

UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGÜÍSTICOS

VANESSA CRISTINA OLIVEIRA WRIGHT NITSCH

**COMPLEXIDADE DOS SINTAGMAS NOMINAIS DO INGLÊS:
UM ESTUDO COMPARATIVO DE *CORPORA* DE APRENDIZES BRASILEIROS E
FALANTES NATIVOS DE INGLÊS**

BELO HORIZONTE

2017

VANESSA CRISTINA OLIVEIRA WRIGHT NITSCH

**COMPLEXIDADE DOS SINTAGMAS NOMINAIS DO INGLÊS:
UM ESTUDO COMPARATIVO DE *CORPORA* DE APRENDIZES BRASILEIROS E
FALANTES NATIVOS DE INGLÊS**

Tese apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial para obtenção de título de doutora em Linguística teórica e descritiva.

Área de concentração: Linguística teórica e descritiva.

Linha de pesquisa: Estudos da língua em uso

Orientadora: Profa. Dra. Adriana Maria Tenuta de Azevedo.

Belo Horizonte

Faculdade de Letras da UFMG

2017

N732c

Nitsch, Vanessa Cristina Oliveira Wright.

Complexidade dos sintagmas nominais do inglês [manuscrito] : um estudo comparativo de *corpora* de aprendizes brasileiros e falantes nativos de inglês / Vanessa Cristina Oliveira Wright Nitsch. – 2017.

165 f., enc.: il., tabs, grafs.

Orientadora: Adriana Maria Tenuta de Azevedo.

Área de concentração: Linguística Teórica e Descritiva.

Linha de Pesquisa: Estudos da Língua em Uso.

Tese (doutorado) – Universidade Federal de Minas

Gerais, Faculdade de Letras.

Bibliografia: f. 146-152.

Anexos: f. 153-165.

1. Língua inglesa - Sintagma nominal – Teses. 2. Redação acadêmica – Teses. 3. Linguística de corpus – Teses. 4. Língua inglesa – Gramática gerativa – Teses. 5. Língua inglesa – Estudo e ensino – Falantes estrangeiros – Teses. I. Azevedo, Adriana Maria Tenuta de. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.

CDD: 425



UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGÜÍSTICOS



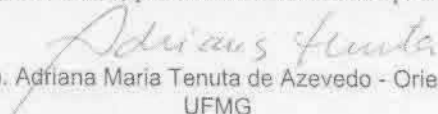
FOLHA DE APROVAÇÃO

**COMPLEXIDADE DOS SINTAGMAS NOMINAIS DO INGLÊS:
um estudo comparativo de corpora de aprendizes brasileiros e
falantes nativos de inglês**

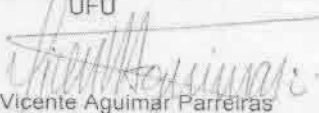
VANESSA CRISTINA OLIVEIRA WRIGHT NITSCH

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGÜÍSTICOS, como requisito para obtenção do grau de Doutor em ESTUDOS LINGÜÍSTICOS, área de concentração LINGÜÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Estudos da Língua em Uso.

Aprovada em 11 de dezembro de 2017, pela banca constituída pelos membros:


Prof(a). Adriana Maria Tenuta de Azevedo - Orientador
UFMG


Prof(a). Guilherme Fromm
UFU


Prof(a). Vicente Aguiar Parreiras
CEFET/MG


Prof(a). Bárbara Malveira Orfanó
UFMG


Prof(a). Marisa Mendonça Carneiro
UFMG

Belo Horizonte, 11 de dezembro de 2017.

AGRADECIMENTOS

À professora Adriana Tenuta, por ter me apresentado aos estudos linguísticos, à sintaxe e à semântica do inglês. Agradeço também pelo tempo de orientação, dedicação, conselhos e paciência.

Aos professores Marisa Carneiro e Guilherme From, pelos comentários e sugestões feitos no exame de qualificação.

A todos os meus professores, que, de alguma forma, contribuíram para a minha formação acadêmica. Professores da E. E. Ari da Franca, professores do CEFET-MG e professores da Faculdade de Letras da UFMG.

À minha família e amigos. Em especial aos meus irmãos, pela disposição em me auxiliar no que fosse preciso; ao meu pai, pelo apoio; à minha mãe, pelo amor incondicional, suporte, carinho e conselhos; e ao Matthias, por seu valioso apoio à minha formação acadêmica.

RESUMO

O presente trabalho busca descrever a complexidade dos sintagmas nominais produzidos por aprendizes brasileiros de inglês, principalmente no uso de modificadores. Os conceitos utilizados nesta pesquisa se baseiam em gramáticas descritivas da língua inglesa, trazendo exemplos da língua em uso, e baseados em *corpora* de referência. Três *corpora* foram utilizados neste trabalho, a saber: o BR-ICLE (*International Corpus of Learner English*, o *corpus* de estudo), o LOCNESS (*Louvian Corpus of Native English Essays*, o *corpus* comparativo) e o BAWE (*British Academic Written English*, o *corpus* de referência), que representam a produção linguística dos aprendizes brasileiros, a produção linguística de falantes nativos americanos e britânicos e a produção linguística de universitários britânicos de várias áreas de estudo, respectivamente. O trabalho propôs a classificação dos níveis de complexidade dos SNs e comparou os dados obtidos no *corpus* de estudo com o *corpus* comparativo, com base na classificação de complexidade. Além disso, utilizamos a ferramenta de linguística de *corpus* *WordSmith Tools* para conduzir a pesquisa. Esta pesquisa foi realizada com base na hipótese (confirmada) de que a escrita dos aprendizes brasileiros se mostraria mais simples do que a dos falantes nativos de inglês. Essa menor complexidade se deu na densidade do SN produzido pelos aprendizes. A escrita do aprendiz se caracterizou por ser fortemente pré-modificada, em detrimento da pós-modificação. Essa pré-modificação foi mais frequente com adjetivos, e menos frequente com substantivos. Além disso, os aprendizes utilizaram menos orações não-finitas e sintagmas preposicionados como pós-modificadores em comparação aos nativos.

Palavras-chave: complexidade do sintagma nominal; modificação; *corpus* de aprendiz; escrita acadêmica.

ABSTRACT

This study describes the complexity of noun phrases produced by Brazilian learners, more specifically, the noun modifiers. The concepts used in this research are based on descriptive grammars of English which have a functionalist bias, bringing examples of language in use, or which are based on reference corpora. Three *corpora* were used in this work: BR-ICLE (International Corpus of Learner English, study *corpus*), LOCNESS (Louvian *Corpus* of Native English Essays, comparative *corpus*) and BAWE (British Academic Written English, reference *corpus*), which represent the linguistic production of Brazilian apprentices, the linguistic production of American and British English native speakers, and the linguistic production of British university students from various fields of study, respectively. I propose a measurement for noun phrase complexity, and analyze data in native and learner *corpus* with this measurement. In addition, WordSmith Tools corpus linguistics tool was used to conduct the research. This research was carried out based on the hypothesis that learner's noun phrases are structured differently from native speakers, and that they tend to be less complex. This lower complexity occurred in the density of the noun phrase produced by the learner. The learner's writing was characterized by being strongly pre-modified, to the detriment of post-modification. Pre-modification occurred more often as adjectives, and it occurred less frequently as nouns. In addition, learners used fewer non-finite clauses and prepositional phrases as post-modifiers, compared to natives.

Key-words: complex noun phrase; noun modification; learner *corpus*; academic writing.

LISTA DE FIGURAS

Figura 1: Nome próprio com variação na estrutura	24
Figura 2: Os tipos de substantivos	25
Figura 3: A ferramenta <i>WordSmith Tools</i>	72
Figura 4: Exemplo de lista de palavras	73
Figura 5: Exemplo de lista de palavras-chave	75
Figura 6: Exemplo de linhas de concordância	75
Figura 7: Exemplo do recurso <i>plot</i> do corpus LOCNESS	76
Figura 8: Exemplo do recurso <i>Patterns</i> do corpus LOCNESS	77
Figura 9: Configurações do <i>WordSmith</i> para geração de palavras-chave.....	79
Figura 10: Geração das listas de palavras-chave	80
Figura 11: Exemplo de lista de palavras-chave do BR-ICLE gerada com o recurso <i>Plot</i>	81
Figura 12: Seleção dos SNs com núcleo -ing sem modificação	87
Figura 13: Seleção dos SNs com núcleo -ing com modificação	87
Figura 14: Interface do programa R	92

LISTA DE QUADROS

Quadro 1: Tipos de determinantes	31
Quadro 2: Fases hipotéticas do desenvolvimento para características de complexidade do sintagma nominal	46
Quadro 3: Os tipos mais comuns de <i>corpora</i>	54
Quadro 4: Os 10 nomes em comum na lista de palavras-chave do <i>LOCNESS</i> e do BR-ICLE	85
Quadro 5: Determinantes pesquisados	88
Quadro 6: Complexidade do sintagma nominal	89
Quadro 7: Resumo das características do sintagma nominal produzido pelos aprendizes	142

LISTA DE GRÁFICOS

Gráfico 1: Relação de SNs com e sem modificação	98
Gráfico 2: Relação dos SNs e sua complexidade	99
Gráfico 3: Relação das complexidades entre os SNs com modificação	100

Gráfico 4: Relação das complexidades dos SNs com núcleo Substantivo	101
Gráfico 5: Relação das complexidades dos SNs com núcleo -ing	102
Gráfico 6: Relação dos núcleos -ing realizados pelos 4 termos mais comuns	103
Gráfico 7: Relação dos pré-modificadores	105
Gráfico 8: Relação dos tipos de pré-modificadores	106
Gráfico 9: Relação dos tipos de pós-modificadores	110
Gráfico 10: Relação dos pós-modificadores realizados por sintagmas preposicionados	112
Gráfico 11: Relação dos tipos de pronomes relativos introdutores de orações finitas como pós-modificadores	114
Gráfico 12: Relação dos tipos de pós-modificadores realizados por orações não-finitas	116
Gráfico 13: Relação dos pronomes pessoais de sujeito	120
Gráfico 14: Relação dos pronomes pessoais de objeto	121
Gráfico 15: Complexidade dos SNs com função de complemento de sintagma preposicionados, modificadores de SN	124
Gráfico 16: Relação dos tipos de modificadores de SN com função de complemento de SPrep	126

LISTA DE TABELAS

Tabela 1: Gêneros de escrita acadêmica por área	68
Tabela 2: Total de palavras dos <i>corpora</i>	95
Tabela 3: Frequência dos termos pesquisados	95
Tabela 4: Total de SNs analisados e modificados	96
Tabela 5: Frequência das complexidades	97
Tabela 6: Frequência dos tipos de pré-modificadores	104
Tabela 7: Frequência dos tipos de pós-modificadores	110
Tabela 8: Frequência dos pronomes pessoais	119

LISTA DE ABREVIações

BR-ICLE – Brazilian International Corpus of Learner English

BAWE – British Academic Written English

Co – Predicativo do objeto

COCA – Corpus of contemporary American English

Cprep – Complemento da preposição

Cs – Predicativo do sujeito

GED – General Educational Development

L1 – Primeira língua

L2 – Segunda língua

LOCNESS – Louvian Corpus of Native English Essays

NN – Substantivo modificador de substantivo

Ora. Fin. – Oração finita

Ora. não Fin. – Oração não-finita

Sadj – Sintagma adjetivo

Sadv – Sintagma adverbial

SN – Sintagma nominal

SPrep – Sintagma preposicional

SV – Sintagma verbal

SUMÁRIO

1 CONSIDERAÇÕES INICIAIS	10
1.1 OBJETO DE ESTUDO	11
1.2 MOTIVAÇÃO	11
1.3 PERGUNTAS DE PESQUISA E HIPÓTESE	13
1.4 OBJETIVOS	13
1.5 JUSTIFICATIVA	14
1.6 PESQUISAS BASEADAS EM PRODUÇÃO DE APRENDIZES	14
1.7 ORGANIZAÇÃO DOS CAPÍTULOS	15
2 O SINTAGMA NOMINAL	17
2.1 OS ELEMENTOS DA ORAÇÃO NA LÍNGUA INGLESA	17
2.2 OS ELEMENTOS DO SINTAGMA NOMINAL DO INGLÊS	20
2.2.1 O núcleo do sintagma nominal	21
2.2.1.1 Os substantivos	22
2.2.1.2 Os pronomes	26
2.2.1.3 Os nomes verbais	28
2.2.2 Determinantes do sintagma nominal	30
2.2.3 Os modificadores do sintagma nominal	32
2.2.3.1 Pré-Modificadores	32
2.2.3.2 Pós-Modificadores	36
2.3 AS FUNÇÕES EXERCIDAS PELO SINTAGMA NOMIAL	40
2.4 COMPLEMENTO DE SUBSTANTIVO	41
2.5 SINTAGMAS NOMINAIS COMPLEXOS	42
3 LINGUÍSTICA DE CORPUS E ESCRITA ACADÊMICA	51
3.1 DEFINIÇÃO DE <i>CORPUS</i>	51
3.2 BREVE HISTÓRICO DA LINGUÍSTICA DE <i>CORPUS</i>	52
3.3 TIPOLOGIA	53
3.4 TAMANHO DO <i>CORPUS</i>	58
3.5 ESCRITA ACADÊMICA E <i>CORPUS</i> DE APRENDIZES	59
4 METODOLOGIA	65
4.1 OS CORPORA DE REFERÊNCIA, ESTUDO E COMPARATIVO	65
4.1.1 O <i>corpus</i> de referência	66
4.1.2 O <i>corpus</i> de aprendizes brasileiros	69
4.1.3 O <i>corpus</i> comparativo de falantes nativos	70

4.2 SOFTWARES UTILIZADOS	71
4.2.1 O <i>Wordsmith Tools</i>	72
4.3 PROCEDIMENTOS METODOLÓGICOS	78
4.3.1 As palavras-chave	78
4.3.2 A busca por termos que sejam núcleos de SN	83
4.3.2.1 A seleção dos sintagmas nominais	85
4.3.2.2 A seleção dos núcleos de sintagma nominal realizados por nomes verbais (-ing)	86
4.3.2.3 A seleção dos sintagmas nominais realizados por pronomes	88
4.3.3 Proposta de classificação das complexidades do sintagma nominal	89
4.3.4 Teste estatístico e normalização dos dados	92
5 DESCRIÇÃO E DISCUSSÃO DOS RESULTADOS	94
5.1 RESULTADOS GERAIS	94
5.1.1 Modificação dos nomes verbais	101
5.2 OS PRÉ-MODIFICADORES	104
5.2.2 Estruturas atípicas encontradas na produção dos aprendizes referentes à pré-modificação de sintagma nominal	107
5.3 OS PÓS-MODIFICADORES	109
5.3.1 Pós-modificadores realizados por sintagmas preposicionados	111
5.3.2 Pós-modificadores realizados por orações	114
5.3.3 Estruturas atípicas encontradas na produção dos aprendizes referentes à pós-modificação de sintagma nominal	117
5.4 O USO DE PRONOMES	119
5.5 OS COMPLEMENTOS DOS SINTAGMAS PREPOSICIONADOS MODIFICADORES DE SNS	123
5.6 DISCUSSÃO DOS RESULTADOS OBTIDOS	126
6 PROPOSTA DE ATIVIDADES DIDÁTICAS	131
6.1 Trabalhando com o conceito de modificação	132
6.2 Criando sintagmas nominais complexos a partir de textos autênticos	137
CONSIDERAÇÕES FINAIS	141
REFERÊNCIAS	146
ANEXOS	153

CONSIDERAÇÕES INICIAIS

Após anos de ensino de inglês como segunda língua para aprendizes brasileiros, notei que os alunos apresentam certa dificuldade para lidar com o sintagma nominal (SN) complexo da língua-alvo. Por SN complexo me refiro a SNs que possuem alguma modificação, principalmente SNs que possuem várias modificações ao mesmo tempo. Essa dificuldade parecia se refletir tanto na leitura quanto na escrita dos meus alunos. Como o texto acadêmico possui uma ampla gama de SNs mais complexos (BIBER et al., 1999), assim como textos de outros gêneros possuem, o SN e seus elementos merecem atenção.

O SN é descrito, de forma sucinta, como uma unidade linguística que possui como núcleo um substantivo ou um pronome. Seus elementos são, além do núcleo, os modificadores e determinantes. O único elemento obrigatório do SN é o núcleo; porém, a presença dos outros elementos pode tornar o SN uma unidade de enorme complexidade. Além disso, o SN exerce inúmeras funções diferentes em uma oração, como sujeito, objetos e complementos. Assim, o SN, como outros sintagmas, serve para desenvolver um único elemento dentro de uma oração ou de outro sintagma (HALLIDAY, 2014).

Buscando entender como o aprendiz brasileiro de inglês estrutura o SN e seus modificadores nessa língua, utilizei, para o presente estudo, um *corpus* de aprendiz brasileiro de inglês, para descrever, portanto, o uso de SNs. Assim, busco trazer importantes contribuições para o entendimento da interlíngua dos aprendizes brasileiros de inglês, principalmente no uso que eles apresentam na complexidade do sintagma nominal.

O termo interlíngua foi cunhado por Selinker (1972), que o definiu como a língua do aprendiz de uma segunda língua, ou uma língua estrangeira, composta por um sistema de regras linguísticas próprio. Ortega (2009, p. 6), que também trata da interlíngua como a “gramática mental dos aprendizes”, afirma que essa gramática gera uma variedade especial de linguagem quando os aprendizes produzem a segunda língua ao falar, cantar, escrever ou interagir com outras pessoas. Essa tentativa do aprendiz de produzir a língua-alvo é, então, baseada em representações mentais da nova gramática (interlíngua) que eles criaram (ORTEGA, 2009).

Neste capítulo, serão apresentados o objeto de estudo, a motivação para a investigação, as perguntas de pesquisa, a hipótese, os objetivos do trabalho, a justificativa para o tema escolhido e um breve panorama de estudos baseados em *corpus*.

1.1 Objeto de Estudo

Visando descrever as características do sintagma nominal produzido pelo aprendiz brasileiro de inglês, investiguei a estrutura dos modificadores do sintagma nominal produzido por ele. Para conduzir tal estudo, comparei a produção de aprendizes universitários brasileiros com a produção de universitários, falantes nativos de inglês. Essa comparação não foi feita de forma normativa, mas, sim, descritiva, para se ter uma maior clareza acerca da frequência dos tipos diferentes de complexidade que um SN pode ter.

A escolha dos SNs como objeto de estudo se deve ao interesse em investigar a escrita de aprendizes brasileiros. Uma das formas de se medir o nível de complexidade de um texto é pelo SN e seus modificadores (BIBER, GRAY, POONPON, 2011). Ou seja, quanto mais SNs complexos são produzidos, mais elaborada a escrita, e eu busco, através deste estudo, entender e descrever a complexidade dos SNs produzidos por aprendizes brasileiros de inglês.

1.2 Motivação

O presente trabalho tem como origem e motivação um trabalho anterior de Wright (2014), minha dissertação de mestrado, em que um *corpus* de aprendizes e

outro de falantes nativos de inglês foram analisados. O objeto de estudo naquele momento foram os determinantes de sintagmas nominais (SN). Foram utilizados dois *corpora*. O primeiro *corpus* era composto por ensaios acadêmicos produzidos por aprendizes brasileiros, enquanto o segundo era composto por ensaios acadêmicos produzidos por falantes nativos de inglês britânico e americano.

Os dados mostraram que a escrita dos aprendizes tende a ser mais simples, apresentando menor variação lexical, e mostrando algumas influências da sua língua materna. Apesar de essa pesquisa ter sido a respeito dos determinantes do SN, um fato chamou atenção da autora durante a análise dos dados: em uma amostra de 200 sintagmas nominais analisados, os aprendizes produziram SNs com modificadores muito mais simples do que os produzidos pelos nativos. Nesse trabalho, Wright (2014) atentou para a diferença apenas entre uma pequena amostra de pós-modificadores produzidos nesses dois *corpora*. Essa diferença foi em relação à frequência de uso, comprovada estatisticamente, já que os nativos produziram uma maior quantidade de pós-modificadores do que os aprendizes, enquanto a produção dos aprendizes apresentou uma maior quantidade de pré-modificadores que a dos nativos. Isso sugere que há diferenças em relação à modificação dos SNs produzidos por esses dois grupos e que esse fato merece mais atenção. Essa constatação foi um dos elementos motivadores para a criação deste projeto.

O segundo elemento motivador para este projeto foi uma constatação feita em minha prática docente em sala de aula. Durante o curso de sintaxe do inglês, ministrada por mim na UFMG, percebi que os alunos tinham uma incrível dificuldade de identificar o núcleo de sintagmas nominais mais complexos, como, por exemplo, aqueles pós-modificados com uma oração finita. E quando conseguiam identificar o núcleo, não conseguiam identificar os pós-modificadores mais complexos como sendo pós-modificadores. Nessas turmas, os alunos da habilitação em inglês já estavam mais avançados em seu curso, possuíam pelo menos um nível intermediário de proficiência na língua e alguns já haviam até tido experiências no exterior, através de intercâmbios em universidades americanas. Isso fez com que eu me perguntasse se eles também deixariam de produzir essas estruturas, em seus próprios textos, já que não as identificavam nos textos que analisavam.

Com base nessas constatações, proponho, através deste estudo, uma análise comparativa, baseada em *corpora*, do uso de modificadores de sintagmas nominais por aprendizes de inglês e por falantes nativos dessa língua.

1.3 Perguntas de pesquisa e hipótese

Com o presente trabalho, pretendo responder às seguintes perguntas de pesquisa:

- Os aprendizes brasileiros produzem SNs complexos em seus textos na língua inglesa?
- Como as estruturas dos SNs produzidos pelos aprendizes se comparam com as estruturas dos SNs produzidos por falantes nativos?

Embora este trabalho seja de ordem descritiva, ele tem aplicação indireta no ensino da língua inglesa, seja para auxiliar na elaboração de materiais didáticos, seja na conscientização de professores sobre as características da produção linguística de seus aprendizes e, portanto, na conscientização da necessidade de se abordar pedagogicamente a estrutura do SN.

Esta pesquisa foi feita com base na hipótese de que os sintagmas nominais produzidos pelos aprendizes brasileiros se mostrarão diferentes dos sintagmas nominais produzidos por falantes nativos de inglês no que se refere à modificação. A hipótese é que os SN mais complexos (principalmente com pré-modificação e com pós-modificação ao mesmo tempo) serão raros na produção dos aprendizes e que estruturas mais complexas, como pós-modificação realizada por orações e sintagmas preposicionados, serão menos utilizadas em comparação com estruturas mais simples. Além disso, espera-se que os aprendizes apresentem frequência significativamente maior de SNs sem modificação, em comparação com os nativos.

1.4 Objetivos

Este estudo tem como objetivo geral descrever os sintagmas nominais produzidos por aprendizes brasileiros, em escrita acadêmica, no que se refere à modificação do núcleo nominal, relativa tanto à estrutura dos pré- quanto à dos pós-

modificadores. E, através dessa descrição, alcançar os seguintes objetivos específicos:

- Propor categorias referentes aos graus de complexidade do SN a partir de sua complexidade estrutural;
- Quantificar os SNs selecionados dos *corpora* utilizados de acordo com as categorias de complexidade propostas;
- Comparar as complexidades dos sintagmas nominais produzidos pelos aprendizes brasileiros com as complexidades dos SNs produzidos pelos falantes nativos de inglês;
- Analisar qualitativamente os elementos modificadores dos sintagmas nominais produzidos pelos aprendizes, focando nas estruturas atípicas produzidas por eles.

