicos.

Outra noção extremamente importante quando se pretende medir a CL, é a noção de “idade da língua”, e junta a ela, sua devida “maturação”. Ora, para linguistas como Dahl (2004), línguas mais antigas, e apenas elas, manteriam traços distintivos mais robustos que só seriam adquiridos pela “maturação” de dada língua. Um desses traços, aliás, é objeto de nosso exercício de análise no

capítulo 4, o qual seja, as construções de tópico no PB que, por uma série de razões discutidas naquele capítulo, eclodem de forma mais saliente no PB do que no português europeu.

Com o passar do tempo e com a maturidade, as línguas vão adquirindo “ornamentos”. Esses ornamentos são traços que ganham as línguas, com o tempo, na medida que vão envelhecendo. Usamos, ao longo da tese, para exemplificar essa importante noção, a marcação redundante de plural, tão disseminada no português normativo. Ao passo que a língua adquire ornamentos, ela onera o seu sistema, pois aumenta seu custo de processamento. Nesse sentido, então, Dahl (2004) diz que tais ornamentos, via de regra, não trazem vantagem computacional para língua, servem para, entre outras possibilidades, garantir a informatividade dos enunciados. McWhorter (2001 e 2005) propõe uma noção semelhante, a de “superespecificação”, que só estaria presente em línguas antigas. Nesse sentido, línguas jovens não apresentariam o traço mais superespecificado.

No capítulo 3, apresentamos alguns modelos para se medir complexidade linguística. Com a finalidade de simplificar nossa exposição, mostramos, então, que se pode medir CL, a princípio, sob três vieses diferentes: o viés da estrutura, o viés do uso e o viés do falante.

Sob o viés da estrutura, nomes importantes como Culicover (2013), Newmeyer & Preston (2014) e Nichols (2009) vêm se inclinados à questão. Para Culicover (2013), por exemplo, fatores sociais regulam complexidade ou simplicidade na língua, embora isso seja refletido em sua estrutura. Nichols (2009) se utiliza de um aparato bastante robusto que privilegia praticamente todos os níveis da estrutura da língua para tentar medir CL. Newmeyer & Preston (2014) apelam para os princípios (da noção gerativista de princípios e parâmetros) para procurar medir CL, já que os universais apontam para uma leitura de equidade, o que não interessa aos estudos sobre CL.

Além disso, no que tange ao viés da estrutura, na parte do capítulo mostramos as dificuldades de se medir, por exemplo, complexidade semântica. Tal feito depende, em dada medida, de dados fornecidos a computadores. Ora, operações de língua natural como ambiguidade e quantificações são noções extremamente difíceis para se processar computacionalmente, portanto difíceis de serem apreendidos para uma eventual métrica. Posição contrária da semântica é, por exemplo, a da morfologia, em que abundam exemplos que podem ser medidos.

Já pelo viés do uso, apresentamos especialmente as ideias de Dahl (2004) que associa língua à comunicação. Sua noção de “ornamentos” é muita cara para fornecer métricas de CL, pois ele mostra que medir CL tem íntima relação com a idade da língua. Ou seja, para adquirir ornamentos, a língua deve ter certa idade. O que mais ou menos diz McWhorter (2001 e 2005): línguas novas, emergentes, não possuem certos traços que línguas antigas possuem. E é justamente através desses traços que as línguas só adquirem com o tempo a chave para medir CL. Na esteira de Dahl (2004), o matemático Vulanovic (2007) também associa eficiência de uma gramática ao menor número de regras que ela apresenta em sua descrição. Assim, tanto para Dahl (2004), quanto para Vulanovic

(2007), uma gramática que apresenta maior descrição é, ao mesmo tempo, informacionalmente menos eficiente. Nesse sentido, para esses dois estudiosos, complexidade é visto como uma espécie de desvantagem do ponto de vista do processamento da língua.

Já pelo viés da aquisição ou aprendizado, há variados e diversos modelos, isso porque, como comentamos no capítulo 3, os estudos sobre aprendizado de segunda língua, há tempos, se utilizam de métodos para medir proficiência. Isso, ao nosso ver (Leal 2018), pode ser utilizado como uma métrica possível de CL. Um parâmetro importante que linguistas perceberam na aprendizagem de L2, relacionando-o à métrica de CL, é o fato de os falantes estrangeiros, muitas vezes, normatizarem a L2 tomando como base sua língua-mãe. Nesse sentido, pode-se perceber que estruturas que apresentam mais dificuldades de aprendizagem, podem ser estruturas mais complexas. Isso obviamente tem diversas variantes, como o eventual parentesco ou não entre L1 e L2. Aliás, dos três níveis de propostas de métricas de CL que vimos aqui, a mais abundante é a do viés da aquisição. Possivelmente porque, como já o dissemos acima, ferramentas eram já utilizadas há tempos para medir proficiência, e que agora também podem ser usadas para medir CL.