1.5 Justificativa

Embora haja trabalhos que descrevam os SNs complexos produzidos por falantes aprendizes de uma segunda língua, há uma falta de pesquisas desse tipo para a produção de aprendizes brasileiros de inglês. Em uma pesquisa no estado da arte, encontrei trabalhos que comparam os SNs complexos do inglês, da língua escrita com a língua oral (BIBER, GRAY, 2010; BIBER, GRIEVE, IBERRI-SHEA, 2010; BIBER, GRAY, POONPON, 2011), trabalhos que comparam a complexidade de SNs produzidos entre estudantes universitários em início e final de curso (PARKINGSON, MUSGAVE, 2014), trabalhos que descrevem a complexidade de SNs na produção de aprendizes espanhóis (PASTOR, 2008) e também trabalhos que analisam a complexidade dos SNs em materiais didáticos de cursos superiores em universidades americanas (O'MALLEY, 2013). Estes trabalhos, junto com os seus resultados, serão descritos mais detalhadamente no próximo capítulo.

1.6 Pesquisas com base em produção de aprendizes

Os estudos baseados em *corpora* têm sido desenvolvidos há muitos séculos, e o avanço tecnológico contribuiu para o desenvolvimento das ferramentas para

estudos linguísticos. Além disso, as ferramentas da linguística de *corpus* viabilizaram um estudo mais detalhado em *corpora* maiores. Esse maior detalhamento se dá pela automatização da busca de termos e construções (BERBER SARDINHA, 2004), além de outras funcionalidades, como busca de colocados, palavras-chave de um *corpus*, lista de frequência das palavras, dentre outros. Com o surgimento dos estudos baseados em *corpora* de aprendizes, passa-se a ter maior conhecimento das características da produção desses falantes, de seus erros, assim como das semelhanças e diferenças dessa produção com a língua dos falantes nativos.

Utilizando ferramentas de linguística de *corpus* e analisando *corpus* de aprendiz, a pesquisa aqui proposta vem se unir a esses trabalhos mencionados e a outros que contribuem para uma maior compreensão da linguagem utilizada por aprendizes da língua inglesa, mais especificamente aprendizes brasileiros. O trabalho visa proporcionar a chance de reflexão sobre as características dessa linguagem, suas motivações e possibilidades de alteração de padrões indesejados.

1.7 Organização dos capítulos

Este trabalho foi dividido em 6 capítulos. O presente capítulo apresenta as considerações iniciais, contendo objetivo, perguntas de pesquisa, hipótese, além de uma breve descrição do objeto de estudo.

O segundo contempla o sintagma nominal do inglês, recorrendo sobre os conceitos básicos, as funções desempenhadas pelo sintagma, as suas realizações etc. Além disso, discuto, nesse capítulo, os estudos feitos até o momento sobre a complexidade do sintagma nominal.

O terceiro capítulo consiste em um levantamento teórico da linguística de *corpus* e da escrita acadêmica. Além disso, nesse capítulo, descrevo trabalhos que lidam com aspectos da escrita de aprendizes de inglês. Abordo trabalhos relacionados com a linguística de *corpus* e também trabalhos relacionados com *corpora* de aprendizes.

O quarto capítulo contém uma descrição detalhada da metodologia adotada no presente trabalho. Trato, também, dos três *corpora* utilizados (o BAWE, o LOCNESS e o BR-ICLE) e dos processos metodológicos que pretendo seguir para

obter os dados para análise. Descrevo também as ferramentas de linguística de *corpus* adotadas neste trabalho, além de outras ferramentas e softwares.

Já o quinto capítulo consiste na descrição dos resultados obtidos, além de uma discussão desses resultados.

Por fim, o sexto e último capítulo contém sugestões de atividades didáticas voltadas para o SN e suas complexidades e os resultados encontrados no presente estudo.

O SINTAGMA NOMINAL

Este capítulo traz um levantamento teórico a respeito dos SNs do inglês. Ele se inicia com uma descrição da estrutura do sintagma e suas realizações. Discorrerei sobre as classificações dos elementos que constituem o sintagma, ou seja, o núcleo, os determinantes e os modificadores. Em seguida, a complexidade dos SNs será discutida junto com a descrição das pesquisas feitas dentro desse tema.

São utilizadas, neste capítulo, gramáticas descritivas da língua inglesa, reconhecidas no meio acadêmico. Essas gramáticas possuem um viés funcionalista, trazendo exemplos da língua em uso, sendo baseadas em *corpora* de referência. Antes de descrever os SNs, uma breve descrição dos elementos que compõem uma oração na língua inglesa será apresentada. Esses elementos serão descritos brevemente, já que grande parte dos elementos que compõem uma oração são exercidos por sintagmas nominais.

2.1 Elementos da oração na língua inglesa

Os sintagmas nominais são um tipo de unidade linguística, dentre outros tipos de unidades de uma língua (DOWNING, 2015), conforme exemplo (1).

(1) *Television violence is prevalent in the United States.* (LOCNESS)

No exemplo 1, é possível observar uma unidade linguística (oração) que possui, como elementos, várias unidades menores, como:

- sintagmas: [*Television violence*] [*is*] [*prevalent*] [*in the United States*]
- palavras: [*Television*] [*violence*] [*is*] [*prevalent*] [*in*] [*the*] [*United*] [*States*]
- morfemas: [STATE] + [plural -s] = *States* (em que temos 2 morfemas; a raiz *state* e o morfema de plural -s).

Em inglês, há duas palavras correspondentes para sintagma: *phrase* e *group*. Halliday e Matthysen (2014) definem *group* como uma expansão de uma palavra, enquanto *phrase* é definida como a contração de uma oração. Ao comparar *group* e *phrase* “começando por lados opostos, as duas alcançam aproximadamente o mesmo status na escala de classificação, como unidades que estão em algum lugar entre a classe da oração e a da palavra”¹ (HALLIDAY, MATHYSSEN, 2014, p. 362-363). Enquanto alguns autores, como Halliday e Matthysen (2014) e Downing (2015) tratam o SN como *noun group*, outras gramáticas, como a de Carter e McCarthy (2015), tratam-no como *noun phrase*. Apesar dessa diferença de nomenclatura, as gramáticas se referem à mesma unidade linguística que, no presente estudo, será referida pelo termo em português, sintagma nominal.

Cada unidade linguística, não sendo uma oração independente, exerce uma função, seja em uma oração, seja em um grupo ao qual ela pertence. Ao observarmos o exemplo (1), notamos que a oração pode ser dividida, primeiramente, em 4 unidades menores, ou seja, há 4 funções sintáticas preenchidas:

- [*Television violence*] é um sintagma nominal (SN) com dois elementos, que tem função de sujeito da oração. Esse sintagma possui, internamente, duas unidades linguísticas, que são o substantivo *violence*, que tem função de núcleo do sintagma, e possui outro SN, *television*, que tem função de modificador do núcleo.
- [*is*] é um sintagma verbal (SV), que tem função de predador da oração.
- [*prevalent*] é um sintagma adjetivo (Sadj) com apenas um elemento. Esse sintagma tem função de predicativo do sujeito da oração.

1 Todas as traduções do trabalho são de minha autoria. No original: “starting from opposite ends, the two achieve roughly the same status on the rank scale, as units that lie somewhere between the rank of a clause and that of a word”.

- [*in the United States*] é um sintagma preposicionado (SPrep), que contém o núcleo *in* e o objeto da preposição, um SN, *the United States*.

Conforme observado anteriormente, uma oração possui algumas funções sintáticas que são realizadas por grupos diferentes. A seguir, descreverei os elementos da oração na língua inglesa.

Uma oração independente, na língua inglesa, é composta por dois constituintes, o sujeito e o predicado (DOWNING, 2015). O predicado pode conter, além do verbo, outros elementos, que são os objetos e os complementos. Os objetos são classificados como direto (Od), indireto (Oi) e preposicionado (Oprep). Apenas o objeto direto tem o termo exatamente correspondente em português, sendo o objeto na presença de verbos transitivos. O objeto indireto é o objeto com papel semântico beneficente ou receptor, e ocorre na presença de um Od. Já o Oprep é o objeto que ocorre com preposição. Os segundos são classificados, no inglês, como complemento do sujeito (Cs), conhecido no português como predicativo do sujeito; complemento do objeto (Co), conhecido no português como predicativo do objeto; ou complemento locativo (Cloc). As funções da oração do inglês mencionadas podem ser vistas nos exemplos a seguir:

(2) *Stagnation also hit **the railway systems of the US and Germany**.* (BAWE)

(3) *It gave **them** a small amount of independence.* (BAWE)

(4) *You must look **after him** devotedly.* (BAWE)

(5) *Women are still **marginal figures in contemporary sport**.* (BAWE)

(6) *He doesn't find Brazil **abhorrent**.* (BAWE)

(7) *When he moved **to Milan**, however, he was not understood.* (BAWE)

No exemplo (2), o objeto direto de *hit* é o SN *the railway systems of the US and Germany*, que é o segundo participante do evento da oração, além do sujeito *stagnation*. Já em (3), temos *them* como o objeto indireto de *gave*, enquanto *a small amount of independence* é o objeto direto. Desta forma, temos três participantes no evento expresso pela oração. Note que, na língua inglesa, o objeto indireto é,

semanticamente, o receptor ou o beneficiário. Para o objeto preposicionado, temos o exemplo (4), em que *after him* é objeto preposicionado de *look*.

Em relação aos complementos, em (5) o predicativo do sujeito *marginal figures in contemporary sport* dá qualidade ao sujeito *women*, enquanto em (6) o predicativo do objeto *abhorrent* dá qualidade ao objeto direto *Brazil*. Já no exemplo (7), *to Milan* é o complemento locativo de *moved*. O complemento locativo (Cloc) é um complemento necessário para alguns verbos que expressam movimento ou direção, como *come* e *drive*. Além desses verbos, de acordo com Downing (2015), o Cloc também é um complemento em orações com verbos que indicam estar em uma localização, como *live* e *stay*. Diferentemente dos objetos, os complementos não são participantes semânticos do evento da oração, embora sejam elementos essenciais tanto sintática quanto semanticamente (DOWNING, 2015).

Além dos elementos da oração descritos, temos o adjunto. Segundo Downing (2015), os adjuntos circunstanciais são os mais frequentes e geralmente expressam modo, tempo, propósito, local, dentre outras circunstâncias, conforme exemplo (8).

(8) *At this point, active assistive motion should be employed.* (BAWE)

Em (8), o SPrep *at this point* é o Adjunto da oração *active assistive motion should be employed*.

Como o foco deste trabalho são os SNs, discutirei a seguir os elementos que compõem o SN, além de apresentar as funções que os SNs podem exercer em uma oração ou em outras unidades linguísticas.

2.2 Os elementos do Sintagma Nominal do inglês

Os elementos que compõem o SN podem ser descritos de duas formas diferentes, dependendo do autor que o descreve. A primeira descrição consiste em dividir os elementos em três: núcleo, pré- e pós-modificadores (HALLIDAY, MATTHIESSEN, 2014), enquanto a segunda divide o SN em quatro elementos: o núcleo, o determinante, pré- e pós-modificadores (HUDDLESTON, PLUM, 2005; CARTER, MCCARTHY, 2015; DOWNING, 2015).

Para ilustrar essa diferença, utilizarei o exemplo (9), que apresenta um SN encontrado em Halliday e Matthiessen (2014, p. 388).

(9) *those two splendid old electric trains with pantographs.*

Esse sintagma seria descrito da seguinte forma, na primeira descrição:

Classificação dos elementos	pré-modificadores	núcleo	pós-modificadores
Elementos do SN	<i>those two splendid old electric</i>	<i>trains</i>	<i>with pantographs</i>

Já na segunda, o sintagma teria a seguinte divisão de elementos:

Classificação dos elementos	determinantes	pré-modificadores	núcleo	pós-modificadores
Elementos do SN	<i>those two</i>	<i>splendid old electric</i>	<i>trains</i>	<i>with pantographs</i>

Observe que, na primeira descrição (HALLIDAY, MATTHIESSEN, 2014), todos os elementos que antecedem o núcleo são considerados pré-modificadores. Já na segunda classificação (CARTER, McCARTHY, 2015; DOWNING, 2015), os dois primeiros elementos (*those* e *two*) são classificados como determinantes, e os três elementos seguintes (*splendid*, *old* e *electric*) são classificados como pré-modificadores. O presente estudo adotou a segunda descrição dos elementos que compõem o SN, ou seja, considera que esse sintagma é composto por determinante, núcleo, pré- e pós-modificadores. Essa descrição foi adotada porque os determinantes têm função diferente da dos pré-modificadores. A seguir, cada um dos elementos que compõem o SN será descrito.

2.2.1 O núcleo do sintagma nominal

Os nomes, ou substantivos², são o núcleo prototípico de um SN. Os substantivos são definidos como uma categoria gramatical que denota todos os

² Utilizarei os termos “nomes” ou “substantivos” para me referir à classe de palavra de substantivos indistintamente.

objetos físicos, substâncias e elementos abstratos. Segundo Downing (2015), os sintagmas nominais se referem aos aspectos do mundo, que, segundo os autores, nós percebemos como entidades. Essas entidades possuem aspectos tanto concretos quanto abstratos. Além dos substantivos, outros tipos de palavras também exercem a função de núcleo de SN, algumas com mais frequência que outras. A outra classe de palavras que se realiza tipicamente como núcleo de SN são os pronomes. Por fim, temos os nomes verbais, que também se realizam frequentemente como núcleo de SN. Menos frequentemente, temos os adjetivos como núcleos de SN. A seguir, será descrita cada uma dessas classes de palavras que ocorrem como núcleo de SN mais frequentemente.

2.2.1.1 Os substantivos

Os substantivos, conforme mencionado anteriormente, são o núcleo prototípico de um SN. Eles são a maior classe de palavras, e não são, geralmente, identificáveis em sua forma (CARTER, McCARTHY, 2015). Apesar disso, há alguns sufixos que geralmente ocorrem em substantivos, como *-ment*, *-tion*, *-age*, *-ness*, dentre outros. Um mesmo sufixo, contudo, pode indicar mais de uma classe de palavras. Esse é o caso do sufixo *-ing*, que pode indicar tanto substantivos (gerúndio) quanto verbos (particípio presente); e do sufixo *-y*, que pode indicar tanto substantivos, como em *discovery*, como adjetivos, como em *windy*.

Os substantivos têm uma forma singular e uma plural. No plural, são geralmente flexionados no final da palavra por *-s*. No entanto, alguns substantivos possuem forma irregular de plural, como é o caso de *person*, cujo plural é *people*. Além disso, os substantivos no inglês, em geral, não possuem gênero masculino ou feminino, como ocorre em línguas como o português, na qual se encontra essa distinção de gênero marcada por um sufixo (como em *porco* e *porca*) (CARTER, McCARTHY, 2015).

É possível reconhecer os substantivos em uma estrutura através da observação de algumas características sintáticas (CARTER, McCARTHY, 2007). Essas características são: pode ser precedido por um determinante, pode ser modificado por adjetivos ou outro substantivo. E quando o substantivo é núcleo de

um sintagma nominal com função de sujeito, o verbo concorda com o substantivo, indicando singular ou plural, e pessoa.

Os substantivos possuem diferentes classificações que se complementam. Inicialmente, eles podem ser classificados como próprio ou comum. De acordo com Greenbaum e Quirk (2004), os nomes próprios dão uma designação ou identificação a pessoas, lugares e instituições. Além disso, segundo esses autores, os nomes próprios podem conter uma, duas e até mais palavras, lembrando que essa definição também se estende para os nomes que designam tempo, estações e comemorações, como os dias da semana, como *Sunday* e *Monday*; os nomes dos meses, como *January* e *February*; e as datas comemorativas, como *Easter* e *Christmas*.

Os nomes próprios são, geralmente, escritos com letra maiúscula, embora haja nomes próprios escritos com letras minúsculas, como *brie*, o nome de um queijo (CARTER, McCARTHY, 2015, p. 351). Por se tratar de um nome que possui um referente único no mundo, os nomes próprios podem possuir comportamento diferente dos substantivos comuns. Primeiramente, eles geralmente não aceitam determinantes ou contrastes com números, como em (*) *the Brazil* ou em (*) *some Belo Horizontes*. Além disso, os nomes próprios que possuem muitos elementos são tratados como uma unidade, não ocorrendo variações dentro dessa unidade. Um exemplo seria o nome do jornal *The New York Times*, um nome próprio que possui 4 palavras. Greenbaum e Quirk (2004) argumentam que o exemplo (10) seria tratado como agramatical:

(10) **The Newer York Times* (GREENBAUM, QUIRK, 2004, p. 87)

O exemplo (10) mostra o nome próprio do jornal *The New York Times* com a palavra *Newer*, que é uma variação de *New*, tornando esse nome próprio, segundo o autor, agramatical. Contudo, dentro de contextos específicos, é possível encontrar variações de substantivos próprios. Nesses contextos, a variação pode cumprir um papel específico; por exemplo, um papel cômico.

Figura 1: Nome próprio com variação na estrutura



Fonte: <https://theinfosphere.org/File:The_New_New_York_Times_7ACV18_1.png>

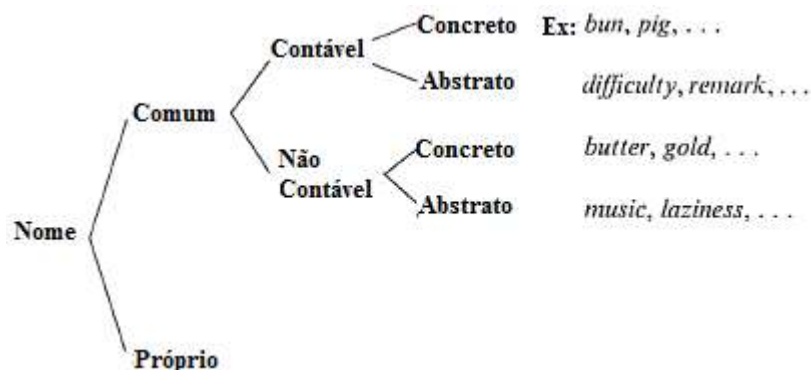
A figura 1 mostra uma variação do nome próprio “The New York Times” com um modificador *new* após o determinante *the*. Esse uso ocorreu no contexto de um desenho animado chamado “Futurama”, indicando um novo jornal no futuro.

Além disso, Leech e Svartvik (2002) atentam para o fato de o substantivo próprio não ser completamente invariável. Esses autores lembram que nomes próprios podem receber artigos definidos. Para eles, isso ocorre quando há mais de uma entidade com o mesmo nome, por exemplo. Com isso, o nome perde sua propriedade de nome próprio (LEECH, SVARTVIK, 2002), por não ter mais um referente único. O mesmo pode ocorrer em relação à possibilidade de o nome próprio aparecer na forma plural (CARTER, McCARTHY, 2015), visto que há substantivos próprios que são contáveis, como as estações do ano, no exemplo (11)

(11) (...) *adapted for life in harsh, snowy **winters***. (BAWE)

Já os nomes comuns possuem dois tipos de classificação. A primeira classificação, que tem maior importância semântica, é em relação a um nome ser concreto ou abstrato. A segunda, que tem importância tanto semântica quanto sintática, é em relação à possibilidade de serem contados, sendo classificados como nomes contáveis ou não contáveis (GREENBAUM, QUIRK, 2004; CARTER, McCARTHY, 2015; DOWNING, 2015).

Figura 2: Os tipos de substantivos



Fonte: Greenbaum e Quirk (2004, p. 70).

Essas duas classificações, conforme figura 2, não são divergentes entre as gramáticas consultadas (HALLIDAY, MATTHIESSEN, 2004; CARTER, McCARTHY, 2015; DOWNING, 2015).

Segundo Downing e Locke (2006), os nomes concretos se referem a entidades como:

- pessoas, como no exemplo (12),
(12) ***students** in the public schools are not properly disciplined.* (LOCNESS)
- objetos, como no exemplo (13),
(13) *the person with the **gun** really doesn't want to kill the other.* (LOCNESS)
- lugares, como no exemplo (14),
(14) *These are now available in a **supermarket** near you.* (LOCNESS)
- e instituições, como no exemplo (15),
(15) *the suing of a tobacco **company**.* (LOCNESS)

E as entidades abstratas podem referir-se a:

- nomes de ações, como no exemplo (16),
(16) *many colleges are experiencing major problems with underage **drinking**.* (LOCNESS)
- abstrações, como no exemplo (17),
(17) *they begin to seek a new **truth** where the problems with the old truth began.* (LOCNESS)
- fenômenos ou processos, como no exemplo (18),
(18) ***divorce** has become a growing problem in the second half of the twentieth century.* (LOCNESS)
- qualidades, como no exemplo (19),
(19) *(...) has in turn led to many of today's problems with **immigrants**.* (LOCNESS)

- emoções, como no exemplo (20),

(20) *he has problems with **love** and with **deceit**.* (LOCNESS)

Em relação aos nomes contáveis e não contáveis, Greenbaum e Quirk (2004) lembram que eles possuem regras e restrições de uso particulares. Essas restrições se aplicam desde as possibilidades de o nome ser expresso no plural até o tipo de determinantes que podem ocorrer no SN. Como o que difere esses dois tipos de substantivo é se referirem a entidades que têm ou não a possibilidade de serem contadas, as diferenças gramaticais são encontradas no uso dos determinantes que acompanham essas palavras.

2.2.1.2 Os pronomes

Conforme mencionado anteriormente, os pronomes também realizam o núcleo de um SN. Apesar de possuírem características semelhantes às dos substantivos (eles desempenham as mesmas funções que um nome), eles raramente são modificados (CARTER, McCARTHY, 2015; LEECH, SVARTVIK, 2002). Os pronomes que recebem algum tipo de modificação mais frequentemente são os pronomes indefinidos, e os pronomes *one* e *those* (BIBER et al., 1999).

Os pronomes que realizam núcleo de SN são os pronomes pessoais, impessoais, possessivos, demonstrativos, interrogativos e reflexivos. Os pronomes pessoais (*I, you, he, she, it, we, they*) são, tipicamente, elementos dêiticos em uma oração. Eles se referem a pessoas, coisas e até mesmo ideias anteriormente mencionadas, como no exemplo (21), que possui o pronome *they*, que faz referência a *Americans*.

(21) *Unlike the citizens of foreign democracies in which the government takes responsibility for voting, **Americans** must prepare for the eventual vote by taking the initiative to register before the election at a time in the political cycle, when political information and interest are at a low point. Permanent registration does not come automatically as an attribute of citizenship; **they** vote according to their place of residence.* (BAWE)

Essa função dêitica também é desempenhada pelos pronomes demonstrativos (*this, that, these, those*), como no exemplo (22) a seguir.

(22) *For example, say you see someone beating up an elderly lady. You see **this** happening and take in this primary information that is external from yourself and analyse it in your brain.* (BAWE)

No exemplo (22), temos o pronome demonstrativo *this*, que se refere ao evento anteriormente mencionado: *you see someone beating up an elderly lady*.

Há pronomes que podem ter um uso gramatical não dêitico. Esse é o caso do pronome *it*. Como a posição de sujeito sempre deve ser preenchida em orações finitas na língua inglesa (exceto imperativo), o pronome *it* desempenha essa função quando uma oração não possui sujeito, conforme exemplo (23).

(23) *the window curtains were closed because **it** was raining.* (BAWE)

No exemplo 23, *it* é semanticamente vazio, usado de forma não referencial apenas para preencher a posição de sujeito sintático da oração.

O pronome impessoal também pode desempenhar um papel não dêitico, como é o caso do pronome *one*, no exemplo (24).

(24) **One** thinks about life a lot more as time goes by. (DOWNING, 2015, p. 370).

No exemplo (24), o pronome *one* é usado para fazer declarações mais gerais. De acordo com Downing (2015), esse recurso é encontrado, às vezes, em estilos formais, para se evitar o uso do pronome pessoal de primeira pessoa.