Posto de uma forma bastante genérica, embora toda contribuição apontada nesta tese acerca da CL, julgamos que três linguísticas merecem mais destaque graças a seus trabalhos originais e que contribuíram de maneira mais veemente para os estudos sobre CL: John McWhorter, Östen Dahl e Wouter Kusters. McWhorter (2001), principalmente, porque, ao investigar sincronicamente o fenômeno das línguas emergentes, especialmente as línguas crioulas, fez perceber através de ferramentas para se determinar o que são línguas emergentes, justamente, que há diferenças de complexidade entre as línguas, sobretudo que há diferenças de complexidade entre línguas mais novas e línguas já há mais tempo estabelecidas. Dahl (2004) apresenta uma noção bastante parecida, chamando de “ornamentos”, como vimos, os traços das línguas adquiridos somente com o tempo. Ora, ele percebe que são justamente esses ornamentos que tornam as línguas mais complexas, conforme procuramos mostrar no nosso capítulo 4. Kusters (2003) nos mostrou, através das desinências verbais de três diferentes variedades do quéchua, sendo uma delas língua franca, que pode haver diferenças de complexidade em uma mesma língua a depender de sua variedade. Com isso, também mostrou que complexidade é um fato que pode ser notado ou medido somente em sua contraparte social, embora reflita em sua estrutura.

Finalmente, no capítulo 4 fazemos uma demonstração, através de um exercício, de como complexidade linguística pode ser medida. Usamos duas noções de Dahl (2004), a de *cumprimento de descrição* e da *verbosidade* para aplicar num fenômeno do PB, as construções de tópico, e assim mostramos que no quesito ordem, o PB está se tornando mais complexo do que o português europeu. Importante que se frise novamente que apenas no quesito ordem, e apenas nesse fenômeno, o PB está ficando mais complexo do que o PE. Como já discutimos anteriormente, não é razoável medir

complexidade global de uma língua. Ao longo da tese mostramos por que razão deve-se medir complexidade apenas dos níveis das línguas e, quando muito, de suas interfaces.

Para trabalharmos num modelo de medida de CL com as CT, no entanto, antes precisamos postular alguns parâmetros operacionais do que sejam as construções de tópico no português brasileiro. Via de regra, definimos CT como um fenômeno relativamente novo em PB e que apresenta características peculiares justamente com relação à ordem dos elementos numa sentença. Mais propriamente, o que tentamos apontar é que a CT estaria deslocando o PB de sua ordem “natural” SVO. Ao invés da ordem SVO o que temos agora é um outro constituinte, o tópico, que aparece na periferia esquerda de uma sentença, mas cuja relação sintática com esta não é propriamente padrão. Dito de outro modo, o que caracteriza uma CT, geralmente, é um SN (na maioria das vezes) que foi movido à posição esquerda do enunciado, e que, embora muitas vezes se confunda com o sujeito da sentença, não se comporta dessa forma, pois geralmente não estabelece relação de concordância com o verbo da oração. Na verdade, linguistas como Tarallo, Kato *et al.* (1991) afirmam que a CT movida para posição de tópico estaria quase que completamente desconectada sintaticamente da sentença, e o que estaria em jogo seria outra forma de organização, agora tópico-comentário, sendo tópico a CT e comentário a sentença, cuja estrutura ainda se preserva em SVO. Propriamente, eles dizem que a CT não mantém relações temáticas com a estrutura SVO.

Para execução do nosso exercício de análise, retomamos alguns conceitos de Dahl (2004) que melhor se encaixam aos fenômenos que aqui escolhemos, os das CT. Primeiramente introduzimos os conceitos de *recursos* e *regulamentos*, em que mostramos que, dada a possibilidade de, em PB, se usar tanto construções de tópico, quanto construções do tipo sujeito-predicado, a escolhas dos falantes são caracterizadas como *recursos* disponíveis nessa língua. Logo depois, procuramos mostrar que as CT são uma espécie de ornamento adquirido pelo português nos últimos séculos (Araújo, 2006). Ora, se há um ornamento no PB, que só é adquirido com o tempo, é lícito que através dele procuremos observar um provável fenômeno de aumento de complexidade. Fizemos isso através do tamanho da descrição sintática das CT e observamos que, além das descrições de CT serem maiores do que as descrições do tipo construção sujeito-predicado (CSP), envolvem também mais operações. Assim, numa dada sentença que seja CSP verificamos apenas um nó IP em nosso exercício, enquanto em construções análogas a essas em CT, há uma operação acima de IP, o que a cartografia sintática (Negrão 2017) cunhou como nó TopP, acima de IP. Além disso, já é conhecida a relação particular que têm as CT com uma prosódia diferente das CSP. Constatamos que em Callou *et al.* (2002) a prosódia das CT se diferencia da prosódia de uma CSP por apresentar mais nuances. Dito de outra forma, para além dos apontamentos de descrição sintática, possivelmente uma eventual descrição prosódica também mostraria que a CT envolve mais operações do que a CSP, o que, nos parâmetros

de Dahl (2004) e de Culicover (2013), são indícios de que, ao menos no quesito ordem, o português brasileiro pode estar se tornando uma língua mais complexa.

Para o exercício nos utilizamos de sete sentenças com construções de tópicos e atribuímos a cada uma delas uma contraparte de construção sujeito-predicado. Feitas as devidas considerações, em todos os casos, através de representações sintáticas parentéticas, notamos que em CT as descrições foram sempre maiores. Tendo maior descrição, portanto, a sentença em CT apresenta maior grau de complexidade do que seu par CSP.

Após mais de meio século, a complexidade volta a ser objeto de estudo da linguística, isso graças a esforços de alguns estudiosos que não se intimidaram e enfrentaram o truísmo estabelecido, justificadamente, pelos estruturalistas de que todas as línguas deveriam ter o mesmo grau de complexidade. Embora vários linguistas, como vimos acima, estejam pesquisando acerca da CL, o tema, talvez pela amplitude, talvez pela discordância em traçar suas reais características, não se encontra ainda como conjunto de estudos bem estabelecidos no meio linguístico. Apesar dos esforços, dos inúmeros trabalhos e das diversas propostas para medir CL, não temos ainda um programa científico coeso para precisar noções mais bem estabelecidas. Ainda que posições filosóficas firmes estejam postas, a complexidade linguística carece de uma atenção especial: primeiramente porque a língua é um objeto, por si só, complexo de se definir; também porque a própria complexidade das línguas tem diversas possibilidades e finalidades. Vimos, por exemplo, no capítulo 4, que parte da complexidade da língua se dá para garantir a informatividade, mas certamente esta não é a única função da CL nas línguas naturais. Além disso, outra dificuldade é a de estabelecer parâmetros precisos para métricas de complexidade. Talvez ferramentas descritivas para avaliar e descrever sintaxe num viés de sua complexidade possa nos ajudar, mas por si só não são capazes de estabelecer padrões definitivos de métricas, pois as línguas naturais são compostas de vários níveis que se conectam de forma indissociável. Por esse motivo que as tipologias das línguas são ferramentas importantes para o estudo da CL.

Muitos dos linguistas citados nesta tese estão ainda ativamente trabalhando com CL. Várias obras, como aqui citamos, além de congressos, encontros ou simpósios foram já dedicados ao assunto, desde, pelo menos, 2015. Só para citar alguns, aconteceram dois congressos: o “CoEDL Seminar: Social Structure and Linguistic Complexity” da Australian National University em 2017 e outros dois estão marcados para acontecer ainda este ano de 2020: o “Acceptability & Complexity Evaluation Task for Italian” do instituto EVALITA, na Itália e o simpósio “Investigating Complexity in L2 Phraseology: methods and applications” em Groningen, na Holanda. Além disso, o professor Frederick Newmeyer fez uma apresentação sobre CL nas conferências virtuais da ABRALIN, no último dia 6 de julho. Nela, o professor Newmeyer fez um apanhado historiográfico da CL, depois apresentou suas principais bases teóricas e de que forma a CL é hoje em dia debatida e estudada.

Newmeyer (2020), depois de uma brevíssima apresentação historiográfica, apresentou três razões pelas quais a CL deixou de ser debatida: i) razão humanista, pois como debatemos no capítulo 1, isso poderia desabonar alguns povos de línguas que possuem traços mais simples e suas capacidades cognitivas; ii) pela chamada equicomplexidade, que indicaria a simplificação de um dado nível da língua, enquanto outro nível se tornaria mais complexo, conforme debatemos no capítulo 2; iii) uma razão teórica que acredita que, por ter universais, uma língua não pode apresentar complexidade maior ou menor que outra, conforme também debatemos no capítulo 2. Ao debater o que os linguistas falaram, até agora, a respeito desses três argumentos, Newmeyer lembra que Chomsky nunca se manifestou claramente sobre eventuais diferenças de complexidade entre as línguas. Nesse sentido, cita Andrea Moro para quem as línguas, por serem biologicamente determinadas, não precisam diferir em sua complexidade.