Já os pronomes reflexivos (*myself, yourself, himself, herself, itself, ourselves, yourselves, themselves*) podem desempenhar funções diferentes (DOWNING, 2015). Primeiro, eles são utilizados em co-referência com o sujeito, como exemplificado em (25).

(25) *the peasants saw **themselves** as rightful owners of the land.* (BAWE)

Observe que o SN *the peasants* é o sujeito da oração, e o SN *themselves* é o Od. Nesse exemplo, o pronome reflexivo faz referência ao sujeito *the peasants*. Segundo, os pronomes reflexivos também podem ser usados para se dar ênfase, sendo utilizados antes ou depois do sujeito, conforme exemplificado em (26):

(26) *Or, as Hegel **himself** states (...)* (BAWE)

Os outros tipos de pronomes que se realizam como núcleo de SN são os interrogativos (*who, whose, which, what*) e os indefinidos *some, any, no* e *every* (*somebody, anyone, everything*, dentre outros), conforme exemplos (27) e (28), a seguir:

(27) **Who** can know for certain that certain knowledge is impossible? (BAWE)

(28) if **someone** claims a breach of his or her rights society starts a witch-hunt amidst a blind panic, in order to blame this so-called crime on anything. (BAWE)

Em (27) o pronome interrogativo é o único elemento de um SN que tem a função de sujeito da oração. Em (28), o pronome *someone*, em negrito, é também o único elemento do SN com função de sujeito da oração subordinada.

2.2.1.3 Os nomes verbais

Os nomes verbais são outro tipo de núcleo de SN e consistem em verbos na forma de gerúndio (terminado em *-ing*). Carter e McCarthy (2015) dizem haver a nominalização dos verbos, quando nessa forma. De acordo com os autores, *V-ing* (verbo com sufixo *-ing*) é muito utilizado no estilo acadêmico, como uma forma de se criarem SNs com uma grande quantidade de informação, uma alternativa para não se ter de escrever orações muito longas. No exemplo (29), vemos um SN (sublinhado) cujo núcleo (em negrito) é um *V-ing*:

(29) Its **dismantling** over the last twenty years has much to do with efforts in the area of conceptual history. (CARTER, McCARTHY, 2015, p. 271).

As gramáticas pesquisadas, em geral, são muito sucintas ao tratarem desse assunto. De acordo com Carter e McCarthy (2015), o gerúndio é tradicionalmente definido como uma forma verbal que é funcionalmente semelhante ao substantivo, e os nomes verbais se realizam nessa forma. Para eles, os nomes verbais podem ocorrer como o núcleo de um sintagma nominal, desempenhando funções diversas, dentre as quais sujeito, objeto e complemento de SPrep. Nos exemplos (30) e (31), a seguir, temos exemplos de SNs cujos núcleos são nomes verbais, desempenhando as funções de sujeito e complemento de SPrep, respectivamente.

(30) When I was a lad, I had to milk cows by hand. Now all the **milking** is done by machines. (CARTER, McCARTHY, 2015, p. 423).

(31) It was a good play, with some very good **acting**. (CARTER, McCARTHY, 2015, p. 423).

No exemplo (30), *milking* aparece como núcleo do SN, que possui, também, como elementos os determinantes “all” e “the”. Já no exemplo (31), *acting* aparece como complemento da preposição *with*, e é determinado por *some* e pré-modificado

por *very good*. Nesses dois exemplos, é possível notar que o SN nucleado por um nome verbal pode possuir os mesmos elementos de um SN cujo núcleo é um substantivo: determinantes e modificadores.

No presente estudo, não fiz distinção entre SNs com núcleo *-ing* muito frequentes, como *beginning*, em relação aos quais pode-se passar despercebido o fato de serem derivados de verbos, e SNs com núcleo *-ing* menos frequentes, como *craving*.

Além do gerúndio, o particípio presente também possui terminação *-ing*. Essa forma verbal, além de entrar na formação dos tempos verbais progressivos, também pode ter a função de modificador (CARTER, McCARTHY, 2015), como em *swimming lesson*.

Huddleston (1984) observa, de forma clara, a importante distinção entre o particípio presente e o gerúndio:

o principal campo de problema concerne as formas tradicionais não finitas de lexemas verbais, os particípios, o gerúndio e o infinitivo. Um particípio é considerado um “adjetivo verbal”, enquanto o gerúndio e – para algumas gramáticas, pelo menos – o infinitivo são considerados “nomes verbais”. De acordo com a doutrina tradicional, um gerúndio como *writing* em *She likes writing letters* é um nome porque ele é o objeto do verbo *like* [...] Isso [...] leva os gramáticos tradicionais a classificar de uma forma única, como nomes, palavras que são sintaticamente muito diferentes.³ (HUDDLESTON, 1984, p. 99).

Huddleston (1984, p. 99), então, exemplifica essa distinção com 2 exemplos em que o verbo *writing* aparece:

(32) *Actually writing the letters took only a couple of hours.* (HUDDLESTON, 1984, p. 99).

(33) *The actual writing of the letters took only a couple of hours.* (HUDDLESTON, 1984, p. 99).

O autor sugere que, no exemplo (32), *writing* deveria ser classificado como um sintagma verbal, por ser o núcleo do predicado verbal *actually writing the letters*. Já o *writing* do exemplo (33) deveria ser classificado como um nome, já que ele é o

3 No original: “The main problem area concerns the traditional non-finite forms of verb lexemes, participles, the gerund and the infinitive. A participle is said to be a ‘verbal adjective’, while the gerund and – for some grammars at least - the infinitive are ‘verbal nouns’. According to traditional doctrine, a gerund like *writing* in *She likes writing letters* is a noun because it is the object of the verb *like* [...] This [...] leads the traditional grammarians to classify together as nouns words which are syntactically very different”.

núcleo do SN estendido: *the actual writing of the letters*. Além disso, as estruturas das unidades, em ambos os exemplos, apresentam diferenças. Em *actually writing the letters* há uma estrutura que remete a SVO⁴, sendo (*she was*) [*writing*] [*the letter*]. Em *the actual writing of the letters*, temos uma estrutura semelhante à de um SN, na qual encontramos det [*the*] pré-mod [*actual*] núcleo [*writing*] pós-mod [*of the letter*]. Outra característica observada por Huddleston (1984) é em relação ao tipo de modificação encontrada nos sintagmas dos exemplos (32) e (33). Enquanto o primeiro *writing* é modificado por um advérbio (característica típica de verbos), o segundo *writing* é modificado por um adjetivo (característica típica de nomes).

Em suma, os nomes verbais são palavras derivadas de verbos, classificados como gerúndio. Eles se realizam como núcleo de SN e possuem os mesmos elementos que um SN cujo núcleo é um substantivo (determinantes e modificadores). Além disso, o SN cujo núcleo é um nome verbal pode realizar as mesmas funções de um SN, dentre eles a função de sujeito, objeto e complemento de uma oração.

2.2.2 Determinantes do sintagma nominal

Os determinantes são elementos que ocorrem apenas em SNs. Eles “identificam o referente de um grupo nominal dizendo-nos qual, ou o quê, ou de quem ele é, quanto, quantos, a qual parte ou nível dele nós estamos nos referindo, o quão grande ou frequente ele é, como ele é distribuído no espaço ou tempo”⁵ (DOWNING, 2015, p. 382). Ou seja, os determinantes particularizam e selecionam o referente do SN dentro do discurso do falante. Há um grande número de determinantes e eles são sub-categorizados. De acordo com Downing (2015), os determinantes dividem-se em grupos que têm função de definir ou particularizar, quantificar e distribuir, numerar e ordenar. Além dos determinantes, há os semi-determinantes, que possuem função um pouco diferente. A seguir, temos um quadro que ilustra os determinantes de acordo com sua função.

4 SVO = Sujeito Verbo Objeto.

5 No original: “identify a nominal group referent by telling us which or what or whose it is, how much, how many, what part or degree of it we are referring to, how big or frequent it is, how it is distributed in space or time”.

Quadro 1: Tipos de determinantes

1 Definidores e particularizadores	2 Quantificadores e Distribuidores	3 Numeradores e ordenadores	4 Semi-determinantes
Definido <i>The</i>	Frações <i>half, (a) quarter, etc.</i>	Cardinais <i>one, two, ten, etc.</i>	<i>such, certain, former, latter, same, other, last, next, own</i>
Indefinidos <i>a(n), some, zero (0)</i>	Multiplicadores <i>double, twice, hundreds of, etc.</i>	Ordinais <i>first, second, third ...</i>	
Demonstrativos <i>this, that, these, those</i>	Não exatos <i>some, any, no, much, (a) little, (a) few, many, etc.</i>		
Possessivos <i>my, your, his, her, their</i> <i>Sam's, mu friend's, etc.</i>	Outros quantificadores <i>a lot of, lots of, plenty of, a great deal of, etc.</i>		
Interrogativos/relativos <i>what, whose, which, whichever</i>	Distributivos <i>all, both, either, neither, each, every, none (of)</i>		
Exclamativo <i>what (a) ...</i>			

Fonte: Downing (2015, p. 390).

Os determinantes definidores e particularizadores abrangem artigos definidos e indefinidos, demonstrativos, possessivos, interrogativos, relativos e exclamativos. Já os determinantes quantificadores e distribuidores abrangem as frações, os multiplicadores, os determinantes não exatos, os distributivos e os quantificadores. Os determinantes numeradores e ordenadores abrangem os números cardinais e ordinais.

Por fim, os semi-determinantes, de acordo com Downing (2015), são elementos que precedem um determinante definido ou indefinido. Embora essas palavras sejam algumas vezes classificadas como adjetivos (com exceção de *such*)

por alguns autores, como Carter e McCarthy (2015), Downing (2015, p. 388) afirma que “elas não descrevem o referente e parecem ter uma função de especificador”.⁶Embora alguns determinantes possuam forma fixa, como é o caso dos definidores e particularizadores, outros determinantes podem variar seus elementos. Este é o caso dos quantificadores e distribuidores, que, em sua maioria, são formados por várias palavras. Um exemplo é o determinante *a lot of*, que possui 3 elementos. Ou seja, há uma infinidade de possibilidades de combinação de elementos dentro da classe dos determinantes.

2.2.3 Os modificadores do sintagma nominal

Os modificadores são aqueles elementos que classificam ou trazem outras informações sobre o referente, diferentes dos tipos de informação trazidos pelos determinantes (DOWNING, 2015). O SN pode receber dois tipos de modificadores: os pré-modificadores, que ocorrem antes do núcleo, e os pós-modificadores, que ocorrem após o núcleo. Cada um desses dois elementos possui função específica e pode ser realizado por um número limitado de unidades.

2.2.3.1 Pré-Modificadores

Os pré-modificadores são classificados em 2 tipos: classificador e epíteto. Halliday e Matthiessen (2014) mostram a posição específica e a função de cada elemento que antecede o núcleo do SN.

O sintagma *those two splendid old electric trains* (HALLIDAY, MATTHIESSEN, 2014, p. 364), mencionado no capítulo 1, exemplo (9), mostra as posições e as funções dos elementos pré-núcleo.

(9) *those two splendid old electric trains*

<i>those</i>	<i>two</i>	<i>splendid</i>	<i>old</i>	<i>electric</i>	<i>trains</i>
elemento dêitico	numeral	epíteto1	epíteto2	classificador	substantivo
determinante	determinante	pré-mod	pré-mod	pré-mod	núcleo

6 No original: “they do not describe the referent and appear to have a specifying function”.

No exemplo, vemos que os dois primeiros elementos são determinantes. Observe que mais de um tipo de determinante e pré-modificador pode ocorrer ao mesmo tempo; nesse caso, temos um determinante dêitico e um numeral, seguido de três pré-modificadores: dois epítetos e um classificador.

O elemento epíteto dá alguma qualidade ao elemento principal. De acordo com Halliday e Matthiessen (2004), essa qualidade pode ser de dois tipos: alguma propriedade do referente (experencial) ou uma opinião do falante a respeito desse referente (interpessoal). O primeiro é realizado através de palavras como *new* ou *old*, e o segundo é realizado através de adjetivos como *splendid* e *beautiful*. Os epítetos interpessoais, segundo os autores, tendem a preceder os experienciais, como no exemplo (34).

(34) *splendid old train* (HALLIDAY, MATTHIESSEN, 2014, p. 364).

No exemplo (34), *splendid* é considerado interpessoal, enquanto *old* é considerado experencial, uma propriedade de *train*. Embora os epítetos apareçam após os determinantes, é possível encontrar SNs em que algum deles ocorra em uma posição pós-dêitica (após um determinante dêitico e antes de um determinante quantificador). Isso ocorre principalmente com adjetivos de qualidade experencial, como no exemplo (35).

(35) *the famous three musketeers*. (em oposição a *the three famous musketeers*). (HALLIDAY, MATTHIESSEN, 2014, p. 364).

Já o elemento classificador, segundo Halliday e Matthiessen (2004), indica uma subclasse do referente. No exemplo (9), o adjetivo *electric* classifica o tipo de trem. Uma palavra como *book* pode ser classificada como *math book*, *history book* ou *geography book*. Halliday e Matthiessen (2004) lembram que a ordem em que o epíteto e o classificador ocupam pode ser invertida, porém o significado também é alterado. Além disso, algumas expressões podem funcionar tanto como epítetos quanto como classificadores, porém com significados distintos. Assim, um adjetivo pode não só classificar um referente, mas também trazer uma característica dele, como em (36)

(36) *fast train* (HALLIDAY, MATTHIESSEN, 2014, p. 377).

Em (36), *fast* pode significar tanto um trem que corre rápido (epíteto) quanto um trem classificado como expresso (classificador). Contudo, os autores salientam que um epíteto admite gradação (ex. *faster train*), enquanto o mesmo não ocorre com um classificador, como no exemplo (37)

(37) *passenger train* (HALLIDAY, MATTHIESSEN, 2014, p. 377).

Em (37), o adjetivo modificador *passenger* não poderia ser **more passenger train*

As outras gramáticas consultadas (DOWNING, 2015; CARTER, McCARTHY, 2015) para este estudo partilham das distinções dos pré-modificadores descritas por Halliday e Matthiessen (2004).

São várias as unidades linguísticas que podem exercer a função de pré-modificador de SN (DOWNING, 2015). Embora alguns mais frequentes que outros, os pré-modificadores podem ser realizados por (em negrito):

- adjetivo (Adj), como no exemplo (38)

(38) *Also, by making **tall** buildings practical, the elevator 'revealed the need (...)* (BAWE)

- particípio passado, como no exemplo (39)

(39) *They found well **established** business houses.* (BAWE)

- particípio presente, como no exemplo (40)

(40) *The united opposition was such an **unconvincing** union of the former enemies.* (BAWE)

- substantivo, como no exemplo (41)

(41) *They found well established **business** houses.* (BAWE)

- sintagma nominal, como no exemplo (42)

(42) *At **low income** levels, there is a positive correlation between GNI per capita and life expectancy.* (BAWE)

- Advérbio (Adv), como no exemplo (43)

(43) (...) who created a mass hysteria thereby frightening the **then** government led by Rajiv Gandhi. (BAWE)

- orações coordenadas, como no exemplo (44)

(44) *If no agreement is reached the proposal goes back to the Council, which can make a "take-it-or-leave-it" offer.* (BAWE)

Embora todas essas unidades possam ocorrer como pré-modificadores de SN, as unidades que exercem essa função mais frequentemente são adjetivos, participios e substantivos.

Vale distinguir os substantivos modificadores dos substantivos compostos. Tanto Downing (2015) quanto Carter e McCarthy (2015) mencionam que os elementos de um substantivo composto são fortemente conectados, e que esses substantivos compostos são melhor considerados como um único núcleo no SN. Os autores também lembram que, em grande parte desses substantivos compostos, os dois elementos que o compõem são escritos juntos, como *postman* e *bathroom*. Para os substantivos compostos que possuem dois elementos nominais, no entanto, ambos os autores afirmam que não fica tão claro se se trata de um substantivo composto ou de um SN com um substantivo como modificador. Nesses casos, Carter e McCarthy (2015, p. 320) argumentam que a entonação ajuda na distinção, afirmando que

o limite entre um substantivo composto e um sintagma nominal agindo como um pré-modificador de um núcleo nominal não é sempre claro. No entanto, o padrão da ênfase preferido para os compostos, com ênfase no primeiro item, é usualmente um indicador que os substantivos são considerados como uma unidade institucionalizada.⁷

Por outro lado, Downing (2015) argumenta que essa distinção por ênfase nem sempre funciona. Segundo o autor, “padronização por ênfase nem sempre é confiável”⁸ (DOWNING, 2015, p. 397). Os autores que utilizam essa padronização

7 No original: “The botherline between compound nouns and noun phrases acting as premodifiers of noun heads is not always clear. However, the preferred stress pattern for compounds, with stress on the first item, is usually an indication that the nouns are considered as an 'institutionalized' unit.

8 No original: “stress-patterning is not always reliable”.

por ênfase, segundo Downing (2015), afirmam que, nos compostos, o primeiro substantivo é tônico. No entanto, “muitos compostos não seguem esse padrão (*cotton wool*, *zebra crossing*), enquanto alguns classificadores seguem (*stream vehicles*, *rose-bush*)”⁹ (DOWNING, 2015, p. 397). Além disso, Downing e Locke (2006) argumentam que, geralmente,

quando a combinação é escrita como palavras separadas, é provável que seja um substantivo com um substantivo modificador (*head waiter*); se escrito como uma única palavra, é mais provável que seja um substantivo composto (*headache*). O uso do hífen sinaliza aqueles elementos que formam um composto (*walkie-talkie*), e que, caso contrário, seriam considerados pré-modificadores separados¹⁰ (DOWNING, LOCKE, 2006, p. 396).

Tendo em vista essa regularidade observada por Downing e Locke (2006), o presente estudo considera substantivos modificadores todos aqueles nomes que ocorrem ao lado de um núcleo nominal (desde que não seja acompanhado de hífen). E considera um substantivo composto aquele que ocorre em uma única palavra.

2.2.3.2 Pós-Modificadores

Os pós-modificadores trazem informações que auxiliam os participantes a identificarem o referente do sintagma nominal e ocupam a posição pós-núcleo em um SN (DOWNING, 2015). Essa função, geralmente, é realizada por uma série de unidades, geralmente sintagmas e orações. Dentre os sintagmas que exercem função de pós-modificadores, temos:

- sintagmas preposicionados (SPrep), como no exemplo (45),

(45) *All these multiethnic culinary sites allowed people **of all cultures and classes** to work together.* (BAWE)

- os sintagmas adjetivais (SAdj), como no exemplo (46),

9 No original: “many compounds do not follow this pattern (*cotton wool*, *zebra crossing*), while some classifiers do (*steam vehicles*, *rose-bush*)”.

10 No original: “when the combination is written as separate words, it is likely to be a noun with a noun modifier (*head waiter*); if written as a single word it is more likely to be a noun compound (*headache*, *headrest*). Hyphenation signals those elements which form a compound (*walkie-talkie*) and which otherwise would appear to be separate pre-modifiers”.

(46) *Other people migrated to London for the educational opportunities **available**.* (BAWE)

- e os SNs apostos, como no exemplo (47),

(47) *Firstly, although 'Camelot', **the National lottery organiser**, takes a profit from the money generated by its funds.* (LOCNESS)

Dentre as orações que têm função de pós-modificador, temos:

- as orações relativas finitas, como no exemplo (48),

(48) *The aim was to find a person **who could fit the job perfectly**.* (BAWE)

- e as orações não finitas, como no exemplo (49),

(49) *the Pleasure Beach is visited by millions of people **seeking escapism from their everyday lives**.* (BAWE)

Além desses dois tipos de orações, temos também outros elementos que podem exercer a função de pós-modificador, tais como:

- os pronomes reflexivos, como no exemplo (50),

(50) *Jackson **himself** has been convinced by this argument and now endorses a form of physicalism based on this response.* (BAWE)

- e os advérbios, como no exemplo (51)

(51) *This may indicate to some degree what the local people **here** thought of the roman influence.* (BAWE)

O tipo de pós-modificador mais comum na prosa acadêmica, de acordo com Biber et al. (1999), é o SPrep, representando mais de 60% dos casos de pós-modificação de SN.

Outro tipo de pós-modificador de SN frequente são as orações, que são dependentes e se dividem em dois tipos: finitas e não finitas (DOWNING, 2015). Na língua inglesa, o tempo carregado pelo verbo pode ser presente ou passado. O elemento verbal que carrega tempo, número e pessoa pode ser o verbo lexical, auxiliar ou modal.

(52) *Continuous works on their houses would be rather problematic for the peasants who **occupied** them.* (BAWE)

No exemplo (52), *occupied* é o elemento que carrega tempo. Ou seja, os elementos que definem o tipo de oração são o sintagma verbal e a presença ou não de um sujeito sintático. As orações finitas são aquelas em que o verbo expressa tempo, e as não finitas são aquelas em que o verbo não expressa tempo (CARTER, McCARTHY, 2015). Além disso, os verbos das orações finitas possuem

concordância de número e pessoa com o sujeito da oração (GREENBAUM, QUIRK, 2004), conforme exemplo (52). No referido exemplo, *occupied* participa como predicador de uma oração, carregando tempo, número e pessoa (*occupied* = passado, plural e terceira pessoa, neste caso). Por outro lado, as orações não finitas geralmente não possuem sujeito, e não possuem essas características, conforme exemplo a seguir:

(53) *Thus, a community **taking its legislative task to the extreme**, might indeed force its citizens into abide by the law.* (BAWE)

No exemplo (53), *taking its legislative task to the extreme* é uma oração não finita, porque o verbo *standing* não carrega tempo ou modo, número ou pessoa. As orações não finitas com função de pós-modificadores podem ser realizadas por várias formas verbais, sendo elas:

- infinitivo com a partícula *to*, como no exemplo (54),

(54) *Who are the people **to decide the general will**?* (BAWE)

- particípio passado, como no exemplo (55),

(55) *the introduction of entire channels **dedicated to them**.* (BAWE)

- e particípio presente, como no exemplo (56),

(56) Particípio presente: *People **living on low incomes** are prone to consume less healthy foods.* (BAWE)

No exemplo (54), a oração não finita *to decide the general will* modifica o substantivo *people*, que é núcleo do SN. Vale ressaltar que o pronome relativo aparece em orações finitas, e que nessas orações ele pode ser omitido quando tiver a função de objeto na própria oração relativa. Nos exemplos (55) e (56), as orações não finitas *dedicated to them* e *living on low incomes* pós-modificam os núcleos *channels* e *people*, respectivamente.

Os pós-modificadores realizados por orações relativas são classificados como restritivos e não restritivos (DOWNING, 2015). O que diferencia essas duas estruturas, graficamente, é a pontuação. O primeiro, que tem a função de restringir o significado de um referente, geralmente é usado sem vírgulas, enquanto o segundo, que tem a função de dar uma informação adicional sobre um referente, é utilizado

entre vírgulas. Segundo os autores, muitas vezes a vírgula, ou a prosódia, não são suficientes para se diferenciar esses dois tipos de pós-modificadores.

Os modificadores restritivos são o que chamamos de encaixados (*embedded*), já que eles são parte integrante do sintagma. O que caracteriza um pós-modificador restritivo são alguns pontos apontados por Downing e Locke (2006), que abrangem vários níveis, desde sintáticos a semânticos.

Sintaticamente, ele é agregado à estrutura matriz do SN. Prosodicamente, ele compartilha o contorno entonacional de seu antecedente, já que os dois juntos constituem uma unidade informacional: *They admitted the immigrants **who had their papers in order*** (= apenas os imigrantes que tinham seus papéis em ordem). Semanticamente, a relativa restritiva é uma parte integral do significado do SN como um todo. Ela ajuda a estabelecer sobre o quê (ou sobre quem) o falante está falando. Ele seleciona o(s) referente(s) de outros possíveis referentes por alguma propriedade distintiva; nesse caso, apenas os imigrantes que tinham seus papéis em ordem foram admitidos. O SN maior, com sua oração relativa, pode ser expandido por mais uma oração relativa: *The umbrella **we bought¹** that has duck's head handle² made a good present.*¹¹ (DOWNING, LOCKE, 2006, p. 450).