Posteriormente, Newmeyer passa a discutir sobre possíveis métricas de CL. A princípio ele apresenta três critérios possíveis: i) aquisição de primeira língua; ii) aprendizado de segunda língua; iii) uso da língua. Um pouco diferente do que debatemos no capítulo 3 desta tese, mas não totalmente. Assim como nós aqui o fizemos, Newmeyer destaca os trabalhos de John McWhorter. Para McWhorter, que observou línguas emergentes, três critérios conjugados podem determinar se uma língua é mais ou menos complexa do que outra: i) superespecificação; ii) tamanho de descrição; e iii) irregularidades. Debates esses aspectos ao longo do capítulo 3, apontando que Dahl (2004) e Culivover (2013) também compactuam com algumas das ideias de McWhorter.

Ao comentar sobre as diferentes maneiras de medir complexidade pela aquisição/aprendizagem, Newmeyer fornece alguns exemplos de construções que seriam mais tardias para crianças adquirirem, não por serem mais complexas, mas por serem mais sofisticadas, estilisticamente falando. Nesse sentido, o autor pergunta se há línguas estrangeiras mais difíceis de se aprender do que outras. Tais indagações nós debatemos no final do capítulo 3 da tese, bem como em Leal (2018), quando falamos que critérios estudados em aquisição de L2 podem também servir de métricas de CL.

Newmeyer também debate um pouco sobre algumas formas de se medir complexidade linguística, mas diferentemente do nosso capítulo 3, ele toma várias evidências de como algumas línguas se simplificaram, enquanto outras se complexificaram em contatos umas com as outras através dos tempos. Para falar disso, toma como base, novamente, McWhorter e, principalmente, Peter Trudgill. Ele finaliza sua fala do mesmo modo como estamos afirmando aqui: não há equicomplexidade entre as línguas. Assim como Newmeyer, nós mostramos aqui as razões para isso ao longo desta tese.

Esta tese é um dos primeiros trabalhos a versar sobre complexidade linguística no Brasil. Embora o assunto tenha sido debatido em diversos congressos, livros, artigos etc., no Brasil ainda

parece haver certa resistência para tratar do assunto. Acreditamos que o caso da CL não tenha sido devidamente resolvido no modelo estruturalista, então vigente no final da primeira metade do século XX, e essa crença levou outros linguistas a tomarem a equicomplexidade como ponto pacífico. Nesse sentido, podemos dizer que a complexidade linguística se torna um truísmo, se não um tabu entre muitos linguistas. Acreditamos, contudo, que os truísmos e os tabus devem ser esclarecidos ou combatidos pela ciência, incluindo, obviamente, a linguística. Procuramos, assim, mostrar que (i) a complexidade linguística pode e deve ser debatida; (ii) quais são as ferramentas e modos que podemos debater CL atualmente.

O que aprendemos nesta tese, entre outras coisas, foi que a complexidade linguística, tabu entre linguista por décadas, pode e deve voltar a ser objeto de estudo privilegiado na linguística moderna, isso porque hoje em dia temos diversas ferramentas, inclusive filosóficas para encarar de maneira completamente diferente do que víamos este fenômeno no passado. Se antes a CL tinha como barreira a capacidade dos seus falantes, hoje, através da genética, por exemplo, descobrimos que todo humano é rigorosamente igual. Linguisticamente falando, não há relatos de línguas cujo grau de complexidade seja muito inferior a qualquer outra língua comparada. O que os linguistas vêm fazendo são métricas minuciosas de complexidade linguística que vêm, cada vez mais, revelando justamente a riqueza das línguas naturais.

Referências Bibliográficas

- ALTMANN, Cristina. “Retrospectivas e perspectivas da historiografia linguística no Brasil”. In *Revista Argentina de Historiografia Linguística*, Vol.1, Nº 2, 2009. Disponível em: www.rahl.com.ar
- AMSTERDAMSKA, Olga. *Schools of Thought*. Norwell: Reidel Publishing Company, 1987.
- ARAGÃO, Shaltom & ARAÚJO, Edivalda. “O tópico retomado na posição de sujeito: diferenças entre o português brasileiro e o português europeu”. In: *A Cor das Letras*. Vol. 18, Nº 2. Feira de Santana: 2017.
- ARAÚJO, Edivalda Alves. *As Construções de Tópico do Português nos Séculos XVIII e XIX*. Tese (Doutorado) – Universidade Federal da Bahia, Instituto de Letras, Programa de Pós-Graduação em Letras e Linguística. Salvador: 2006.
- AUROUX, Sylvain. *A Revolução Tecnológica da Gramaticização*. Campinas: Editora da UNICAMP. 1992.
- BAERMAN, Matthew; BROWN, Dunstan & CORBETT, Gerville. *Understanding and Measuring Morphological Complexity*. London: Oxford University Press. 2015.
- BATISTA, Ronaldo de Oliveira. *Introdução à Historiografia da Linguística*. São Paulo: Cortez, 2013.
- BECCARI, Alessandro Jocelito. *Uma tradução da Grammatica Speculativa de Tomás de Erfurt para o português: acompanhada de um estudo introdutório, notas e glossário*. Tese (doutorado) - Universidade Federal do Paraná, Setor de Ciências Humanas, Letras e Artes, Programa de Pós-Graduação em Letras. Curitiba: 2013.
- BECCARI, Alessandro Jocelito & LEAL, Ednei de Souza. “Uma breve revisão dos antecedentes históricos da pressuposição de dois níveis da linguagem na sintaxe das gramáticas racionalistas portuguesas do final do século XVIII”. Trabalho apresentado no *VII MiniEnapol de Historiografia Linguística*. USP. São Paulo 2013. Disponível em: << http://cedoch.fflch.usp.br/sites/cedoch.fflch.usp.br/files/u65/CHLC1_0.pdf>>
- BOAS, Franz. *Antropologia Cultural*. Rio de Janeiro: Zahar. 2014.
- BORGES NETO, José. *Ensaio de Filosofia da Linguística*. São Paulo: Parábola. 2002.
- BULTÉ, Bram & HOUSEN, Alex. “Defining and Operationalizing L2 complexity”. In: HOUSEN, Alex, KUIKEN, Folkert & VEDDER, Ineke (orgs.). **Dimensions of L2 Performance and Proficiency**. Amsterdam: John Benjamins. 2012.
- CALLOU, Dinah *et al.* “Topicalização e Deslocamento à Esquerda: Sintaxe e Prosódia”. In: CASTILHO, Ataliba Teixeira de (org.). *Gramática do Português Falado. Vol.III: As Abordagens*, pp. 315-362. Campinas: UNICAMP: 2002.

- CHABOT, Yoan & NICOLLE, Christophe. “Semantic Measures”. In: KHOSROW-POUR, Mehdi (org.). *Encyclopedia of Information Science and Technology*. IGI Global: Hershey. 2015.
- CHOMSKY, Noam. *Aspectos de la Teoria de la Sintaxis*. Aguillar: Madrid, 1970 [1965].
- _____. *Linguística Cartesiana*. Vozes: Rio de Janeiro. 1973 [1966].
- CULICOVER, Peter W (org.). *Grammar & Complexity – Language at Intersection of Competence and Performance*. New York: Oxford. 2013.
- DAHL, Östen. *The Growth and Maintenance of Linguistic Complexity*. Amsterdan: John Benjamins. 2004.
- DAWKINS, Richard. *O Gene Egoísta*. São Paulo: Companhia das Letras. 2004 [1976].
- DIAS, Augusto Epiphanio. *Sintaxe histórica portuguesa*. Lisboa: Livraria Clássica Editora de A. M. Teixeira, 1918.
- DEUTSCHER, Guy. *Syntactic Change in Akkadian. The Evolution of Sentential Complementation*. Oxford University Press. Oxford. 2000.
- _____. “Overall Complexity: a Wild Goose Chase”. In: SAMPSON, Geoffrey; GIL, David; TRUDGILL, Peter (orgs.). *Language Complexity as an Evolving Variable*. New York: Oxford. 2009.
- _____. *Through the Language Glass*. New York: Metropolitan Book, 2010.
- _____. *O Desenrolar da Linguagem*. Tradução Renato Basso e Guilherme Henrique May. Campinas: Mercado de Letras, 2014.
- DEVITT, Michael. “Methodology in the Philosophy of Linguistics”. In: *Australasian Journal of Philosophy* Vol. 86, No. 4, pp. 671–684; December 2008.
- DIXON, R. M. W. *Searching for aboriginal languages: Memoirs of a field worker*. Chicago: University of Chicago Press, 1989.
- EDMONDS, Bruce. *Syntactic Measures of Complexity*. Tese (Doutorado). Universidade de Manchester, Departamento de Filosofia, Manchester. 1999.
- ELLIS, Rod. *The study of second language acquisition*. London: Oxford University Press. 1997.
- FARACO, Carlos Alberto. *Linguística Histórica*. São Paulo: Parábola. 2005.
- GALVES, Charlotte. “Tópicos, sujeitos, pronomes e concordância no português brasileiro”. In: *Cadernos de Estudos Linguísticos*. Editora da UNICAMP, n.34. Campinas: 1998.
- GELL-MANN, Murray. “What is Complexity”. In *Complexity*, Vol. 1, Nº. 1, John Wiley and Sons, Inc: London. 1995.
- _____. *O Quark e o Jaguar. As aventuras no simples e no complexo*. Rocco. Rio de Janeiro: 1996.
- GONÇALVES, Rodrigo Tadeu. *Perpétua prisão órfica ou Ênio tinha três corações: o Relativismo Linguístico e o Aspecto Criativo da Linguagem*. Tese (Doutorado) – Universidade Federal do