Ou seja, o exemplo (57), dado pelos autores, possui dois modificadores restritivos que pós-modificam *umbrella*.

(57) *the umbrella **we bought that has duck's head handle** made a good present.*
(DOWNING, LOCKE, 2006, p. 450).

O primeiro é a oração relativa finita *we bought* e o segundo é a oração relativa finita *that has duck's head handle*. Assim como ocorre com os pré-modificadores, não há limite para o número de pós-modificadores que podem ocorrer em um único SN.

Conforme discutido nesta seção, os modificadores dos sintagmas nominais possuem comportamentos específicos e o pré- e o pós-modificador são, em geral, exercidos por elementos distintos. A seguir, serão descritas as funções que os sintagmas nominais podem exercer dentro de uma oração.

11 No original: "Syntactically it is embedded in the NG matrix structure. Prosodically, it shares the intonation contour of its antecedent, as the two together constitute one information unit: They admitted the immigrants *who had their papers in order*. (= only the immigrants who had their papers in order). Semantically, the restrictive relative is an integral part of the meaning of the whole nominal group. It helps to establish what (or whom) the speaker is talking about. It picks out the referent(s) from other possible referents by some distinguishing property; in this case, that only the immigrants who had their papers in order were admitted. The larger NG unit with its relative clause can be expanded by a further relative clause: The umbrella ***we bought¹*** *that has a duck's head handle²* made a good present".

2.3 As funções exercidas pelo sintagma nominal

Conforme visto no início deste capítulo, uma oração possui vários elementos e muitos deles são realizados por SNs. Nesta sessão, veremos as funções desempenhadas por SNs.

As funções mais frequentemente desempenhadas por SNs, segundo Downing (2015), são de

- sujeito, como no exemplo (58),
(58) **the people** are obviously involved in these policies. (BAWE)
- objeto direto, como no exemplo (59),
(59) a virus like grey-goo is killing **people**. (BAWE)
- objeto indireto, como no exemplo (60),
(60) I tell **people** I'm a law student. (BAWE)
- predicativo do sujeito, como no exemplo (61),
(61) they are **the people**. (BAWE)
- predicativo do objeto, como no exemplo (62),
(62) we consider ourselves **a tolerant society**. (BAWE)
- e adjunto, como no exemplo (63),
(63) **One day**, someone banks in \$1000 of dollar bills into a local bank. (BAWE)

Além dessas funções, de acordo com Carter e McCarthy (2015), os SNs, frequentemente, também exercem funções dentro de sintagmas, como no caso de complemento de sintagmas preposicionais, no exemplo (64), e determinantes possessivos genitivos de SN, no exemplo (65), conforme exemplos abaixo:

- (64) There was a group of **people present**. (BAWE)
- (65) The **music industry's** current fall in sales. (BAWE)

Em número mais reduzido, encontramos outras funções desempenhadas por SNs, quais sejam:

- pré-modificador de sintagma adjetivo (Sadj), conforme exemplo (66),
(66) The train was **an hour** late. (CARTER, McCARTHY, 2015, p. 319).
- pré-modificador de sintagma adverbial (Sadv), conforme exemplo (67),
(67) pré-modificador de sintagmas adverbiais: **An hour later** they are afraid they have been taken in. (BAWE)

- pré-modificador de sintagma preposicionado, conforme exemplo (68)

(68) pré-modificador de sintagmas preposicionados: *He lives **three houses** along the street.* (CARTER, McCARTHY, 2015, p. 319).

Esses três últimos exemplos mostram que SNs também podem ter função de modificador em outros tipos de sintagmas além do SN.

2.4 Complemento de substantivo

Com uma estrutura sintática parecida com a dos pós-modificadores, os complementos de substantivos são muito comuns na escrita acadêmica (DOWNING, 2015; CARTER, McCARTHY, 2015). Essas unidades, diferentemente dos modificadores, complementam o significado do núcleo do SN. Elas geralmente ocorrem na presença de alguns substantivos específicos, como *fact*, *claim* e *idea*, por exemplo. (CARTER, McCARTHY, 2015).

A unidade linguística que mais frequentemente desempenha essa função de complemento é a oração encabeçada por *that*. Embora menos frequente, outras unidades linguísticas também realizam essa função de complemento do SN. Elas são os sintagmas preposicionados e orações infinitivas com partícula *to* (DOWNING, 2015; CARTER, McCARTHY, 2015), como no exemplo a seguir.

(69) complemento *that*: *The fact **that the majority of young delinquents come from bad-structured families** must not be ignored.* (BRICLE)

Sintagma preposicionado: *a rise **in wages*** (BAWE)

Oração infinita com partícula *to*: *this gain can influence the decision **to have children**.* (BAWE)

No exemplo (69), as estruturas em negrito complementam o significado dos núcleos dos SNs (*fact*, *rise* e *decision*), de acordo com Carter e McCarthy (2015) e Downing (2015).

Como, na superfície, essas estruturas são idênticas aos pós-modificadores, Carter e McCarthy (2015) mostram, a partir de exemplos, a diferença entre modificadores e complementos. Para os autores, um SPrep com função de complemento não pode ser parafraseado com uma oração relativa, como em *the girl in jeans* → *the girl who is wearing jeans*. (CARTER, McCARTHY, 2015, p. 329). Além disso, diferentemente dos modificadores, um complemento *that* não pode ser

substituído por uma oração relativa com *which*. Downing (2015) também atenta para o fato de que os substantivos que são complementados por SPreps geralmente possuem um verbo correspondente.

2.4 Sintagmas nominais complexos

As características da escrita acadêmica podem ser o foco de interesse de estudiosos de várias áreas, seja do ensino, seja da descrição linguística. Observa-se que, com o crescimento de estudos baseados em *corpus*, o avanço tecnológico das ferramentas de pesquisa e o crescente número de *corpora* de referências, houve mais facilidade para o desenvolvimento de estudos que buscam identificar essas características. Um tema, por exemplo, amplamente estudado sobre a língua inglesa e a escrita acadêmica é o da relação entre o gênero acadêmico e a complexidade do SN produzido.

O estudo da complexidade de um texto através dos SNs se baseia na capacidade do SN de aglomerar uma grande quantidade de informações através dos pré- e pós-modificadores. Um SN pode ser estruturalmente simples ou complexo, de acordo com a presença ou não de modificadores. Um SN simples não possui modificadores, como é o caso de SNs nucleados por pronomes e nomes próprios (FANG, SCHLEPPEGRELL, COX, 2006).

A complexidade na escrita tem sido foco de estudo pelo menos desde a década de 1930, de acordo com Biber, Gray e Poonpon (2011). Segundo os autores, esses estudos da década de 1930 eram centrados no desenvolvimento da escrita de alunos do ensino básico, e também voltados para como a escrita desses mesmos alunos se tornava mais complexa com o passar dos anos escolares.

Nas décadas de 1970 e 1980, os estudos começaram a abranger, também, a produção de falantes de inglês como segunda língua, ao estudar a complexidade sintática da escrita acadêmica e de textos de outros gêneros. Hiatt (1978), por exemplo, analisou a complexidade da escrita em prosa de autores masculinos e femininos, utilizando a medida do tamanho das orações. Os resultados de seu estudo mostraram que o tamanho médio das orações (por número de palavras)

escritas em prosa por escritores masculinos não difere estatisticamente do tamanho médio das orações escritas em prosa por escritoras femininas.

No âmbito acadêmico, Witte e Faigley (1981) pesquisaram complexidade da escrita em ensaios. Os autores analisaram ensaios acadêmicos escritos por universitários na Universidade do Texas, comparando a complexidade sintática a partir da T-Unit dos ensaios com melhores e piores notas. Essa medida se baseia na quantidade de orações, variação entre tamanho da unidade (oração), quantidade de orações subordinadas e coordenadas, dentre outras medidas (LU, 2011). Os resultados desse estudo de Witte e Faingley (1981) mostraram que os ensaios com melhores notas possuíam um tamanho médio de T-Unit maior do que o dos ensaios com piores notas. Além disso, os ensaios dos alunos com melhor nota apresentaram em média menos orações por T-Unit. Na análise dos resultados, os autores apontam para o fato de a medida T-Unit não ser capaz de trazer mais informações sobre a complexidade da oração. Ou seja, eles mostram que, apesar de a medida T-Unit trazer um valor a ser interpretado, ele não é o suficiente para indicar quais elementos são usados pelos alunos com melhor nota que tornam sua escrita mais complexa. Os autores, então, concluem, com base em suspeitas, que os alunos com melhores notas escrevem um número maior de informações contidas em orações menores.

Nas décadas seguintes, de acordo com Biber, Poonpon e Gray (2011), a complexidade da escrita de aprendizes de inglês como segunda língua continuou a ser estudada, utilizando-se a T-Unit para investigar fluência, precisão e complexidade da escrita de aprendizes de inglês até o presente momento. Os autores citam pesquisas como as de Ellis e Yuan (2004), Larsen-Freeman (2006) e Nelson & Van Meter (2007). Segundo Biber, Poonpon e Gray (2011)

a grande maioria desses estudos apoiaram-se no construto da T-unit: uma oração principal e todas as orações dependentes associadas. Duas medidas específicas tornaram-se especialmente populares: medida média da T-Unit (MLTU), que depende do tamanho geral em palavras da T-Unit, a partir de uma média de todas as T-Units em um texto, e orações por T-Unit (C/TU), que depende do número de orações dependentes por T-Unit, novamente a partir da média de todas as T-Units em um texto. Por exemplo, a seguinte oração consiste de uma única T-Unit que é relativamente curta (11 palavras), mas inclui 2 orações encaixadas: I don't know [[why I was

expecting [to see something else]]¹² (BIBER, POONPON, GRAY, 2011, p. 7).

Ou, seja, esses estudos se baseiam na crença de que quanto mais complexa a escrita de um aprendiz ou falante nativo de inglês, maiores são as orações independentes e subordinadas produzidas. No entanto, Biber, Poonpon e Gray (2011) questionam esse tipo de medida para se analisar a complexidade na escrita, pois, segundo esses autores, não há muitas evidências empíricas para suportar essa medida. Outros estudos dão suporte para esse posicionamento de Biber, Poonpon e Gray (2011). Lu (2011), por exemplo, comparou 14 medidas de complexidade sintática que são mais comumente usadas em estudos sobre o desenvolvimento de L2 em um *corpus* de escrita acadêmica de aprendizes chineses de inglês. A autora conclui que as medidas no nível sintagmático são melhores indicadores para uma escrita de qualidade do que as medidas de orações por T-Unit e T-Unit por sentenças.

Em outro estudo focando o estudo da complexidade da escrita acadêmica de aprendizes de uma segunda língua, Bardovi-Harlig (1992, p. 391) afirmou que “avaliando a complexidade sintática das composições escritas por adultos aprendizes avançados de segunda língua, a análise por T-Unit não parece refletir exatamente o conhecimento do aprendiz”.¹³ Para o autor, essa forma de análise da complexidade da escrita do aprendiz desconsidera características de orações coordenadas e subordinadas produzidas pelos aprendizes. Desconsidera, por exemplo, o uso excessivo e atípico de orações subordinadas. Uma das principais críticas a estudos que medem a complexidade da escrita a partir das medidas baseadas na T-Unit é a inconsistência encontrada nos resultados desses estudos (LU, 2011; BIBER, POONPON, GRAY, 2011). Biber, Poonpon e Gray, por exemplo,

12 No original: “The vast majority of these studies have relied on the construct of the T-unit: a main clause and all associated dependent clauses. Two specific measures have been especially popular: mean length of T-unit (MLTU), which relies on the overall length in words of the T-unit, averaged across all T-units in a text, and clauses per T-unit (C/TU), which relies on the number of dependent clauses per T-unit, again averaged across all T-units in a text. For example, the following sentence comprises a single T-unit that is relatively short (11 words) but includes two embedded dependent clauses: don’t know [[why I was expecting [to see something else]]”.

13 No original: “in evaluating the syntactic complexity of compositions written by advanced adult second language learners, T-unit analysis does not seem to reflect accurately the knowledge of the learner”.

fazem um levantamento de 18 estudos que analisam a relação de orações por T-Unit com a complexidade na escrita publicados em Quintero et al. (1998). Nesse levantamento, dos 18 estudos revisados, apenas 7 mostraram uma relação significativa entre complexidade e orações por T-Unit, enquanto 11 estudos não mostraram.

Esses estudos, no entanto, são consistentes quando apontam que há relação entre complexidade de sintagmas nominais e complexidade da escrita acadêmica. Parkinson e Musgave (2014) atentam para o fato de que na década de 1990 uma série de autores conduziram estudos que dão suporte para o estudo da complexidade do SN na escrita e o reflexo que essa unidade linguística tem no texto como um todo. Eles lembram que

estudos de escritores de L1 e L2 demonstraram que o desenvolvimento em crianças vai de fragmentos de sentenças a sentenças simples (Ishikawa, 1995), de sentenças à coordenação oracional e da coordenação à subordinação (Cooper, 1976). Nos últimos anos escolares e na universidade, a tendência é em direção a uma maior concisão, com complexidade construída cada vez mais no nível sintagmático do que no nível oracional (ver, por exemplo, Ferris, 1994; Grant & Ginther, 2000). Wolfe-Quintero et al. (1998, p.72) observam que escritores mais avançados tendem a usar formas mais reduzidas. Exemplos dessas formas reduzidas incluem: sintagmas preposicionados ou adjetivos ao invés de uma oração adjetiva; um particípio ao invés de uma oração adverbial; e um gerúndio ou infinitivo ao invés de uma oração nominal. Da mesma forma, Halliday (1993) observa que a escrita, embora oracionalmente mais simples do que a fala, favorece grupos nominais encaixados complexos¹⁴ (PARKINSON, MUSGAVE, 2014, p. 49).

De fato, o que todos esses estudos mencionados até o momento têm em comum é o fato de confirmarem que a complexidade do SN também reflete a complexidade na escrita. Com base nesses achados, Biber, Poonpon e Gray (2011) fazem uma investigação baseada em *corpus* para propor uma nova forma de se

14 No original: "Studies of L1 and L2 writers have shown that development in children is from sentence fragments to simple sentences (Ishikawa, 1995), from sentences to clausal coordination, and from coordination to subordination (Cooper, 1976). In the later school years and at university, the trend is towards greater conciseness, with complexity increasingly constructed at the phrasal rather than the clausal level (see, for example, Ferris, 1994; Grant & Ginther, 2000). Wolfe-Quintero et al. (1998, p. 72) note that more advanced writers tend to use more reduced forms. Examples of these reduced forms include: prepositional phrases or adjectives rather than an adjective clause; a participle rather than an adverbial clause; and a gerund or infinitive rather than nominal clause. Similarly Halliday (1993) notes that writing, while clausally simpler than speech, favours complex embedded nominal groups".

estudar a complexidade dos SNs em escrita acadêmica. Em seu trabalho, os autores compararam um *corpus* de escrita acadêmica formal (produzida por profissionais, ao invés de produzida por universitários) com um *corpus* oral. Essa distinção dos dois registros se deu, pois, segundo os autores, “é importante identificar complexidades que são comuns na escrita acadêmica, porém raras em conversação, e distingui-las das complexidades que são comuns na conversação”¹⁵ (BIBER, POOMPON, GRAY, 2011).

Em outras palavras, distinguir o que é comum apenas na escrita acadêmica, o que é comum apenas na conversação e o que os dois registros têm em comum. Os resultados dessa pesquisa mostraram que os elementos que caracterizam a complexidade dos SNs em escrita acadêmica são orações relativas não finitas e orações relativas com pronomes relativos *Wh-*, para orações com função de modificadores de SN. Já os elementos pré-modificadores de SN mais comuns são os adjetivos atributivos, nomes e SPreps. Além disso, os resultados mostraram que as medidas de T-Unit são melhores para se medir a complexidade da língua oral.

Biber, Poonpon e Gray (2011) sugerem, então, estágios hipotéticos do desenvolvimento de complexidade na escrita de aprendizes do inglês como segunda língua. Nessa proposta, os autores não dizem que uma estrutura é mais ou menos complexa, porém criaram etapas no desenvolvimento da escrita acadêmica, em que características mais comuns da fala se encontram nos primeiros estágios e características mais comuns da escrita acadêmica nos últimos. Esses estágios foram ilustrados no quadro a seguir:

Quadro 2: Fases hipotéticas do desenvolvimento da escrita acadêmica para características de complexidade do sintagma nominal

Estágio	Estrutura gramatical	Exemplo(s)
2	Incorporação de sintagmas simples em sintagmas nominais: adjetivos atributivos.	2g <i>It certainly has a nice flavor</i> (oral) 2h <i>Tom Jones is apparently a real name</i> (oral)

15 No original: “In particular, it is important to identify complexities that are common in academic writing but rare in conversation, and to distinguish those from complexities that are common in conversation”.

3	orações relativas <i>that</i>, especialmente com núcleo substantivo animado. incorporação de sintagmas simples em sintagmas nominais: substantivos como pré-modificadores. sintagmas <i>of</i> como pós-modificadores. sintagmas preposicionados simples, especialmente com outras preposições além de <i>of</i>, quando eles têm significado concreto/locativo.	3e ...<i>the guy that made that call</i> (fict) 3f ...<i>some really obscure cable channel</i> (fict) 3h ...<i>editor of the food section</i> (fict) 3i ...<i>house in the suburbs</i> (fict)
4	orações relativas não finitas. mais incorporação de sintagmas no SN por adjetivos atributivos e Substantivos como pré-modificadores. sintagmas preposicionados simples, especialmente com outras preposições além de <i>of</i>, quando eles têm significado abstrato.	4d ... <i>the method used here should suffice...</i> (acad) 4g <i>positive propagule size effects have been demonstrated for both plant and animal systems.</i> (acad) 4i <i>The specific growth rate at small (acad) population sizes</i>
5	preposição + oração não finita apositivo realizado por SN extensa incorporação de sintagmas: múltiplos sintagmas preposicionados como pós-modificadores, com níveis de incorporação	5a <i>The idea of using a Monte Carlo approach</i> (acad) 5c <i>The CTBS (the fourth edition of the test) was administered in 1997-1998</i> (acad) <i>The [presence of layered [[structures] at the [[[borderline] of cell territories]]]</i> (acad)

Fonte: Biber, Poonpon, Gray (2011, p. 30 e 31).

Esse quadro foi adaptado de forma a mostrar apenas os estágios que se relacionam aos modificadores de SN. Observa-se que, embora os tipos de modificação não sejam classificados como mais ou menos complexos, alguns tipos são comuns já nos estágios iniciais, em que características da língua oral são

também frequentes. Isso ocorre com os pré-modificadores realizados por adjetivos (estágio 2) e substantivos (estágio 3). À medida que os estágios ficam mais avançados, os pós-modificadores (dos tipos mais comuns em escrita acadêmica) vão surgindo (a partir do estágio 3).

Em estudo posterior ao de Biber e Gray (2011), Hundt, Denison e Schneider (2012) comparam a escrita acadêmica em inglês britânico e americano do *corpus* ARCHER. Nesse estudo diacrônico, os autores analisaram as mudanças no uso de modificadores restritivos de SNs realizados por orações relativas. Além disso, examinaram como as orações relativas contribuem para a complexidade do SN, além de tamanho e estrutura das orações. Os autores argumentam que as mudanças nos últimos 300 anos se deram pelo aumento da profissionalização do processo de publicação dos textos científicos.

As descobertas em relação à contribuição das orações relativas para os SN, em Hundt et al. (2012), mostraram que, assim como discutido por Biber e Gray (2011), houve uma mudança dramática no uso de pré-modificadores, apresentando um crescimento de mais de 100% nas ocorrências dos casos de pré-modificadores realizados por nomes.

No estudo de Hundt, Denison e Schneider (2012), que analisaram os casos de modificadores sintagmáticos (orações não finitas com *-ing* e *to* infinitivo) em um *corpus* histórico, foi descoberta uma troca em que mais modificadores sintagmáticos passaram a ser usados em detrimento dos modificadores oracionais, nos últimos 3 séculos. Essas descobertas contribuem para a caracterização do SN complexo que ocorre na escrita acadêmica contemporânea.

Em um estudo sobre a complexidade dos SNs nos materiais didáticos em universidades americanas, Capstone (2013) faz um levantamento comparativo entre textos de livros didáticos universitários e textos do curso preparatório para a prova GED (*General Educational Development*, uma prova feita por falantes não nativos de inglês que visam a estudar em universidades Americanas). O que motivou a autora a pesquisar tal assunto foi a dificuldade que os alunos tinham de entender os textos dos materiais da graduação, principalmente pela dificuldade de identificarem SNs complexos. Nesse trabalho, a autora considera SNs sem modificação como simples,

e modificados (pré- ou pós) como complexos. Os dados de O'Malley (2013) sugerem que os textos dos livros do curso preparatório para a prova GED apresentam uma quantidade de SNs complexos menor do que a dos textos dos livros dos cursos de mestrado. Além disso, os livros preparatórios para os alunos do curso de GED apresentavam SNs com menos elementos encaixados. A autora conclui que o material preparatório deveria apresentar SNs com estrutura mais complexa para que os alunos se preparassem melhor para o curso de mestrado.

Em outro estudo, Oliveira (2014) investigou a relação da estrutura do SN do inglês com a sua função sintática na oração em resumos acadêmicos. Nesse trabalho, OLIVEIRA (2014) classifica o SN de acordo com seus elementos em 3 complexidades, diferentemente de O'Malley (2013) que utiliza 2. Para o autor, o SN sem modificação é o mais simples; o SN com pré- ou pós-modificação são classificados igualmente, possuindo uma complexidade média; e o SN com pré- e pós-modificação são classificados como sendo mais complexos. Os resultados mostraram uma alta frequência de SNs de complexidade média e mais complexos nas posições de objeto, no gênero resumo acadêmico.

Em estudo sobre o uso de modificadores na escrita acadêmica, Parkinson e Musgave (2014) analisaram textos escritos por alunos em uma universidade na Nova Zelândia. Dois *corpora* foram comparados. O primeiro grupo era de alunos do curso de inglês para fins acadêmicos (IFE). Esse curso era feito por alunos que pretendiam cursar o mestrado, porém não possuíam a proficiência exigida da língua. O outro grupo era de alunos que cursavam o mestrado. Para passar na seleção, deveriam obter uma nota mínima no TOEFL, exame aplicado internacionalmente como forma de mensurar o nível de proficiência na língua inglesa. Os alunos de IFE demonstraram, segundo os autores, uma frequência baixa de substantivos e adjetivos participio como modificadores de SNs. Além disso, escolheram formas nominais atípicas como núcleo do SN, quando modificado por outro nome, como, por exemplo, o uso do termo “*generator*” em vez de “*generation*”, e utilizaram construções N N (substantivo modificador de substantivo) com um dos elementos no plural, que não seria uma construção frequente.

Em estudo posterior, que investiga as características dos SNs compostos por N N, Parkinson (2015) analisa as diferenças entre produções de aprendizes de inglês cuja língua materna prevê ou não o uso de N N. Para tanto, a autora compara a produção de dois grupos de aprendizes distintos: o primeiro formado por falantes nativos de mandarim e o segundo formado por falantes nativos de espanhol. A autora defende a hipótese de que o fato de a língua materna de um aprendiz permitir ou não a formação de SNs formados por N N interfere na produção desses SNs.

Para conduzir tal estudo, Parkinson (2015) comparou três *subcorpora* de aprendizes de inglês do projeto ICLE. Ela tinha como objetivo analisar, de forma mais precisa, o que pode influenciar aprendizes de inglês na produção de SNs pré-modificados por substantivos, além do nível de proficiência da língua inglesa que eles possuem. A autora conclui que a produção em inglês do aprendiz cuja língua materna admite a estrutura N N apresenta uma quantidade maior dessa estrutura, em comparação com a produção em inglês do aprendiz cuja língua materna não a admite. Por outro lado, muitos desses SNs produzidos não serão combinações típicas utilizadas pelos falantes nativos.

Em suma, um dos fatores que caracterizam a complexidade dos SNs do inglês, na escrita acadêmica profissional, são os modificadores que os acompanham. Os modificadores mais típicos são aqueles realizados por adjetivos atributivos, participios e substantivos (um terço dos SNs que recebem alguma modificação) como pré-modificadores, e sintagmas preposicionados e orações não finitas (além das finitas) como pós-modificadores.

LINGUÍSTICA DE *CORPUS* E ESCRITA ACADÊMICA

Neste capítulo, os conceitos de linguística de *corpus* serão discutidos, e será apresentado um breve histórico do seu surgimento e desenvolvimento para o que temos hoje, bem como será apresentada uma tipologia de *corpus*. Além disso, abordarei alguns trabalhos desenvolvidos com utilização de *corpora* de aprendizes, aprendizes brasileiros e escrita acadêmica, nos últimos anos.

3.1 Definição de *corpus* na linguística de *corpus*

Uma série de autores definiram *corpus* na linguística de *corpus*. Dentre as variadas definições (ATKINS, CLEAR, 1992; BACKER, 1995; SANCHES, 1995; BIBER et al., 1998; SARDINHA, 2004), uma que engloba as características centrais de um *corpus* é a definição de Sinclair (2005). Segundo o autor,

um corpus é uma coleção de partes de um texto linguístico em forma eletrônica, selecionado de acordo com critérios externos para representar, o melhor possível, uma língua ou variedade linguística como uma fonte de dados para pesquisa linguística.¹⁶ (SINCLAIR, 2005, p. 16).

De acordo com Berber Sardinha (2004), há alguns pontos importantes sobre o *corpus* que devem ser levados em consideração, e essa definição de Backer (1995) leva em consideração todos esses pontos, que são: a origem, o propósito, a composição, a formatação, a representatividade e a extensão.

Os estudos baseados em *corpus*, então, têm como objetivo estudar uma língua, ou variedade linguística, explorando a linguagem através de evidências encontradas nos *corpora* por meio de computadores (SARDINHA, 2004).

16 No original: "A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research."

3.2 Breve Histórico da linguística de *corpus*

O uso de um *corpus* para descrição linguística não é algo atual. De acordo com McCarthy e O’Keeffe (2010), o uso de *corpus* pode ser datado de muitos séculos, por volta do século XIII, quando estudiosos tiveram o trabalho de indexar manualmente a bíblia. Esse processo foi minucioso e demorado e teve por objetivo tornar disponível para os estudiosos da bíblia um índice das palavras que ocorriam nesse texto, com informação da página e da passagem bíblica. Outros trabalhos semelhantes foram desenvolvidos, incluindo um trabalho feito por 500 monges, em 1230, em um convento em Paris. Esses monges compilaram um índice de palavras do Vulgare (McCARTHY, O’KEEFFE; 2010, p. 03). Já no século XVIII, foram feitos trabalhos de *corpus* com as obras de Shakespeare, para ajudar outros estudiosos do famoso escritor. Outros trabalhos mais atuais também foram elaborados manualmente, dentre eles a criação do famoso dicionário *Oxford English Dictionary*, a partir do qual foram feitas análises em mais de três milhões de páginas, para se verificar o uso das palavras.

Conforme exemplificado anteriormente, durante muitos séculos, os trabalhos de *corpus* foram feitos manualmente. Foi apenas na década de 1950 que estudiosos começaram a criar formas eletrônicas de indexação (McCARTHY, O’KEEFFE, 2010). O padre jesuíta Roberto Busa começou a criar, nesse período, um índice eletrônico do trabalho completo de São Tomás de Aquino (*Index Thomisticus*), precisando de mais de uma década para completar o projeto (McCARTHY, O’KEEFFE, 2010).

No final da década de 1940, foram realizadas duas invenções importantes para o surgimento e desenvolvimento das ferramentas de linguística de *corpus*: o gravador de fita e o computador (HALLIDAY, MATTHISSEN, 2014). Enquanto o primeiro tornou possível captar e armazenar o discurso espontâneo, os computadores tornaram possível guardar e deixar disponível uma quantidade enorme de dados. A partir desse momento, uma série de linguistas começou a projetar *corpora* escritos e orais, como Randolph Quirk e W. Freeman Twaddell, da Universidade de Brown, que projetaram e começaram a implementar o primeiro *corpus* de textos escritos, na década de 1940, e como Halliday, que começou a

gravar o discurso espontâneo para análises na década de 1960 (HALLIDAY, MATTHISSEN, 2014).

No início da década de 1970, o trabalho de cientistas da informação e bibliotecários ajudou no desenvolvimento dos *softwares*, a partir do interesse em palavras-chave em contexto. Porém, de acordo com os autores, foi a partir da década de 1980 que os *softwares* de *corpus*, como os que conhecemos hoje, começaram a surgir. Desde então, vários *corpora* de referência, como COCA (*Corpus of Contemporary American English*), BAWE (*British Academic Written English*) e CIC (*Cambridge International Corpus*) foram criados. E o que ajudou a impulsionar o desenvolvimento de *corpora* como vemos hoje foi o rápido crescimento da capacidade de processamento dos computadores e das tecnologias da informação (BONELLI, 2010).

Com o avanço da tecnologia, a linguística de *corpus* ganhou mais importância entre os pesquisadores e gramáticos. Os estudos de *corpus*, gradativamente, tomaram um lugar central em investigações teóricas da língua. Como há inúmeros tipos de *corpora*, além de inúmeras formas na qual um *corpus* pode ser explorado, foi possível estabelecer perfis probabilísticos de sistemas gramaticais através das evidências encontradas nos *corpora*, baseadas em frequências (HALLIDAY, WEBSTER, 2005).

3.3 Tipologia

Hoje temos um número gigantesco de tipos diferentes de *corpora* (BONELLI, 2010). Por ser quase impossível listar os tipos de *corpora* com seus propósitos específicos, é possível falar, de forma genérica, sobre alguns tipos básicos de *corpora*. Os *corpora* são classificados de acordo com seu objetivo, de modo que podemos ter: *corpus* de amostra, *corpus* comparativo, *corpus* especial, *corpus* através do tempo e *corpus* bilíngue e multilíngue (BONELLI, 2010). E cada um desses tipos de *corpus* possui subtipos diversos.

Teixeira (2008) faz um levantamento dos tipos mais comuns de *corpus* discutidos por vários autores (ATKINS, CLEAR, OSTLER, 1992; BIBER, CONRAD, REPPEN, 1998; SARDINHA, 2002; SARDINHA, 2004; SINCLAIR, 2005; TAGNIN,

2007; ATKINS, RUNDELL, 2008). Levando-se em consideração Teixeira (2008), Granger (2004) e Bonelli (2010), apresento os principais tipos de *corpus* de acordo com sua finalidade, descritos no quadro 3.

Quadro 3: Os tipos mais comuns de *corpora*

Tipo	Sub-tipo	Descrição
Língua	Monolíngue	Contém textos escritos em apenas uma língua ou uma variedade linguística.
	Bilíngue/multilíngue	Contém textos escritos em mais de uma língua.
	Multivarietal	Contém textos escritos provenientes de variantes regionais ou nacionais distintas, de uma mesma língua.
Data de publicação	Sincrônico	Contém textos que compreendem um período de tempo.
	Diacrônico	Contém textos que compreendem vários períodos de tempo.
	Histórico	Contém textos de valor documental.
	Contemporâneo	Contém textos atuais.
	Fechado	<i>Corpus</i> para o qual, após a conclusão da compilação, não há possibilidade ou intenção de se adicionar mais textos.
	Aberto	<i>Corpus</i> constantemente atualizado, com a adição de novos textos, a fim de refletir o estado atual de uma dada língua ou variedade linguística.
Modo	Escrito	Contém textos escritos, transcrições de textos orais ou ambos.
	Oral	Contém áudios de uma dada língua com sua respectiva transcrição.

Conteúdo	Especializado	<i>Corpus</i> cujos textos podem ser representativos da língua geral (número balanceado de textos em gêneros e tipos textuais diversos).
	Regional	<i>Corpus</i> cujos textos são provenientes de uma ou mais variedades linguísticas.
Autoria	Nativo	<i>Corpus</i> cujos textos são produzidos por falantes nativos.
	Aprendiz	<i>Corpus</i> cujos textos são produzidos por aprendizes de uma língua estrangeira ou segunda língua.
	De autoria individual, coletiva ou institucional	<i>Corpus</i> cujos textos são artigos científicos ou textos de empresas, órgãos governamentais etc.
Organização interna	Paralelo (alinhado)	<i>Corpus</i> cujos textos são originalmente escritos em uma língua, acompanhados de suas respectivas traduções para uma ou mais línguas.
	Comparável	Dois ou mais <i>subcorpora</i> , cada um contendo textos originalmente escritos em uma dada língua ou variante linguística, mas com características semelhantes.
Uso na pesquisa	De estudo	O <i>corpus</i> que se pretende analisar.
	De referência	Um <i>corpus</i> usado para efeito de comparação, em geral maior do que o <i>corpus</i> de estudo.
Tamanho	Pequeno, pequeno-médio, médio, médio-grande e grande	<i>Corpus</i> assim caracterizado com referência ao número de palavras que contém.
Nível de codificação	Anotado	<i>Corpus</i> cujos textos são acrescidos de diversas informações extratextuais e linguísticas por meio da inserção de cabeçalhos e etiquetas.
	não anotado	<i>Corpus</i> cujos textos não têm o acréscimo de etiquetas e cabeçalhos.

Observa-se, a partir dos tipos de *corpus* apresentados no quadro 3, que a tipologia se refere a algumas categorias, sendo elas língua, data de publicação, modo, conteúdo, autoria, organização interna, uso na pesquisa, tamanho e nível de codificação.

O *corpus* de aprendiz, que é o *corpus* de estudo do presente trabalho, consiste em uma coleção de dados linguísticos na forma eletrônica, produzida por aprendizes de uma segunda língua. Esses dados são produzidos tanto por aprendizes de segunda língua (quando uma segunda língua é aprendida estando o aprendiz imerso no contexto dessa mesma língua) quanto por aprendizes de uma língua estrangeira (quando uma língua estrangeira é aprendida estando o aprendiz imerso no contexto da língua materna do aprendiz ou imerso no contexto de uma terceira língua) (GRANGER, 2012). Os textos coletados para esse fim devem ser textos autênticos, reproduzidos no *corpus* da forma exata em que foram escritos pelos aprendizes. Além disso, o *corpus* deve ser codificado de forma homogênea, seguindo padrões pré-estabelecidos. Alguns exemplos de *corpora* de aprendiz são os *corpora* compilados no projeto ICLE, de aprendizes de inglês, e no projeto MERLIN, de aprendizes de alemão, tcheco e italiano.

A questão da autenticidade dos textos em relação ao *corpus* de aprendiz é de certa forma problemática (GRANGER, 2002). Sinclair (1996) define um texto autêntico como aquele produzido de situações naturais de comunicação, não incluindo situações artificiais como experimentos, dentre outros. Porém, a língua estrangeira produzida por um aprendiz se concentra basicamente em sala de aula, em situações que podem ser consideradas, de certa forma, artificiais, pois a criação do *corpus* envolve algum tipo de controle por parte dos compiladores – por exemplo, na definição de um tópico a ser escrito, com um tamanho específico, em um tempo limitado. Por isso, o termo “autêntico” possui graus diferentes quando se trata de um *corpus* de aprendiz, podendo variar entre a produção natural, até uma produção de uma atividade produzida em sala de aula (GRANGER, 2002).

O *corpus* de nativos consiste em uma coleção de textos produzidos por falantes nativos de uma língua. Grande parte dos *corpora* disponíveis são *corpus* de nativos. Alguns *corpora* de nativos são usados em pesquisa para servirem de

comparação com *corpus* de aprendizes, como é o caso do *corpus* de nativo LOCNESS.

Os *corpora* comparativos são aqueles estudados de forma contrastiva. Eles são compilados seguindo critérios semelhantes e são de tamanhos parecidos. Além disso, há várias formas de se compararem *corpora*, como, por exemplo, geograficamente, comparando *corpus* do inglês americano e *corpus* do inglês britânico, *corpus* de aprendiz e *corpus* de nativo (BONELLI, 2010). Um exemplo de *corpora* compilados para serem comparativos são os *corpora* do projeto ICLE, que compilam *corpora* de aprendizes de falantes nativos de variadas nacionalidades, seguindo os mesmos critérios de criação, número de palavras, dentre outros critérios.

Um *corpus* de referência é um *corpus* representativo usado para ser comparado com um *corpus* de estudo (SARDINHA, 2004). Nessa classificação, um *corpus* pode ser classificado como sendo de vários tipos ao mesmo tempo. Um *corpus* monitor, por exemplo, pode ser utilizado como um *corpus* de estudo e um *corpus* especializado pode ser utilizado como um *corpus* de referência. Os motivos que levam um tipo específico de *corpus* a ser usado como *corpus* de referência dependem, principalmente, do objetivo do trabalho. Em Pan, Reppen e Biber (2016), um *corpus* especializado (escrita acadêmica profissional em inglês, da área de Administração, escrito por falantes nativos de inglês) é utilizado como *corpus* de referência em um estudo que analisa a escrita acadêmica profissional em inglês de falantes nativos de chinês. Já em Lúcio (2013), o *corpus* monitor COCA foi utilizado como *corpus* de referência. Nesse trabalho, a autora analisou dois *corpora* de estudo: o BR-ICLE e o LOCNESS.

Um ponto importante levantado por Granger (2002) em relação ao estudo com *corpus* de aprendiz se refere ao *corpus* utilizado para comparação. O pesquisador deve decidir pontos como o dialeto do *corpus* comparativo, já que há inúmeras variações de inglês, como americano, britânico, sul-africano, dentre outros. Outro ponto a ser considerado é o grau de proficiência no gênero estudado. Granger (2002) lembra que tanto a comparação com a escrita do falante nativo profissional

quanto a comparação com a escrita do nativo estudante podem trazer contribuições igualmente importantes.

De acordo com Granger (2002), a comparação do *corpus* de nativos com a de aprendiz pode contribuir para a descrição de uma série de características da interlíngua do falante não nativo. E essas características não são apenas de erros, mas de estruturas que são muito ou pouco utilizadas pelo aprendiz, além de estruturas que são atípicas para o falante nativo. Além disso, a autora lembra que se a pesquisa com a língua do aprendiz for estudada com um objetivo aplicado ao ensino, a comparação com a língua do nativo é essencial, visto que o objetivo do ensino de todas as línguas estrangeiras é que o aluno adquira o máximo de proficiência possível. E, para alcançar tal objetivo, a produção do aprendiz precisa se aproximar de algumas normas da língua apresentadas pelo nativo. Porém, quando o *corpus* de falantes nativos estudantes é utilizado, o pesquisador deve ser criterioso ao compará-lo com um *corpus* de aprendiz. Os estudantes nativos não necessariamente apresentam uma produção linguística que os professores desejam que seus alunos sigam. Ou seja, embora o *corpus* seja usado como controle, as diferenças encontradas entre os *corpora* de estudantes nativos e de aprendizes devem ser analisadas de forma descritiva, não prescritiva.

3.4 Tamanho do *corpus*

Um fator importante ao se fazer uma análise linguística com *corpus* é o seu tamanho. Reppen (2010) sugere que a questão do tamanho do *corpus* se relaciona a dois fatores principais, quais sejam, a representatividade e a praticidade. O autor argumenta que, embora em alguns casos seja possível representar um tipo de linguagem a ser estudada, em sua grande maioria, uma língua não pode ser 100% representada. Ele também salienta que, dependendo dos objetivos de estudo, um *corpus* pequeno pode ser o suficiente para atender às demandas de uma pesquisa.

De forma semelhante, Koester (2010) afirma que nem sempre um *corpus* grande precisa ser utilizado para certas análises linguísticas. Segundo a autora, um *corpus* pequeno ou médio pode ser não só suficiente para os objetivos do trabalho, mas até melhor do que um *corpus* grande, com milhões de palavras. Koester (2010)

argumenta que, ao trabalharmos com um *corpus* grande, temos um volume de dados muito grande também, o que dificulta o estudo linguístico em contexto. Por outro lado, ao trabalharmos com um *corpus* pequeno, temos a chance de estudar todos os casos. A autora conclui que

corpora menores e mais especializados têm uma vantagem distinta: eles permitem uma ligação mais próxima entre o corpus e o contexto em que os textos do corpus foram produzidos. Onde corpora muito grandes, através de sua contextualização, permitem compreensão a respeito de padrões gramaticais lexicais de uma língua como um todo, corpora pequenos especializados permitem compreensão a respeito de padrões de uso de uma língua em cenários particulares.¹⁷ (KOESTER, 2010, p. 67).

Ou seja, a análise de *corpora* pequenos e especializados também é uma boa fonte para se encontrar padrões linguísticos, conectando-os ao contexto de uso. Segundo Koester (2010), o uso de corpora especializados é muito relevante para os estudos de línguas para fins específicos. Isso ocorre porque

corpora menores, mais focados, que foram criados para uma pesquisa específica, ou com um propósito pedagógico, são muito mais prováveis de suscitar *insights* que são diretamente relevantes para o ensino e aprendizado para fins específicos¹⁸ (KOESTER, 2010, p. 67).

Além da relevância mencionada, professores e pesquisadores encontram muitas dificuldades para criarem *corpora* de aprendizes muito grandes.

3.5 Escrita acadêmica e *corpus* de aprendizes

Tanto alunos quanto pesquisadores e professores buscam, de várias formas, melhorar sua habilidade de escrita em uma língua estrangeira. Em um trabalho de Gea-Valor, Rey-Rocha e Moreno (2014), os autores buscaram identificar os maiores problemas enfrentados por pesquisadores espanhóis para publicarem estudos em inglês. Isso foi feito após constatarem que os pesquisadores espanhóis encontravam dificuldades para publicar trabalhos em periódicos internacionais em inglês. Os

17 No original: “smaller, more specialized *corpora* have a distinct advantage: they allow a much closer link between the *corpus* and the contexts in which the texts in the *corpus* were produced. Where very large *corpora*, through their de-contextualisation, give insights into lexico-grammatical patterns in the language as a whole, smaller specialized *corpora* give insights into patterns of language use in particular settings”.

18 No original: “Smaller, more focused corpora, which have been set up for a specific research or pedagogical purpose, are much more likely to yield insights that are directly relevant for teaching and learning for specific purposes.”

autores aplicaram questionários para alunos de 5 universidades espanholas, das áreas de Psicologia, Educação e Sociologia. Segundo Gea-Valor, Rey-Rocha e Moreno (2014), os pesquisadores espanhóis são avaliados de acordo com suas publicações internacionais, o que influencia diretamente seus aumentos salariais. E como grande parte dos periódicos internacionais são em inglês, os autores viram a necessidade de avaliar os problemas enfrentados pelos pesquisadores para publicar e as estratégias por eles adotadas. Essa exigência de publicação fez com que a busca por cursos como inglês para fins acadêmicos crescesse muito nos últimos anos.

Os informantes do estudo eram compostos, principalmente, por professores. Os resultados do questionário mostraram que menos de 30% dos participantes consideravam ter uma escrita avançada e mais de 40% afirmaram ter apenas uma escrita básica. A grande maioria dos informantes (89%) queria publicar em inglês para levar o resultado de suas pesquisas ao conhecimento dos colegas internacionais. Por outro lado, o motivo que levou 68% dos participantes a não tentarem publicar seus trabalhos em periódicos internacionais em inglês foi o fato de os informantes não acreditarem ter uma habilidade de escrita no padrão esperado pelos periódicos.

No questionário apresentado por Gea-Valor, Rey-Rocha e Moreno (2014) aos informantes, havia uma lista de itens que representavam dificuldades para publicar em inglês. Dentre os 6 itens apresentados (tema de interesse do periódico, resultados representativos, normas da revista, dentre outros), o que correspondeu à maior dificuldade dos informantes foi o referente às características da escrita em inglês que os informantes possuíam. Ou, seja, eles tinham dificuldades gramaticais, de vocabulário, e erros de estilo, como o uso de estruturas complexas. Eles não acreditavam possuir o conhecimento gramatical necessário para escrever um artigo. Ao analisar o percentual de artigos aceitos em inglês, Gea-Valor, Rey-Rocha e Moreno (2014) constataram que apenas 26% tiveram seus artigos aceitos para publicação com alterações. Dessa forma, os autores mostraram que a maior dificuldade que os pesquisadores espanhóis encontram para publicar seus estudos em inglês é a falta de habilidade na escrita acadêmica.

Embora tenha sido conduzido na Espanha, esse estudo é um bom exemplo de como um pesquisador, aluno ou professor se sente ao tentar escrever um texto em língua estrangeira, principalmente quando o objetivo é a publicação de trabalhos em periódicos estrangeiros. O estudo mostra, assim, a relevância do conhecimento linguístico para a vida acadêmica e profissional do pesquisador, professor e aprendiz.

Tendo em vista a importância do conhecimento acerca das características da escrita do aprendiz, linguistas buscam descrevê-las a partir de corpora de aprendiz falante de inglês como segunda língua. Os estudos baseados em *corpus* de aprendiz visam, principalmente, fornecer dados empíricos sobre a língua do aprendiz, com o objetivo de ajudar os alunos a desenvolverem suas habilidades em uma dada língua estrangeira. Conforme apontado por Almeida (2014), quando *corpora* de aprendizes são analisados e comparados com *corpora* de nativos são revelados:

quais características linguísticas da língua-alvo são mais (sobreuso) ou menos (subuso) usadas em comparação com falantes nativos, quais são as áreas mais problemáticas para os aprendizes de um determinado país ou qual a intensidade da influência da língua materna na produção dos aprendizes. Em outras palavras, a análise dos dados dos corpora de aprendiz, ou a análise da interlíngua dos aprendizes pode revelar algumas das dificuldades linguísticas dos alunos e, uma vez cientes das necessidades dos mesmos, professores podem melhor planejar e definir suas ações pedagógicas (ALMEIDA, 2014, p. 35).

Atualmente, uma série de trabalhos tem sido realizada para se estudar as características da escrita acadêmica. O gênero acadêmico possui uma série de subgêneros, dentre os quais ensaios, resumos, estudos de caso, resenhas, artigos e relatórios. Esses estudos focam aspectos diferentes da escrita, analisando questões voltadas para o léxico, sintaxe/semântica, dentre outros.

Os estudos baseados em *corpus* de aprendiz comparam *corpus* de aprendizes com *corpus* de referência para tomar conhecimento de estruturas sobreutilizadas, subutilizadas, mais comumente utilizadas e até mesmo estruturas que o aprendiz não utiliza. Além disso, esse tipo de pesquisa permite comparar *corpus* de aprendiz que fala uma língua materna específica com *corpus* de aprendiz que possui uma língua materna diferente. Essa comparação permite ao pesquisador

descobrir as características do aprendiz em geral e as características que são específicas de aprendizes com uma língua materna específica (GRANGER, 2003).

No Brasil, muitas pesquisas voltadas para a escrita acadêmica focam o uso de verbo e de advérbio (RECSKI, 2004; TENUTA, ORFANO, OLIVEIRA, 2015), na modalidade (VIANA, 2006), em estudos de colocados (BÉRTOLI, SHEPHERD, 2015), em uso de *corpus* em sala de aula (ORENHA-OTTAIANO, 2012), dentre outros.