- Paraná, Setor de Ciências Humanas, Letras e Artes, Programa de Pós-Graduação em Letras. Curitiba: 2008.
- GREENBERG, Joseph. “Some Universals of Grammar with articular Reference to the Order of Meaningful Elements. *In: GREENBERG, Joseph. **Universals of Language**. London: MIT Press. 1963.*
- GUIMARÃES, Márcio Renato. “O Termo *Ariano* e a narrativa indo-europeia”. *In: Revista **Línguas e Letras**. UNIOESTE. Vol. 19, Nº 43. Cascavel: 2018.*
- _____. *Linguística no século XIX: Cientificidade e Narrativa das Origens*. Manuscrito gentilmente fornecido pelo autor. UFPR/DELLIN. Curitiba. 2020.
- HAJICOVA, Eva; PARTEE, Barbara H., SGALL, Petr (Orgs.). *Topic-Focus Articulation, Tripartite Structures, and Semantuc Content*. Springer-Science: 1998.
- HAWKINS. John A. “Major contributions from formal linguistics to the complexity debate”. *In: NEWMEYER, Frederick; PRESTON, Laurel (orgs.). **Measuring Grammatical Complexity**. London: Oxford. 2014.*
- HEIDERMAN, Werner; WEININGER, Markus. *Humboldt: Linguagem, Literatura e Bildung*. Florianópolis: UFSC. 2006.
- HEYLIGHEN, Francis; CILLIERS, Paul; GERSHENSON, Carlos. “Philosophy and Complexity”. *In: BOGG, Jan & GEYER, Robert. **Complexity, Science and Society**. CRC Press. London: 2007.*
- HOCKET, Charles. *A Course in Modern Linguistics*. Oxford: IBH Publishing & Co. 1958 [1954].
- HONG HAN, Zhao & WAI MAN, Lew. “Acquisitional Complexity – What defies complete acquisition in Second Language Acquisition”. *In: HOUSEN, Alex, KUIKEN, Folkert & VEDDER, Ineke (orgs.). **Dimensions of L2 Performance and Proficiency**. Amsterdam: John Benjamins. 2012.*
- HOUSEN, Alex, KUIKEN, Folkert & VEDDER, Ineke (orgs.). *Dimensions of L2 Performance and Proficiency*. Amsterdam: John Benjamins. 2012.
- ILARI, Rodolfo. *Perspectiva funcional da frase portuguesa*. 2. ed. Campinas: Ed. UNICAMP, 1992.
- JIMÉNEZ, Carmem Conti. “La Teoria de los Estadios de la Lengua y la Grafación de la Complejidad Linguística”. *In **Revista Española de Lingüística**, Nº 41, Fasc. 2, 2011.*
- JAKOBSON, Roman. *Linguística e comunicação*. Sao Paulo: Cultrix, 1979.
- JOSEPH, John & NEWMEYER, Frederick. “All Language are Equally Complex’ The Rise and Fall of a Consensus”. *In: **Historiographia Linguistica** xxxix: 2/3. (pp.: 341–368). Washington: John Benjamins Publishing Company. 2012. <doi 10.1075/hl.39.2.08jos>*
- JUBRAN, Clélia Cândida Abreu Spinardi, “Tópico discursivo”, *In: JUBRAN, Clélia Cândida Abreu Spinardi & KOCH, Ingedore Grunfeld Villaça. **Gramática do Português Culto Falado no Brasil – Volume 1: Construção do Texto Falado**. Campinas: Editora da UNICAMP. 2006.*