No campo da modalidade, Ramos (2005) discute o uso de expressões modais, incluindo-se verbos modais e advérbios, por aprendizes brasileiros de inglês. Nesse trabalho, o autor analisa o *corpus* de aprendizes brasileiros (BR-ICLE, com 33 mil palavras na época), com o auxílio de 2 outros *corpora* de falantes nativos do espanhol (SP-ICLE) e tcheco (TC-ICLE), e compara-o com um *corpus* de falantes nativos do inglês (LOCNESS). Os resultados mostraram que os aprendizes brasileiros utilizaram muito mais os verbos modais *can* e *will* do que os nativos. Ramos (2005) argumenta que esses dois verbos são mais frequentes na língua oral, em comparação com o gênero acadêmico, e aponta a possibilidade de os aprendizes terem dificuldade de diferenciar o gênero acadêmico do registro oral.

Em trabalho mais recente, Tenuta, Oliveira e Orfano (2014) estudaram o uso de expressões modais por aprendizes brasileiros em textos acadêmicos. Nesse estudo, as autoras analisam o uso de advérbios com valor modal. Os dados mostraram que, além de usarem mais advérbios com valor modal do que os aprendizes brasileiros, os nativos também os usam com frequência mais equilibrada do que os aprendizes, não demonstrando uma frequência muito alta ou baixa por advérbio. Além dos advérbios, as autoras também analisaram os pacotes lexicais que possuem advérbio indicando modalidade, como *likely to*, *almost certainly*, *would probably*, dentre outros. Os resultados mostraram que os aprendizes apresentaram uma frequência baixíssima destes pacotes, indicando que eles dependem mais dos verbos modais. Assim como Ramos (2005), Tenuta, Oliveira e Orfano (2014) constataram um maior uso dos verbos modais *will* e *can* na escrita dos aprendizes brasileiros.

Em Oliveira (2015), o uso dos advérbios intensificadores por aprendizes brasileiros de inglês é investigado. Nesse estudo, a autora compara a escrita de universitários brasileiros (BRICLE) com a escrita de universitários falantes nativos do inglês (LOCNESS). Dentre as descobertas, estão um maior uso de advérbios amplificadores, como *very* e *really*, por parte dos aprendizes, e um maior uso de advérbios atenuadores, como *nearly* e *merely*, por parte dos nativos. Porém, após realizar uma análise qualitativa, a autora constatou uma influência da língua oral na escrita tanto dos estudantes brasileiros quanto dos nativos, visto que formas mais recorrentes na fala foram encontradas em suas produções escritas, tornando-as mais informais. Desta forma, a autora conclui que ambos os grupos, nativos e brasileiros, ainda apresentam uma escrita inexperiente, tendo em vista o uso de advérbios intensificadores.

Ainda no campo de estudos dos advérbios, Costa (2015) investiga o uso de advérbios de ligação em artigos acadêmicos escritos por brasileiros e publicados em periódicos na língua inglesa. Após criar dois *corpora*, um de referência (escritores nativos) e um de estudo (escritores brasileiros), o autor analisou 7 tipos diferentes de advérbios e concluiu que, mesmo entre os escritores mais experientes e publicados, há uma diferença significativa entre os escritores de L1 e de L2.

Outros autores, como Almeida (2014), buscam utilizar linguística de *corpus* no ensino de línguas. Nesse estudo, que foca as colocações com verbos de alta frequência, Almeida (2014) contrastou um *corpus* em inglês de universitários brasileiros com dois *corpora* de referência em inglês (COCA e BNC). A autora demonstrou que os aprendizes brasileiros apresentaram dificuldades com as colocações, evidenciadas por alguns desvios colocacionais, como combinações de verbo com advérbios (uso de *get out* em vez de *get away*), ou com nomes (uso de *get a work* em vez de *get a job*). Finalmente, a autora propõe atividades que seriam benéficas para uma maior conscientização linguística por parte dos estudantes.

Observa-se que as pesquisas envolvendo a escrita acadêmica em inglês de aprendizes brasileiros têm enfoques mais comuns nos estudos no verbo e na modalidade. O presente trabalho visa a ajudar a preencher essa lacuna nos estudos do sintagma nominal, mais especificamente a complexidade do sintagma nominal

produzido por aprendizes brasileiros, juntando-se a esses outros trabalhos mencionados. Vale ressaltar que, apesar de este trabalho ser de origem descritiva, busco contribuir também para a área de ensino de inglês como língua estrangeira para brasileiros, identificando padrões de uso comuns, incomuns ou até incorretos. Assim, discutirei, no capítulo seguinte, os procedimentos metodológicos para se atingir os objetivos do trabalho proposto.

METODOLOGIA

Neste capítulo, foram descritos os processos metodológicos envolvidos na elaboração desta pesquisa. Três *corpora* foram utilizados neste trabalho, a saber: o BR-ICLE (*corpus* de estudo), o LOCNESS (*corpus* comparativo) e o BAWE (*corpus* de referência), que representam a produção linguística dos aprendizes brasileiros, a produção linguística de falantes nativos de inglês americano e britânico e a produção linguística de universitários britânicos de várias áreas de estudo, respectivamente. Assim, descreverei o processo de compilação de cada *corpus*, os pesquisadores envolvidos em sua criação, além dos procedimentos metodológicos acerca da criação dos *corpora* e as informações pertinentes a respeito dos informantes.

Além disso, a ferramenta de *corpus* utilizada nesta pesquisa, o *WordSmith Tools*, é descrita e suas funções explicadas. Da mesma forma, as configurações utilizadas durante o processo da seleção de dados são explicadas. Em seguida, são descritos os critérios utilizados para a seleção e análise dos dados, dentre eles as classificações adotadas e os critérios de complexidade de SN propostos. Finalmente, os testes estatísticos aplicados neste estudo para a validação de resultados estatisticamente relevantes são descritos.

4.1 Os *corpora* de referência, de estudo e comparativo

Trabalhei com três *corpora* diferentes para a realização desta pesquisa. O *corpus* de referência de escrita acadêmica, o British Academic Written English (BAWE), que será utilizado apenas em uma parte do trabalho; o *corpus* de estudo BR-ICLE; e o *corpus* comparativo LOCNESS.

O *corpus* de aprendiz brasileiro foi comparado ao *corpus* de falantes nativos do inglês, pois essa comparação pode ajudar na descrição da produção do aprendiz, conforme apontado por Granger (2002).

É bem claro, portanto, que a descrição enriquecida da língua-alvo propiciada pelo corpus de nativo é um extra para o ensino da língua estrangeira. No entanto, a visão que eu gostaria de apresentar é a de que isso não é suficiente. Corpora de falantes nativos proporcionam informações valiosas sobre a frequência e o uso de palavras, sintagmas e estruturas, mas não dão nenhuma indicação da dificuldade que elas apresentam para o aprendiz em geral ou para uma categoria específica de aprendizes. Eles vão, portanto, sempre ser de valor limitado e podem, inclusive, levar a decisões pedagógicas mal julgadas, a não ser que eles sejam complementados com o tipo de dado igualmente rico e pedagogicamente mais relevante proporcionado pelo corpus de aprendiz¹⁹. (GRANGER, 2002, p. 21-22).

Ou seja, a análise conjunta de um *corpus* de falantes nativos com um *corpus* de aprendizes visa trazer informações relevantes para o ensino dessa dada língua como língua estrangeira. Ao usar um *corpus* de aprendiz BR-ICLE e compará-lo ao *corpus* de nativos LOCNESS, visto encontrar padrões que auxiliem na descrição da língua inglesa produzida por aprendizes brasileiros.

Nas seções a seguir, os *corpora* de aprendiz e nativos utilizados no presente estudo são descritos.

4.1.1 O corpus de referência

O *corpus* de referência de falantes nativos do inglês selecionado para este trabalho foi o BAWE (British Academic Written English), que faz parte do projeto *Uma investigação dos gêneros de escrita avaliadas na educação superior britânica*.²⁰ O *corpus* BAWE foi desenvolvido nas Universidades de Warwick, Reading e Oxford Brookes e foi dirigido por Hilary Nesi e Sheena Gardner, da Universidade de Warwick, Paul Thompson, da Universidade de Reading, e Paul Wickens, da Universidade de Oxford Brookes, com financiamento do ESRC (RES-000-23-0800)²¹.

19 No original: "it is quite clear therefore that the enriched description of the target language provided by native corpora is a plus for foreign language teaching. However, the view I would like to put forward is that it is not sufficient. Native corpora provide valuable information on the frequency and use of words, phrases and structures but give no indication whatsoever of the difficulty they present for learners in general or for a specific category of learners. They will therefore always be of limited value and may even lead to ill-judged pedagogical decisions unless they are complemented with the equally rich and pedagogically more relevant type of data provided by learner corpora".

20 No original: "An Investigation of Genres of Assessed Writing in British Higher Education".

21 HEUBOECK, A; HOLMES, J; NESI, H. The BAWE corpus manual for the project entitled 'An

Vale ressaltar que esse *corpus* está sendo usado como *corpus* de referência, conforme definição apresentada por Teixeira (2008), com o objetivo de criar a lista de palavras-chave do *corpus* de estudo (BR-ICLE) e do *corpus* comparativo (LOCNESS). Acredito que as diferenças lexicais entre inglês americano e britânico não serão um problema para este estudo, já que os próprios aprendizes brasileiros produzem ensaios tanto em um quanto em outro. Além disso, o *corpus* LOCNESS também é uma mistura de ensaios de inglês americano e britânico, como veremos a seguir.

Conforme salientado por Sardinha (2004), a escolha do *corpus* de referência para a criação de palavras-chave de um texto depende do objetivo do trabalho. O autor argumenta que

a composição do corpus de referência influencia os tipos de palavras que podem se tornar chave. Um corpus de características genéricas semelhantes ao corpus de estudo tende a filtrar, ou seja, eliminar os elementos genéricos (relativos a um mesmo gênero) em comum, resultando em uma lista de palavras-chave que não inclui esses elementos. Alguns traços linguísticos que podem ser filtrados são, entre outros, marcadores discursivos privilegiados, escolhas lexicais típicas e formas verbais flexionadas em comum (SARDINHA, 2004, p. 101-102).

Ou seja, ao se utilizar um *corpus* de gênero diferente do *corpus* de estudo, as palavras-chave serão termos típicos do gênero ou registro estudado. E ao se utilizar um *corpus* semelhante como *corpus* de referência para geração de palavras-chave, essas palavras típicas não serão incluídas na lista de palavras-chave. Por esse motivo, utilizei um *corpus* de ensaio acadêmico como *corpus* de referência, buscando evitar que palavras comuns nos ensaios acadêmicos, como palavras integrantes de marcadores discursivos, e expressões comuns que possuem uma forma fixa. Um exemplo é o substantivo *addition*, que faz parte do conector discursivo *in addition*, comum no gênero acadêmico.

O *corpus* BAWE foi compilado entre os anos de 2004 a 2007 e possui uma coletânea de textos escritos por alunos universitários britânicos. Ele está disponível em três formatos: uma versão .xml, em que o *corpus* está anotado; uma versão .txt, que contém uma quantidade mínima de notações e marcações; e uma versão em PDF, que apresenta os documentos originais. Ele também pode ser acessado pela

internet na interface de busca do próprio projeto.²² Para o presente estudo, foi utilizada a versão .txt do *corpus*, visto que o *corpus* de aprendizes e o *corpus* de nativos utilizados também se encontram nesse formato.

Esse *corpus* foi escolhido como *corpus* de referência porque, além de ser grande (6.506.995 palavras), é composto por ensaios acadêmicos de alunos de cursos variados. Os textos que o compõem foram coletados de alunos que cursavam 35 disciplinas em 4 grandes áreas, quais sejam: artes e ciências humanas, ciências sociais, ciências exatas e ciências biológicas. Os estudantes que produziram os ensaios integrantes do *corpus* eram de todos os anos do curso de graduação e também de cursos de mestrado, conforme tabela 1.

Tabela 1: Gêneros de escrita acadêmica do corpus BAWE por áreas

	Artes e Humanas	Ciências Biológicas	Ciências Exatas	Ciências Sociais	Total
Estudo de caso	0	91	37	66	194
Crítica	48	84	76	114	322
Especificação de <i>design</i>	1	2	87	3	93
Escrita empática	4	19	9	3	35
Ensaio	602	127	65	444	1238
Exercícios	14	33	49	18	114
Explicação	9	117	65	23	214
Crítica literária	7	14	4	10	35
Recontagem metodológica	18	158	170	16	35
Recontagem narrativa	10	21	21	19	75
Pergunta problema	0	2	6	32	40
Proposta	2	26	19	29	76
Relatório de pesquisa	9	22	16	14	61
Total	724	720	624	791	2858

Fonte: Manual do *corpus* BAWE, p. 7

É interessante observar que o gênero que obteve um maior número de textos e, conseqüentemente, um maior número de palavras foi o gênero “ensaio”. Dentre

²² Informações retiradas do Manual do Corpus BAWE.

os 13 subgêneros produzidos, na área de Ciências Humanas e Ciências Sociais, o ensaio foi o gênero mais produzido; na área de Ciências Biológicas, foi o segundo mais produzido e, na área de Ciências Exatas, foi o quarto mais produzido. Por ser um gênero comum e bem difundido dentro da comunidade acadêmica, o ensaio acadêmico facilita a criação de um *corpus* tanto de falantes nativos quanto de aprendizes de uma segunda língua.

Foram selecionados, para o presente estudo, apenas os textos que correspondiam a ensaios acadêmicos, conforme legenda disponibilizada no manual de uso do *corpus*. No total, o BAWE possui 1.206 arquivos <.txt> de ensaios acadêmicos, o que totaliza 3.120.263 palavras (e 62.664 tokens).

4.1.2 O *corpus* de aprendizes brasileiros

O CECL (*Centre for English Corpus Linguistics*) é responsável pela compilação de vários *corpora*, dentre os quais os *corpora* de aprendizes ICLE (*International Corpus of Learner English*), FRIDA (*French Interlanguage Database*), LINDSEI (*Louvain International Database of Spoken English Interlanguage*), LONGALE (*Longitudinal Database of Learner English*) e VESPA (*The Varieties of English for Specific Purposes dAtabase*); o *corpus* pedagógico TeMa, que consiste em materiais didáticos; os *corpora* multilíngues PLECI (*Poitiers-Louvain Échange de Corpus Informatisés*) e MULT-ED (*Multilingual Editorial Corpus*); e também *corpora* de falantes nativos do inglês, como o LOCNESS (*Louvain Corpus of Native English Essay*).

O Projeto ICLE consiste em vários *subcorpora* compostos por textos de escrita acadêmica em inglês, produzidos por falantes aprendizes dessa língua, cuja língua materna seja outra. Esse projeto possui pesquisadores espalhados pelo mundo, ligados a universidades de vários países. Há um total de 24 *subcorpora*, representantes de línguas maternas diferentes, a saber: alemão, árabe, búlgaro, chinês, esloveno, espanhol, finlandês, francês, grego, holandês, húngaro, italiano, japonês, lituano, norueguês, paquistanês, polaco, português de Portugal, português do Brasil, russo, sueco, tcheco e turco. O BR-ICLE, é o *subcorpus* que representa o português do Brasil. O projeto ICLE possui vários *subcorpora* ainda em construção,

uma vez que o objetivo do projeto é que cada *subcorpus* tenha um tamanho de 200 mil palavras, o que consiste em um *corpus* pequeno.

O *corpus* foi compilado por professores da PUC de São Paulo e já foi concluído. No *site* do projeto ICLE²³, porém, o *status* do *subcorpus* BR-ICLE não consta nem como completo, nem como em andamento. De acordo com o professor responsável pela compilação do *corpus* no Brasil, o BR-ICLE completo ainda será disponibilizado na terceira versão do CD do ICLE, sem data definida até o presente momento.

Os informantes que contribuíram para esse *corpus* são estudantes universitários que apresentam pelo menos nível intermediário da língua inglesa. Para escrever o ensaio, esses estudantes receberam uma lista de tópicos variados, que pode ser vista no Anexo 1. Os ensaios deveriam apresentar, no mínimo, 500 palavras e no máximo 1.000. A versão do *corpus* utilizado no presente estudo possui aproximadamente 160 mil palavras, com um total de 181 informantes.

Para a produção dos ensaios, os alunos poderiam consultar ferramentas de referência linguística como dicionários e gramáticas, com a condição de ser uma produção unicamente do aluno. Ou seja, embora o informante pudesse consultar uma ferramenta, ele não poderia consultar falantes nativos nem outras obras, como livros e artigos. Além disso, o tempo para a elaboração desses artigos não foi limitado.

Foi utilizado, no presente estudo, uma versão .txt do *corpus* BR-ICLE. Esse arquivo é composto pelos ensaios acadêmicos separados pela palavra *fileseparator* e por um código, que identifica o aluno. Esses dois códigos não foram considerados para o cálculo do total de palavras do *corpus*.

4.1.3 O *corpus* comparativo de falantes nativos

O terceiro *corpus* utilizado nesta pesquisa é o LOCNESS, composto por ensaios argumentativos de estudantes americanos e britânicos com nota máxima. Ele possui 324.304 palavras, sendo 60.209 palavras de estudantes britânicos,

23 Informações em <<https://uclouvain.be/en/research-institutes/ilc/cecl/icle-partners.html> último acesso em 25/04/2017>

95.695 palavras de universitários britânicos e 168.400 palavras de universitários americanos²⁴.

Esse *corpus* foi utilizado para ser comparado com o *corpus* de estudo (BR-ICLE). O LOCNESS foi escolhido por vários motivos. O primeiro é o fato de ter sido compilado pelo mesmo laboratório responsável pela compilação do BR-ICLE e por ter sido compilado de forma semelhante ao BR-ICLE. O segundo é o fato de ser um *corpus* disponibilizado gratuitamente através do *site* da Universidade Católica do Louvain.²⁵ Além desses dois motivos, optei por utilizar esse corpus para compará-lo ao corpus de aprendiz, porque as condições para a produção dos textos foram semelhantes. Um texto publicado, por exemplo, passou por um processo de revisão pelo autor, pelos pareceristas, e até mesmo por revisores profissionais, o que não ocorre com o texto do corpus de aprendiz.

Todos os informantes que escreveram os ensaios para o LOCNESS eram falantes nativos de inglês e, em mais de 90% dos casos, ambos os pais desses informantes também falavam inglês como primeira língua, totalizando 322 informantes. Os textos selecionados são de ensaios cuja nota foi a mais alta (nota A). Os tópicos para elaboração dos textos acadêmicos sugeridos aos alunos que forneceram textos para a composição do LOCNESS também são variados e estão listados no anexo 2. Foi pedido aos alunos que escrevessem ensaios de, no mínimo, 500 palavras.

Observa-se que os alunos que escreveram ensaios para o *corpus* LOCNESS puderam optar por assuntos diversos: temas polêmicos, problemas da sociedade moderna, tópicos que fazem parte do dia a dia dos alunos, dentre outros. De forma semelhante, os tópicos sugeridos para elaboração de ensaios acadêmicos para o *corpus* de aprendizes BR-ICLE seguiram a mesma variação do LOCNESS.

4.2 Softwares utilizados

A ferramenta computacional escolhida para se trabalhar com os *corpora* selecionados foi o *Wordsmith Tools*. Embora existam outras ferramentas disponíveis,

24 Informações retiradas do *site* <<http://www.uclouvain.be/en-cecl-locness.html>>.

25 Endereço do *site*: <<http://www.uclouvain.be/en-cecl-locness.html>>.

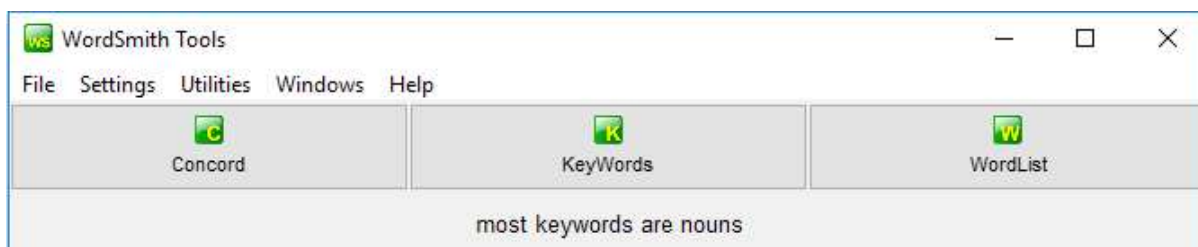
esta foi escolhida por atender a todas as necessidades referentes à geração de lista de palavras, lista de palavras-chave, criação de linhas de concordância, dentre outras. Além disso, ela possui uma interface simples e de fácil manejo.

Além dessa ferramenta, utilizei planilhas do Excel para realizar a análise das linhas de concordância, somar valores e criar os gráficos e as tabelas utilizados neste trabalho.

4.2.1 O *Wordsmith Tools*

Conforme mencionado, com o *Wordsmith* é possível analisar um ou vários *corpora* a partir de várias funções, incluindo-se concordâncias, listas de frequências, colocados. Esse *software* não é gratuito, podendo ser adquirido na internet, via *download*, no endereço <http://www.lexically.net/wordsmith/>. Apesar de a ferramenta não ser gratuita, é possível utilizar a versão demo gratuitamente, porém com restrições na quantidade de resultados exibidos. Ela possui um *layout* simples, como pode ser visto na figura 3, a seguir.

Figura 3: A ferramenta *WordSmith Tools*



Conforme mostrado na figura 3, as 3 (três) funções básicas da ferramenta são o *WordList*, o *Keywords* e o *Concord*. Para o processo de seleção dos dados do presente estudo, utilizei todas as três funcionalidades, que serão descritas a seguir.

O *Word List* é uma funcionalidade do *WordSmith* que gera uma lista contendo todas as palavras de um *corpus* selecionado. Esse mecanismo contabiliza o número de ocorrências de cada uma das palavras e as apresenta em ordem decrescente de frequência (utilizado neste trabalho) ou na ordem em que o pesquisador preferir. A lista de palavras possui várias utilidades, sendo uma delas a criação das palavras-chave.

Figura 4: Exemplo de lista de palavras da ferramenta WordSmith Tools

N	Word	Freq.	%	Texts	%
1	THE	21,105	6.51	14	100.00
2	TO	10,753	3.32	14	100.00
3	OF	10,728	3.31	14	100.00
4	AND	8,325	2.57	14	100.00
5	A	6,846	2.11	14	100.00
6	IN	6,367	1.97	14	100.00
7	IS	6,307	1.95	14	100.00
8	THAT	4,919	1.52	14	100.00
9	IT	3,219	0.99	14	100.00
10	BE	3,196	0.99	14	100.00
11	FOR	3,145	0.97	14	100.00
12	AS	2,832	0.87	14	100.00

Conforme mostrado na figura 4, a lista gerada pelo *WordList* com o corpus LOCNESS exibe a palavra, seguida do número de ocorrências da mesma no *corpus*. Em seguida, aparece uma porcentagem indicadora do quanto essa frequência representa no número total de palavras do *corpus*. E, finalmente, há uma indicação de em quantos textos esse termo aparece e outra de quanto essa frequência representa no número total de textos do *corpus*. Nessa figura, a palavra *the* ocorreu 9.036 vezes, o que representa 5,68% das palavras do *corpus*, tendo aparecido em 1.206 textos diferentes, o que representa 100% dos textos. Essa informação é importante para se observar se um termo ou expressão ocorreu em apenas um texto isolado ou se sua frequência pode ser algo regular.

Como essa funcionalidade do *WordSmith Tools* inclui todas as palavras do *corpus*, é esperado que as palavras mais frequentes do *corpus* sejam palavras gramaticais, como artigos, preposições e verbos auxiliares, já que esses são elementos que existem em número mais limitado na língua (em comparação com as palavras lexicais) e que são essenciais em um texto, ou seja, de grande

funcionalidade na expressão de conteúdos linguísticos. Observe, na figura 4, que 10 das 12 palavras mais frequentes são gramaticais.

Por esse motivo, também, a seleção dos dados foi feita a partir da função *Keywords*, conforme será discutido na próxima seção deste capítulo. Para gerar essa lista de palavras-chave, é necessário, antes, gerar a lista de palavras de todos os *corpora* estudados.

Com a utilização do *Keyword List*, pode-se criar uma lista de palavras-chave do *corpus* de estudo que pode ser comparada com uma lista de palavras de um *corpus* de referência. Esse instrumento de análise pode ser utilizado para se identificarem palavras características em um texto ou gênero. De acordo com Evison (2010),

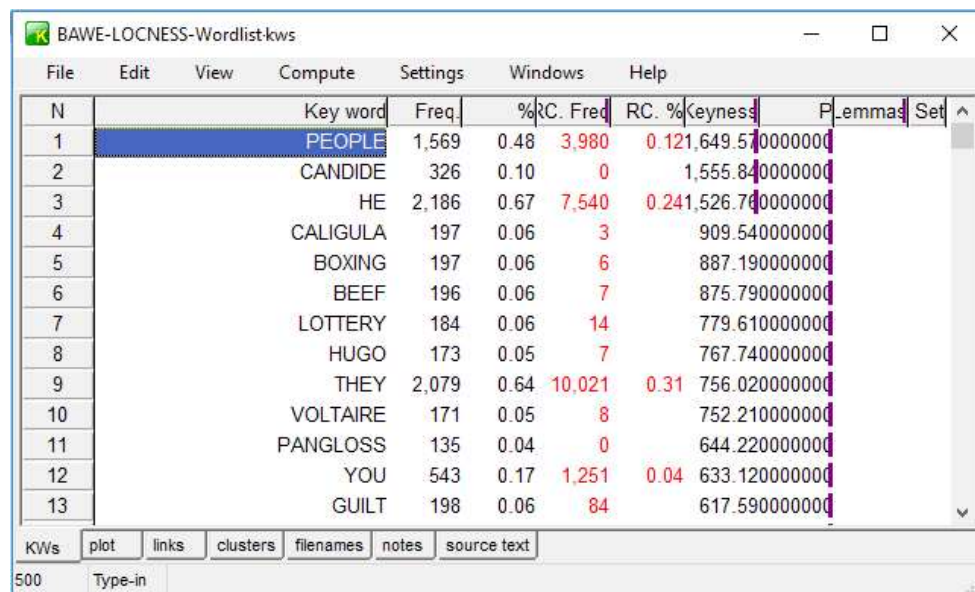
palavras-chave não são, necessariamente, as palavras mais frequentes em um corpus, mas são aquelas palavras que são identificadas por comparação estatística de um corpus 'alvo' com outro corpus maior, que é referido como o corpus de 'referência'. Esta identificação envolve a comparação automática de listas de palavras usando softwares como o Keyword do programa WordSmith Tools. A lista de palavras-chave inclui itens que são ou significativamente relevantes (palavras chave positivas) ou infrequentes (palavras-chave negativas), e é um ponto de partida útil para muitas análises linguísticas de corpus (SCOTT, 1999 e este volume; HUNSTON, 2002; REPPEN; SIMPSON, 2002; McENERY et al. 2006). Apesar de haver disponível várias formas de calcular a significância estatística, um teste de 'keyness' é especialmente útil para a análise de dados de corpus, pois, sendo baseado em um teste de log-likelihood (LL) (DUNNING, 1993), ele não se baseia na suposição de que os dados têm uma distribuição normal.²⁶ (EVISON, 2010, p. 127).

Conforme mencionado por Evison (2010), o *WordSmith Tools* compara duas listas de palavras: a primeira do *corpus* de referência e a segunda do *corpus* de estudo.

26 No original: "Key words are not necessarily the most frequent words in a corpus, but they are those words which are identified by statistical comparison of a 'target' corpus with another, larger corpus, which is referred to as the 'reference' or 'benchmark' corpus. This identification involves the automatic comparison of word lists using software such as the WordSmith Tools Keyword program. A key-word list includes items that are either significantly frequent (positive key words) or infrequent (negative key words), and is a useful starting point for many corpus linguistic analyses (Scott 1999 and this volume; Hunston 2002; Reppen and Simpson 2002; McEnery et al. 2006). Although there are several ways of calculating statistical significance available, a test of 'keyness' is especially useful for the analysis of corpus data because, being based on a log-likelihood (LL) test (Dunning 1993), it is not predicated on the assumption that data have a normal distribution".

Para se utilizar a função *Keyword List*, são necessários dois *corpora*: um de referência, que deve ser maior do que o *corpus* de estudo, e o *corpus* de estudo. A partir dessa função, é possível gerar a lista de palavras-chave, exibida na figura 5.

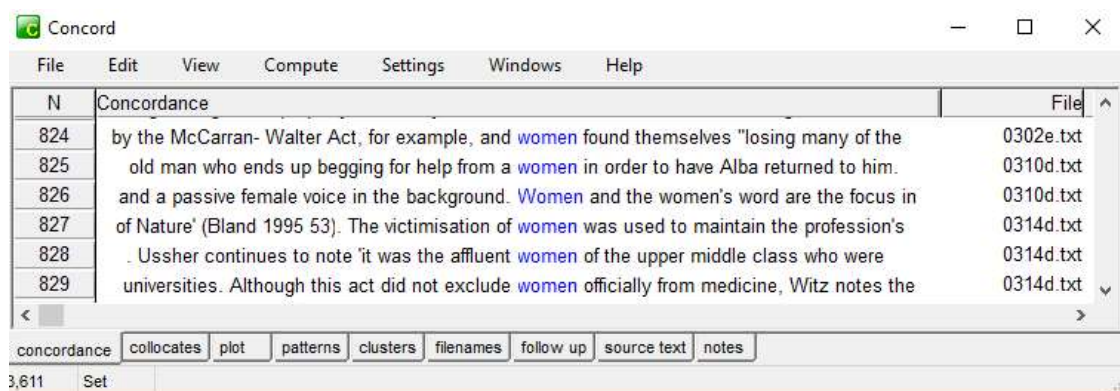
Figura 5: Exemplo de lista de palavras-chave do corpus LOCNESS com o corpus de referência BAWE



N	Key word	Freq.	%RC. Fred	RC. %	Keyness	P.lemmas	Set
1	PEOPLE	1,569	0.48	3,980	0.121,649.57	00000000	
2	CANDIDE	326	0.10	0	1,555.84	00000000	
3	HE	2,186	0.67	7,540	0.241,526.76	00000000	
4	CALIGULA	197	0.06	3	909.54	00000000	
5	BOXING	197	0.06	6	887.19	00000000	
6	BEEF	196	0.06	7	875.79	00000000	
7	LOTTERY	184	0.06	14	779.61	00000000	
8	HUGO	173	0.05	7	767.74	00000000	
9	THEY	2,079	0.64	10,021	0.31	756.02	00000000
10	VOLTAIRE	171	0.05	8	752.21	00000000	
11	PANGLOSS	135	0.04	0	644.22	00000000	
12	YOU	543	0.17	1,251	0.04	633.12	00000000
13	GUILT	198	0.06	84	617.59	00000000	

O a análise qualitativa dos dados em linguística de *corpus* é feita através das linhas de concordância. Linhas de concordância consistem em uma palavra, sintagma ou expressão aparecendo junto com o respectivo contexto de um texto (HUNSTON, 2010). A figura 6 mostra o resultado da busca da palavra *women* no corpus BAWE, utilizando a função *Concord*.

Figura 6: Exemplo de linhas de concordância geradas com a função *concord*



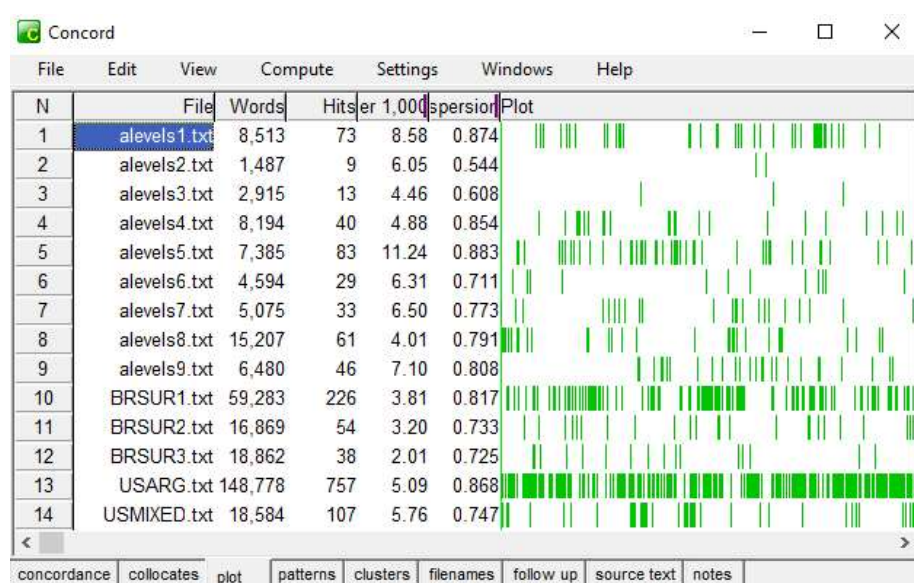
N	Concordance	File
824	by the McCarran- Walter Act, for example, and women found themselves "losing many of the	0302e.txt
825	old man who ends up begging for help from a women in order to have Alba returned to him.	0310d.txt
826	and a passive female voice in the background. Women and the women's word are the focus in	0310d.txt
827	of Nature' (Bland 1995 53). The victimisation of women was used to maintain the profession's	0314d.txt
828	. Ussher continues to note 'it was the affluent women of the upper middle class who were	0314d.txt
829	universities. Although this act did not exclude women officially from medicine, Witz notes the	0314d.txt

Os programas de concordância, incluindo o *WordSmith Tools*, ajustam a quantidade de contexto a ser exibida junto com o termo buscado, permitindo aumentar ou diminuir a quantidade nas configurações dos programas. O termo pesquisado fica centralizado em uma cor diferente, com contexto tanto do lado esquerdo quanto do lado direito do termo. Através do contexto ao redor do termo pesquisado e da exibição de todas as ocorrências de um termo seguidas em seu contexto, facilita-se a identificação de padrões gramaticais não antes observados, quando visualizados separadamente em um texto.

O resultado da busca do *Concord* também fornece muitas informações a respeito do termo pesquisado. Dentro da janela de resultados, o *WordSmith Tools* dá a possibilidade de se verificarem várias características do termo pesquisado: os *Collocates*, *Plot*, *Patterns*, *Clusters*, *Filenames*, *Source text* e *Notes*.

- Os *Collocates* são palavras que ocorrem imediatamente antes e depois do termo. O resultado mostra uma lista de palavras que ocorrem ao redor do termo pesquisado, a frequência em que ocorrem etc.
- O *Plot* mostra o resultado da busca a partir de um gráfico, em um formato de código de barras, conforme a figura 7.

Figura 7: Exemplo do recurso *plot* do *Corpus LOCNESS*



Esse recurso também permite ver a posição em que a palavra buscada está em cada um dos textos do *corpus*, mostrando, assim, a sua distribuição através dos textos.

- O *Patterns* cria uma lista com os itens lexicais que ocorrem mais vezes ao redor de uma palavra pesquisada. Essa função é parecida com a de *collocates*. Porém, em vez de mostrar em forma de lista de palavras que antecedem e sucedem um termo pesquisado, ele mostra um mapa das palavras que aparecem em todas as posições, sendo a primeira à direita, a segunda à direita, a terceira à direita, a primeira à esquerda, a segunda à esquerda e assim sucessivamente.

Figura 8 – Exemplo do recurso Patterns do corpus LOCNESS

The screenshot shows the Concord software interface with the 'Patterns' view selected. The main window displays a table of word patterns around the search term 'THE'. The table has columns for positions N, L5, L4, L3, L2, L1, Centre, R1, R2, R3, R4, and R5. The rows show the following patterns:

N	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	THE	THE	THE	OF	THE	PEOPLE	ARE	THE	THE	THE	THE
2	OF	OF	TO	TO	MANY		WHO	TO	TO	TO	TO
3	TO	A	OF	THE	OF		TO	NOT	OF	AND	AND
4	A	AND	AND	AND	THESE		IN	THAT	THAT	OF	OF
5	IN	TO	A	THAT	THAT		OF	ARE	THEIR	A	A

At the bottom of the window, there are tabs for 'concordance', 'collocates', 'plot', 'patterns', 'clusters', 'filenames', 'follow up', 'source text', and 'notes'. The 'patterns' tab is currently active.

- O *Clusters* é um mecanismo que computa automaticamente grupos de palavras com dois ou mais termos que ocorrem junto ao termo pesquisado. Com isso, a ferramenta busca, de forma automática, padrões gramaticais, lexicais, expressões idiomáticas, dentre outras possibilidades.
- O *Filenames* mostra todos os nomes dos arquivos em que o termo buscado ocorreu.
- O *Source text* mostra o contexto expandido, com o texto completo de uma ocorrência. Para se visualizar o contexto expandido, basta clicar duas vezes sobre a linha de concordância desejada.
- O *Notes* indica as configurações adotadas no *Wordsmith Tools*.

Essas funcionalidades do *Concord* descritas foram utilizadas nesta pesquisa seja para se decidirem quais os próximos passos a serem seguidos, seja para se verificar se um termo foi produzido por vários alunos etc. Quando esses recursos mostraram resultados significantes para o trabalho, o procedimento, assim como os resultados encontrados, foi descrito na seção a seguir, ou nos resultados.

4.3 Procedimentos metodológicos

Conforme mencionado anteriormente, o *corpus* BAWE é utilizado, neste trabalho, como um *corpus* de referência de escrita acadêmica para a geração das palavras-chave. Para se gerar uma lista de palavras-chave são necessários 2 *corpora* diferentes. Como esta pesquisa utiliza 2 *corpora*, um de estudo (BRICLE) e um comparativo (LOCNESS), um terceiro *corpus* (de referência), maior do que os dois citados, foi necessário para se gerar a lista de palavras-chave (*Keyword list*) no *WordSmith Tools*.

Foram selecionados apenas os ensaios do *corpus* BAWE, que possui mais de 6 milhões de palavras em sua totalidade. Essa seleção nos deixou ainda com um *corpus* consideravelmente grande (aproximadamente 3 milhões de palavras). Desta forma, utilizei como *corpus* de referência apenas os ensaios acadêmicos do BAWE, contando ainda com um *corpus* de tamanho grande.

Na documentação do *corpus*, encontra-se uma legenda em que é informado o tipo de texto de cada arquivo. Cada arquivo corresponde a um texto escrito por um aluno e o nome do arquivo corresponde a um código. Através dessa legenda criada pelos compiladores do *corpus* é possível identificar o código do aluno que escreveu o texto e o gênero de cada texto. O *corpus* selecionado possui um total de 1.284 ensaios acadêmicos, totalizando 3.120.263 palavras, o que corresponde a aproximadamente metade do *corpus* BAWE na sua totalidade.

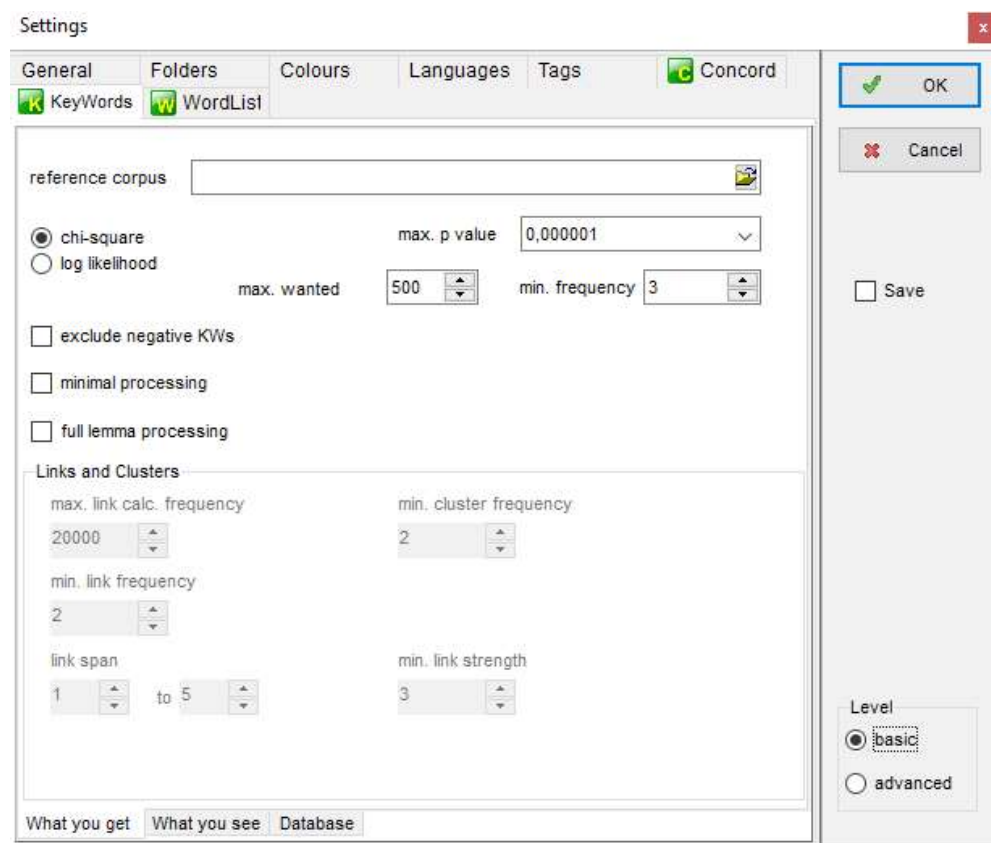
4.3.1 As palavras-chave

Como esta pesquisa consiste em analisar os modificadores de sintagmas nominais, selecionei os sintagmas nominais a serem analisados, com seus contextos, ou seja, a oração em que eles estão inseridos. Por isso, fiz a seleção dos

sintagmas nominais a partir do núcleo desses elementos. A decisão sobre quais elementos seriam selecionados foi feita a partir da função *Keywords* do *WordSmith Tools*. Decidi escolher os nomes (que são os núcleos dos SNs) a partir dessa funcionalidade, pois, diferentemente da *WordList*, a *Keyword List* traz uma lista que consiste em palavras geralmente lexicais.

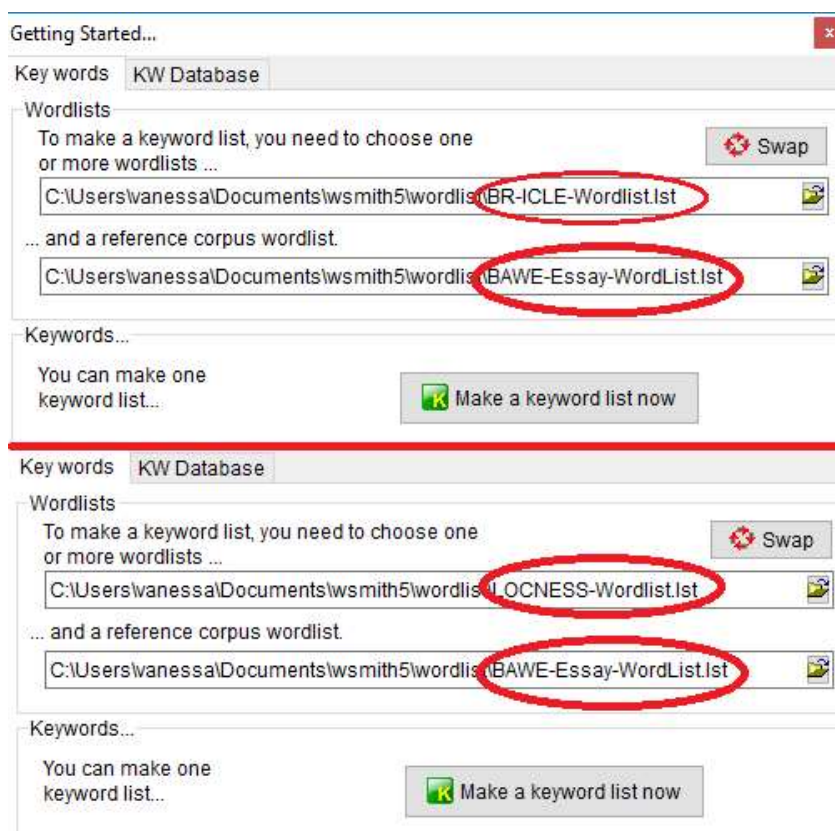
As configurações utilizadas são mostradas na figura 9. O teste estatístico utilizado foi o chi-quadrado, com máximo de p value de 0,000001. O campo *max. wanted* indica o número máximo de palavras-chave desejado e o campo *min. frequency* indica a frequência mínima para uma palavra ser considerada chave, e mantive os valores padrão (sugerido pelo *WordSmith*) de 500 para o primeiro e 3 para o segundo.

Figura 9: Configurações do *WordSmith* para geração de palavras chave



Conforme mencionado anteriormente, a função da *Keyword list* é criar uma lista de palavras-chave de um texto. A partir de uma lista de palavras de um *corpus* grande de referência e da lista de palavras de um *corpus* de estudo, o WordSmith calcula (a partir de um cálculo estatístico e a partir da frequência das palavras nos dois *corpora*) o quão chave uma palavra é. Desta forma, ele deixa de fora palavras que são igualmente frequentes em ambos os *corpora*. Com esse cálculo, o programa desconsidera, principalmente, palavras gramaticais: artigos (como *the*, *a*, *some*), preposições (como *in*, *on*, *from*) e partículas (como *to*).

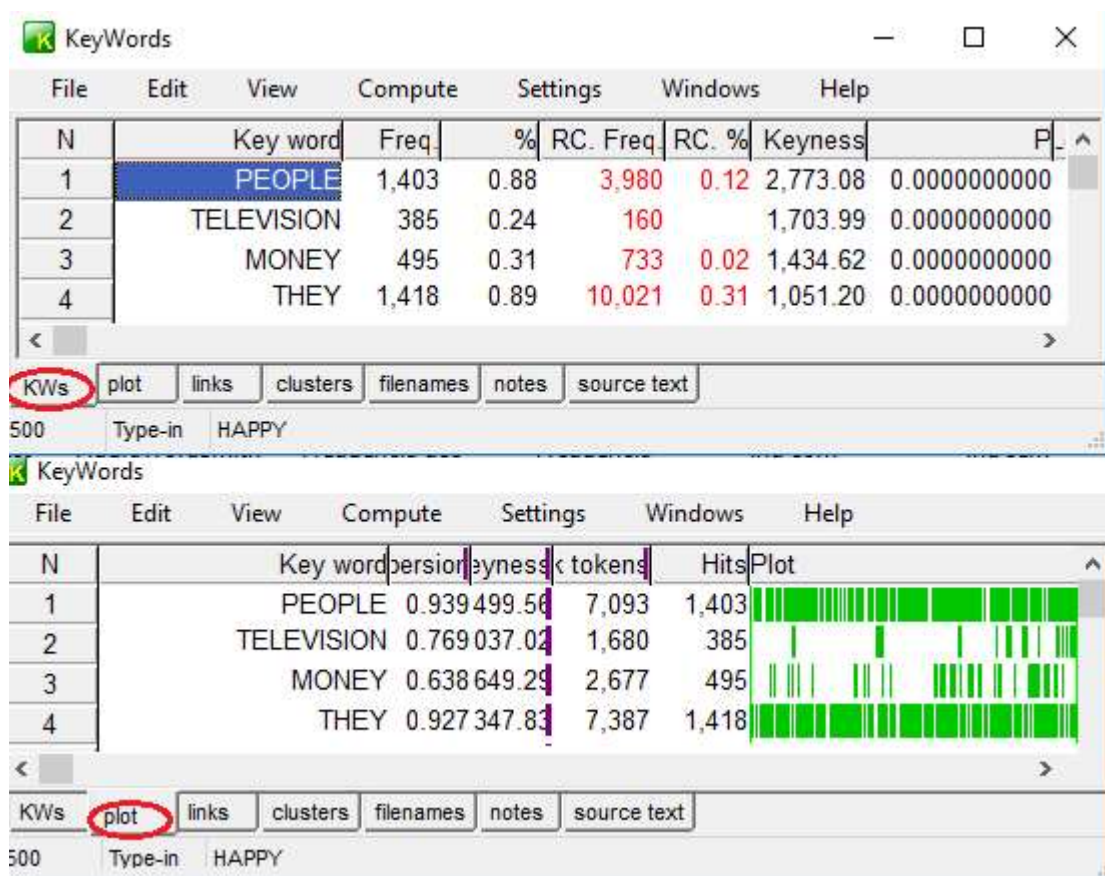
Figura 10: Geração das listas de palavras-chave



Foram geradas duas listas de palavras-chave: uma, utilizando-se o *corpus* de aprendizes brasileiros BR-ICLE; e outra, utilizando-se o *corpus* de falantes nativos do inglês LOCNESS. Em ambas as listas, utilizei o BAWE como *corpus* de referência, conforme figura 10.