- KATO, M. A. “Tópico e sujeito: duas categorias na sintaxe?” *In: Cadernos de Estudos Linguísticos*. Campinas: Editora da Unicamp, n. 17. 1989.
- KEMPSON, Ruth; FERNANDO, Tim; ASHER, Nicholas. *Philosophy of Linguistics*. Oxford: Elsevier. 2012.
- KILARSKI, Marcin. “Complexity in the history of language study”. *In: Poznań Studies in Contemporary Linguistics* 50: 2. (pp.: 157–168). Faculty of English, Adam Mickiewicz University, Poznań, Poland. 2014. <doi:10.1515/psicl-2014-0011>
- KOERNER, E.F.K & ASHER, R.E. *Concise History of the Language Sciences – From the Sumerians to the Cognitivists*. London: Pergamon, 1995.
- KUHN, Thomas S. **A estrutura das revoluções científicas**. 2ª ed. São Paulo: Perspectiva, 1982.
- KUSTERS, Christiaan Wouter. *Linguistic Complexity: The Influence of Social Change on Verbal Inflection*. Tese (Doutorado). Universidade de Leiden: LOT. 2003.
- LEAL, Ednei de Souza. *Pressupostos epistemológicos na "phraseologia" dos Serões Gramaticaes de Ernesto Carneiro Ribeiro*. Dissertação (mestrado) - Universidade Federal do Paraná, Setor de Ciências Humanas, Letras e Artes, Programa de Pós-Graduação em Letras: Curitiba, 2014.
- _____. “Complexidade Linguística e Aquisição de L2: Panorama Teórico”. *In: Working Papers – Universidade Federal de Santa Catarina*. Florianópolis: 2018. Disponível em << <https://periodicos.ufsc.br/index.php/workingpapers/article/view/1984-8420.2018v19n1p136>>>.
- LEITE, Yonne *et al.* “Tópicos e Adjuntos”. *In: CASTILHO, Ataliba Teixeira de; BASÍLIO, Margarida (orgs.). Gramática do Português Falado – Volume IV: Estudos Descritivos*. Campinas: Editora da UNICAMP. 2002.
- LI, Charles & THOMPSON, Sandra. *Subject and Topic: A New Typology of Language*. New York: Academic Press. 1976.
- MARTINS, Janete. *Aspectos Epistemológicos das Teorias sobre a Formação do Português Brasileiro*. Dissertação (Mestrado) – Universidade Federal de Santa Catarina. Programa de Pós-graduação em Linguística. 2013.
- McREADY, Eric. “Expressive and Expressivity”, *In: Open Linguistics*, Nº 1. 2014. Disponível em: << <https://www.degruyter.com/view/journals/opli/open-issue/article-10.2478-opli-2014-0004/article-10.2478-opli-2014-0004.xml>>>.
- McWHORTER, John. “The world’s simplest grammars are creole grammars”. *In: Linguistic typology* 5, 125–166. 2001.
- _____. *Defining Creole*. New York: Oxford. 2005.
- _____. *Language Interrupted: Signs of Non-Native Acquisition in Standard Language Grammars*. New York: Oxford. 2007.

- MIESTAMO, Matti; SINNEMÄKI, Kaius; KARLSSON, Fred (orgs.). *Language Complexity: Typology, contact, change*. Amsterdam: John Benjamins. 2008.
- MIESTAMO, Matti. “Implicational hierarchies and grammatical complexity”. In: SAMPSON, Geoffrey; GIL, David; TRUDGILL, Peter (orgs.). *Language Complexity as an Envolving Variable*. New York: Oxford. 2009.
- MODESTO, Marcelo. “Sujeitos Nulos em Línguas de Tópico Proeminente”. In: *Revista ABRALIN*, vol. III, nº 1. 2004.
- MOUNIN, Georges. *História da linguística: das origens ao século XX*. Porto: Despertar, 1970.
- MUFWENE, S.S. “The Emergence of Complexity in Language: An Evolutionary Perspective”. In: MASSIO-BONET À., BASTARDAS-BOADA A. (Orgs.) *Complexity Perspectives on Language, Communication and Society. Understanding Complex Systems*. Berlin, Heidelberg: Springer. 2013.
- NEELEMAN, Ad & VERMEULEN, Reiko. *The Syntax of Topic, Focus, and Contrast: An Interface-based Approach*. Boston: De Gruyter Mouton. 2012.
- NEGRÃO, Esmeralda Vailati. “A Cartografia Sintática”. In: FIORIN, José Luiz (Org.). *Novos Caminhos da Linguística*. São Paulo: Contexto. 2017.
- NEWMEYER, Frederick & PRESTON, Laurel (orgs.). *Measuring Grammatical Complexity*. London: Oxford. 2014.
- NEWMEYER, Frederick. “Can One Language Be ‘More Complex’ Than Another?” Palestra proferida no evento Abralín ao vivo: Linguists online. 2020.
- NICHOLS, Johanna. “Linguistic complexity: a comprehensive definition and survey”. In: SAMPSON, Geoffrey; GIL, David; TRUDGILL, Peter (orgs.). *Language Complexity as an Envolving Variable*. New York: Oxford. 2009.
- NIEMINEN, Lea. “MLU and IPSyn Measuring Absolute Complexity”, In EESTI Rakenduslingvistika Ühingu aastaraamat. *Estonian Papers in Applied Linguistics*. Vol 5. 2009.
- OKRENT, Arika. *In the Land of Invented Languages: Esperanto Rock Stars, Klingon Poets, Loglan Lovers, and the Mad Dreamers Who Tried to Build A Perfect Language*: Bantam Books, 2009.
- ORSINI, Mônica Tavares. *As Construções de Tópico no Português do Brasil: Uma Análise Sintático-discursiva em tempo real*. UFRJ. Rio de Janeiro, 2014. Disponível em << <http://www.filologia.org.br/viiicnlf/anais/caderno05-07.html>>>
- OTHERO, Gabriel d’Ávila. *Mitos de Linguagem*. São Paulo: Parábola. 2017.
- OTHERO, Gabriel & MENUZZI, Sérgio de Moura. *Linguística Computacional: Teoria e Prática*. São Paulo: Parábola. 2005.

- PASSOS, Equeni Sobrinha Rios & FONSECA, Fernanda Figueira. “As interfaces nas estratégias de Construção de Tópico”. In: *Revista Letra Magna*. Nº 14. 2011.
- PONTES, Eunice. *O Tópico no Português do Brasil*. Campinas: Pontes. 1987.
- PRIA, Albano Della. “Tipologia Linguística: Línguas Analíticas e Línguas Sintéticas”. In: *Revista Soletras*. Ano 6, Nº 11. UERJ. São Gonçalo: 2006.
- ROBERTS, Graige. “Topics”. In: MAEINBORN, Claudia; HEUSINGER, Klaus von & PORTER, Paul. *Semantics: An International Handbook of Natural Language Meaning – Volume 3*. Berlin: De Gruyter Mouton. 2012.
- ROBINS, R. H. (Robert Henry). *Pequena história da Linguística*. Rio de Janeiro: Ao Livro Técnico, 1979.
- SAMPSON, Geoffrey; GIL, David; TRUDGILL, Peter (orgs.). *Language Complexity as an Evolving Variable*. New York: Oxford. 2009.
- SAID ALI, Manoel. *Gramática secundária e gramática histórica da língua portuguesa*. Brasília: Ed. Universidade de Brasília, 1964.
- SAPIR, Edward. *Linguística como Ciência*. Rio de Janeiro: Livraria Acadêmica. 1969.
- SILVA, Jair Barbosa da; MAURÍCIO, Nadir Marques & SILVA, Aguimario Pimentel. “Construções de Tópico Marcado em Gramáticas do Português: Revisão da Literatura”. In: Atas do XVII Congresso Interntacional Asociación de Linguística y Filología da América Latina (ALFAL 2014). João Pessoa: 2017.
- SOKAL, Alan & BRICMONT, Jean. *Imposturas Intelectuais: O abuso da Ciência pelos filósofos pós-modernos*. Rio de Janeiro: Record. 2006.
- SWIGGERS, Pierre. “La historiografía de la lingüística: apuntes y reflexiones”. In: *Revista argentina de historiografía lingüística*. Nº 1, Vol. I. Buenos Aires: 2009.
- SZMRECSANYI, Benedikt & KORTMANN, Bernd (orgs.). *Linguistic complexity – Second Language Acquisition, indigenization, contact*. Berlin: DeGruyter. 2012.
- TARALLO, Fernando & KATTO, Mary et al. “Rupturas na Ordem de Adjacências Canônicas no Português Falado”, In: CASTILHO, Ataliba Teixeira de (org.). *Gramática do Português Falado. Vol.I: A Ordem*, pp. 25-52. Campinas: UNICAMP: 1991.
- VULANOVIC, Relja. ‘On measuring language complexity as relative to the conveyed linguistic information’. In: *SKY Journal of Linguistics*. 2007.