Ao seguir esses procedimentos, pude selecionar os termos para esta pesquisa com base nos mesmos parâmetros, permitindo uma comparação confiável entre os resultados.

Figura 11 – Exemplo de lista de palavras-chave do BR-ICLE mostrando o Plot



Conforme a figura 11, o recurso *Keywords* mostra, além das palavras com seus respectivos valores, o *Plot* das palavras. Assim, é possível ver a distribuição do termo pelo *corpus*. Com a lista pronta, os resultados das listas de palavras-chave tanto do BR-ICLE quanto do LOCNESS foram comparados. Isso permitiu que uma ampla gama de informações pudesse ser visualizada, conforme a figura 11, em apenas uma lista de palavras, a saber:

- uma lista de palavras lexicais do *corpus* (diferentemente da lista de palavras gerada pela *WordList*, que traz uma lista de todas as palavras, lexicais e gramaticais);

- a frequência do termo no *corpus* de estudo;
- e a frequência do termo no *corpus* de referência;
- a porcentagem que representa a frequência de ocorrência do termo no *corpus* de referência;
- a porcentagem que representa a frequência de ocorrência do termo no *corpus* de estudo;
- a *Keyness* do termo;
- o *Plot* do termo.

Diante dessas informações, foi possível escolher os núcleos dos SNs a serem investigados no trabalho. Os critérios para a escolha dos termos foram as seguintes:

- Os termos deviam aparecer tanto na lista de palavras-chave do BR-ICLE quanto na do LOCNESS.
- Os termos deviam ter uma *keyness* positiva (caso seja negativa, quer dizer que o termo foi mais frequente no *corpus* de referência do que no de estudo).
- Os termos deviam ter ocorrido de forma homogênea (distribuição mostrada pela função *plot*) no BR-ICLE e no LOCNESS.

Seguindo esses passos, busquei evitar que os problemas encontrados no meu trabalho de mestrado, Wright (2014), ocorressem novamente. Wright (2014) selecionou os termos mais frequentes, utilizando a lista de frequência. Desta forma, ela selecionou termos diferentes nos *corpora* de nativos e de aprendizes, ocasionando problemas diferentes.

A primeira limitação do trabalho de Wright (2014) foi em relação à possível contaminação dos dados. Como os ensaios possuem tópicos ou enunciados para a sua elaboração, alguns dos termos pesquisados só eram frequentes em decorrência desses enunciados, já que os alunos repetiam palavras do enunciado na elaboração do texto. Embora esse problema não possa ser resolvido, já que não há como evitar que palavras apareçam no enunciado, é possível tentar contornar essa situação analisando os SNs cujo núcleo é realizado pelos mesmos termos nos dois *corpora*.

A segunda limitação daquele trabalho foi, também, em relação à metodologia adotada para a seleção de dados. Os núcleos de SN selecionados em cada *corpus*

não eram termos comuns aos dois *corpora*, pois a seleção desses núcleos foi feita com base na maior frequência de ocorrência em cada *corpus*. Embora esse detalhe não fosse um problema para o trabalho, isso dificultou o alcance de algumas conclusões. Por exemplo, notou-se que os determinantes e modificadores variam, também, de acordo com o tipo de núcleo de SN. No estudo citado, a palavra *society* foi o termo mais frequente no *corpus* de aprendizes, enquanto o termo *life* foi mais frequente no *corpus* de nativos. Embora ambos os substantivos tenham classificação semelhante (ambos são contáveis e abstratos), a ocorrência de determinantes que ocorreram nos SNs pesquisados apresentou diferenças. O determinante possessivo *our* foi produzido com uma frequência 5 vezes maior acompanhado do núcleo *society* do que acompanhado do núcleo *life*. E essa diferença ocorreu devido aos termos pesquisados e não devido à comparação *corpus* de aprendiz *versus corpus* de nativo. Em ambos os *corpora*, *our* é um determinante que ocorre com mais frequência junto de *society*. E o fato de SNs com núcleo *life* ter sido analisado, e não *society*, como no *corpus* de aprendizes, fez com que a frequência final do determinante *our* fosse mais alta entre os aprendizes.

Em outros trabalhos, como o de Parkinson e Musgrave (2014), os autores atentaram para o fato de a diferença nos tópicos abordados em cada *corpus* influenciar nos resultados. E essa diferença, para eles, pode ter influenciado na quantidade e qualidade (tipo) de sintagmas nominais produzidos. Justamente para se evitar essa possível influência, os sintagmas nominais com os mesmos núcleos foram analisados em ambos os *corpora* no presente estudo.

Seguindo passos anteriormente descritos, pretendo alcançar os resultados esperados para esta pesquisa, indicando as características dos modificadores produzidos em ensaios acadêmicos de aprendizes brasileiros de inglês.

4.3.2 A busca por termos que sejam núcleos de SN

Como trabalhei com um número grande de palavras em mais de um *corpus* (cada *corpus* possui mais de 100 mil palavras) e busquei, ao mesmo tempo, estudar um número grande de dados quantitativa e qualitativamente, optei por selecionar uma amostra composta por SNs com núcleos pré-estabelecidos e analisá-los, a fim

de encontrar padrões e características da escrita dos aprendizes referente aos modificadores de SN.

Optei por analisar uma amostra de sintagmas nominais em vez de usar *corpora* anotados, pois, para se realizar uma análise qualitativa, é necessário fazê-lo de forma manual, caso a caso. E seria um trabalho inimaginável analisar todos os SNs dos *corpora* de estudo e de referência, manualmente.

Outro empecilho para se fazer o estudo em um *corpus* anotado é o próprio processo de anotação. Em um estudo comparativo diacrônico do inglês americano com o inglês britânico, Biber et al. (2009) analisam os modificadores de sintagmas nominais de um *corpus* de reportagens de jornal de um período de 4 séculos, no *corpus* ARCHER. Mesmo fazendo uma análise linguística baseada em textos anotados, os autores tiveram de abrir exceções na seleção dos dados, pois não foi possível identificar todos os tipos de modificadores com base nas anotações. Eles encontraram esse problema principalmente com os sintagmas preposicionados com função de pós-modificador, afirmando não haver “nenhum método automático que pudesse distinguir, de forma precisa e confiável, sintagmas preposicionados com função de adverbiais daqueles com função de modificadores de nomes”²⁷ (BIBER et al., 2009, p. 186). Vale lembrar que, nesse estudo de Biber et al. (2009), foi feita uma análise quantitativa apenas.

Granger (2002) também aponta problemas quanto à anotação de *corpus* de aprendiz. Segundo a autora, um anotador automático de *corpus* se baseia em *corpora* de língua natural (aquela produzida por falantes nativos). Por isso, quanto menor o domínio da língua do aprendiz, menor o nível de precisão da anotação.

Experimentos mostraram que se o a produção do aprendiz é avançada, com uma baixa proporção de erros ortográficos e morfológicos, a taxa de sucesso do anotador seria similar ao obtido quando anotando dados de falantes nativos. Porém, quanto mais os dados desviam da norma, menos precisa será a anotação, ao ponto de tornar o uso do anotador impraticável.²⁸ (GRANGER, 2002, p. 18).

27 No original: “There are no automatic methods that reliably and accurately distinguish between prepositional phrases functioning as adverbials and those functioning as noun modifiers”.

28 No original: “Experiments have shown that if the learner output is quite advanced, with a low proportion of spelling and morphological errors, the success rate of the tagger is similar to that obtained when tagging NS data. But the more deviant the data, the less accurate the tagging will be, to the point of making the use of the tagger impracticable”.

Ou seja, o processo de anotação de *corpus* de aprendiz é aconselhável quando apresenta uma escrita próxima da utilizada pelos nativos, com poucos erros de norma e ortografia.

4.3.2.1 A seleção dos sintagmas nominais

Tendo os objetivos da pesquisa em vista, decidi selecionar uma amostra de SNs a partir de seu núcleo. A partir da lista de palavras-chave geradas com o BR-ICLE e com o LOCNESS, selecionei os 10 substantivos com maior *keyness* e que possuísssem frequência significativa para o *corpus* de referência.

Para chegar a uma quantidade significativa de SNs modificados a serem analisados (os 10 substantivos mais frequentes, somados aos nomes verbais), baseei-me em dois dados. O primeiro foi obtido a partir da leitura de pesquisas de outros autores da área, como Parkinson e Musgrave (2011), que indicaram que aprendizes de inglês, falantes maternos de outras línguas diferentes do português, apresentaram aproximadamente 70% de SNs não modificados. O segundo foi a partir de uma análise de amostras de ensaios do *corpus* pesquisado, verificando que o número de SNs que recebem algum tipo de modificação ficava em torno de 30 a 35%.

Tendo em vista que o objeto de estudo desta pesquisa são os modificadores e que há um número enorme de elementos diferentes que podem realizar a função de modificador, selecionei um número de SNs maior do que o estipulado estatisticamente. Ao selecionar 10 substantivos a partir da *Keyword List*, haveria SNs suficientes para encontrar pelo menos 1000 SNs com algum tipo de modificação. Os 10 termos que satisfizeram os critérios estabelecidos foram os seguintes, apresentados no quadro 4:

Quadro 4 - Os 10 nomes mais chave e frequentes em comum na lista de palavras-chave do LOCNESS e do BR-ICLE

1	<i>people</i>	6	<i>life</i>
2	<i>money</i>	7	<i>job</i>
3	<i>problems</i>	8	<i>person</i>
4	<i>things</i>	9	<i>children</i>
5	<i>students</i>	10	<i>television</i>

A partir dessa lista de palavras, selecionei as linhas de concordância de todas as ocorrências desses nomes no *corpus* de aprendizes e no *corpus* de falantes nativos. Após a seleção das linhas de concordância, fiz uma limpeza nos resultados para analisar apenas os sintagmas nominais em que esses termos ocorrem como núcleo.

4.3.2.2 A seleção dos núcleos de sintagma nominal realizados por nomes verbais (-ing)

Os SNs cujos núcleos são realizados pelos nomes verbais foram selecionados de forma diferente dos núcleos realizados por substantivos. Conforme discutido no capítulo 1, essa categoria de nomes é formada por verbos no gerúndio. Para analisar apenas os nomes verbais que se encaixam perfeitamente na categoria de SN, decidi estudar os casos em que os núcleos *-ing* ocorrem em um SN com determinante. Desta forma, os SN, como do exemplo (81), entraram na análise dos dados.

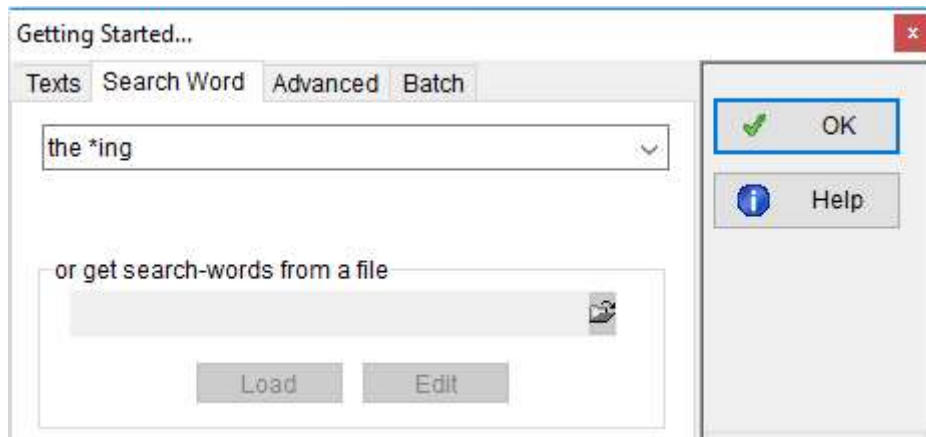
(81) *His **shrinking** is thus symbolic of a broader personal turmoil.* (BAWE)

No exemplo (81), observa-se um caso de SN com determinante *his*, cujo núcleo *shrinking* é um nome verbal.

Nos SNs cujos núcleos são realizados pelos nomes verbais temos palavras muito comuns, com origem verbal, que já foram dicionarizadas como substantivos, que ocorrem com frequência nos *corpora* pesquisados: *beginning*, *feeling*, *meeting*, dentre outras. Essas palavras, desde que precedidas de determinantes, serão consideradas para efeito de minha análise.

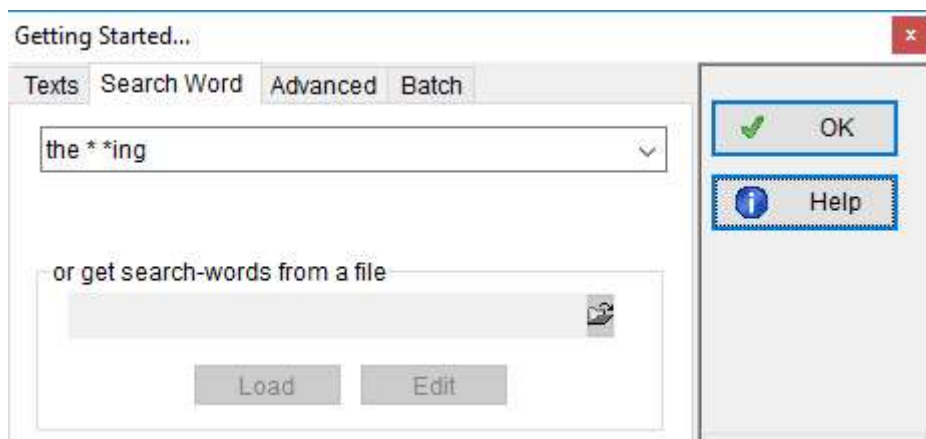
Para facilitar a busca de dados, já que considerei apenas os casos de nomes verbais com determinantes, a seleção desses nomes verbais foi feita a partir da busca dos termos, utilizando a funcionalidade *Concord* da ferramenta *WordSmith Tools*, seguindo dois passos diferentes. Primeiro, foram selecionados os nomes verbais que não possuem pré-modificação. Essa busca foi feita a partir da seleção de um determinante seguido por verbo na forma *-ing*, conforme figura 12.

Figura 12: Seleção dos SNs com núcleo *-ing* sem modificação utilizando *Concord*



Em seguida, foram selecionados os nomes verbais que possuem 1 pré-modificador. Essa busca foi feita a partir da seleção de um determinante seguido por qualquer termo, que foi seguido por termo na forma *-ing*, conforme figura 16.

Figura 13: Seleção dos SNs com núcleo *-ing* com modificação utilizando *Concord*



Procurei por SNs com núcleo *-ing* que possuísem mais de um elemento pré-modificador, porém não obtive nenhum resultado. A busca descrita foi realizada com os seguintes determinantes mostrados no quadro 6:

Quadro 5: Determinantes pesquisados

<i>a</i>	<i>an</i>	<i>another</i>	<i>any</i>
<i>each</i>	<i>few</i>	<i>more</i>	<i>some</i>
<i>that</i>	<i>the</i>	<i>this</i>	<i>those</i>
<i>own</i>			

O primeiro comando trouxe, como resultado de busca, todas as linhas de concordância que possuem a estrutura [determinante] + [termos terminados em *-ing*]. Após essa seleção, foi necessário fazer uma limpeza dos dados, uma vez que há termos que não representam a estrutura desejada. Observe os exemplos de (82) a (85):

(82) *I will briefly mention previous reforms from the beginning of the Fifth Republic.* (BAWE)

(83) *This is reason enough for the banning of boxing.* (BAWE)

(84) *These days this is becoming increasingly important in science.* (BAWE)

(85) *With this excellent reasoning they use to support their incredible claims.* (BAWE)

Durante a seleção de dados, SNs como os dos exemplos (82) e (83) entraram na análise. Isso ocorreu porque a seleção foi feita a partir da busca de determinante seguido por qualquer palavra terminada em *-ing*. Os dois exemplos foram mantidos, pois constituem um SN.

Obtive, também, SNs como os dos exemplos (84) e (85), fazendo uma seleção a partir da busca de determinante seguido por um termo qualquer seguido de palavra terminada em *-ing*. Nesse caso, excluí os resultados como o exemplo (84) por não constituir um SN cujo núcleo é um nome verbal, e mantive o exemplo (85) por ser um SN cujo núcleo é um nome verbal.

Com base na classificação descrita acima, os resultados encontrados no BR-ICLE e no LOCNESS são comparados, descritos e analisados no capítulo seguinte. Os núcleos *-ing* analisados podem ser vistos nos anexos 3 e 4.

4.3.2.3 A seleção dos sintagmas nominais realizados por pronomes

Como os pronomes geralmente não são modificados, eles foram analisados quantitativamente como um caso de SN de complexidade 0. A frequência do uso dos

pronomes pessoais e indefinidos por parte dos aprendizes foi comparada com a frequência por parte dos nativos.

Utilizando a função *concord* do *WordSmith Tools*, todos os pronomes pessoais (sujeito e objeto) foram pesquisados, e sua frequência nos *corpora* foi contabilizada. Os pronomes pesquisados foram:

- Pessoais (sujeito): *I, you, he, she, it, we* e *they*.
- Pessoais (objeto): *me, you, him, her, it, us* e *them*.

Os pronomes impessoais, indefinidos, demonstrativos e interrogativos não foram incluídos nesta pesquisa porque os pronomes mais frequentes são os pessoais. E considero que a seleção escolhida foi suficiente para a análise quantitativa.

4.3.3 Proposta de classificação das complexidades do sintagma nominal

Feita a limpeza dos dados, os sintagmas nominais foram classificados de acordo com sua complexidade. Quanto maior o número de elementos em um SN, maior sua complexidade. Por isso, foi dado um valor para cada tipo/quantidade de modificação encontrada no SN conforme Quadro 5:

Quadro 6: Complexidade do sintagma nominal

Tipo de SN	Complexidade
(det) núcleo	0
(det) pre-mod núcleo	1
(det) núcleo pos-mod	2
(det) pre-mod núcleo pos-mod	3

Optei por dar valores diferentes aos tipos de modificadores que um SN possui. Os casos em que o núcleo de SN ocorre sozinho ou com um determinante são considerados como de complexidade 0.

Os SNs com valor 1 são aqueles que possuem um ou mais pré-modificadores. Esses pré-modificadores são, geralmente, realizados por adjetivos, nomes ou verbos na forma de particípio ou gerúndio. Esses SNs podem ocorrer com ou sem determinantes.

Já os SNs com valor 2 são aqueles que possuem um ou mais pós-modificadores. Eles foram classificados como mais complexos do que os SNs com pré-modificadores por causa da natureza de sua estrutura. Os pós-modificadores são realizados, geralmente, por sintagmas preposicionados e orações. Esses dois tipos de unidades são considerados mais complexos por causa de sua estrutura. Os elementos obrigatórios dos SPreps são a preposição e um complemento, que pode ser um SN ou uma oração. As orações são muitas vezes formadas por vários elementos, incluindo sujeito, verbo e objeto. Já os SNs são elementos que podem conter apenas um núcleo ou ocorrer com vários outros elementos. A estrutura do pós-modificador do SN pode fazer com que ele tenha inúmeros elementos encaixados sucessivamente (BIBER, POONPON, GRAY, 2011). Os SNs pós-modificados podem ocorrer com ou sem a presença de determinantes. Os complementos de substantivos também recebem classificação de complexidade 2, visto que sua estrutura sintática é idêntica à estrutura dos pós-modificadores.

Por fim, os SNs com valor 3 são aqueles que possuem tanto pré-modificadores quanto pós-modificadores. A presença desses dois elementos dentro do SN faz com que ele possua uma estrutura com mais elementos encaixados, tornando esse SN mais complexo do que os outros casos anteriormente citados. Vale lembrar que esses SNs também podem ocorrer com ou sem a presença de determinantes.

De acordo com o critério de classificação proposto no presente trabalho, os SNs teriam a seguinte classificação.

(70) *Additionally, **people** are encouraged to take a greater responsibility.* (BAWE)

O exemplo (70) possui o SN *people*, que é realizado apenas pelo núcleo, sem qualquer tipo de modificador. Ele é classificado como de complexidade 0.

(71) *The desire to avoid HIV can be used to motivate **sexually active people** into practicing safe sex.* (BAWE)

O exemplo (71) possui o SN *sexually active people*, que é realizado pelo núcleo *people*, possui um pré-modificador, *sexually active*, e não possui um pós-modificador. Esse SN é classificado como de complexidade 1.

(72) *Yet it will always be fought for by **people who believe in autonomy and choice**.* (BAWE)

O exemplo (72) possui o SN *people who believe in autonomy and choice*, que é realizado pelo núcleo *people*, possui um pós-modificador, *who believe in autonomy and choice*, e não possui um pré-modificador. Esse SN é classificado como de complexidade 2.

(73) *It is necessary to investigate subjects who have no previous experience of certain sensations, such as **blind people who have no experience of visual stimuli**.* (BAWE)

O exemplo (73) possui o SN *blind people who have no experience of visual stimuli*, que é realizado pelo núcleo *people*, possui um pré-modificador, *blind*, e também um pós-modificador, *who have no experience of visual stimuli*. Esse SN é classificado como de complexidade 3.

Como alguns SNs são muito grandes, não foi possível determinar seu início e/ou seu fim apenas pela linha de concordância. Para esses casos, foi feita uma busca individual do SN específico para ser analisado dentro do seu contexto expandido.

4.3.4 A realização dos modificadores de sintagma nominal

Na segunda parte da análise dos dados, os modificadores dos SNs foram classificados. Essa classificação foi feita manualmente, de acordo com a realização do elemento modificador, podendo ser esse elemento um adjetivo atributivo ou um particípio, um nome, uma oração finita ou não finita ou um sintagma preposicionado, conforme exemplos 74 a 80. Os pré-modificadores foram classificados como:

- adjetivo atributivo:

(74) ***depressive** people may well ignore the fact that they have things in their lives which they should feel are important and beneficial.* (BAWE)

- adjetivo particípio:

(75) *the numbers of **deceased** people who remained on the register balanced out.* (BAWE)

- substantivo:

(76) *A major difference, therefore, is the absence of authentic, generative situations in the **grammar** book.* (BAWE)

E os pós-modificadores foram classificados como:

- oração finita:

(77) *the numbers of deceased people **who remained on the register** balanced out.* (BAWE)

- oração não finita:

(78) *People **using computers** will not be bound to a single location.* (BAWE)

- sintagmas preposicionados:

(79) *people **from different social classes** choose different words to represent a certain thing.* (BAWE)

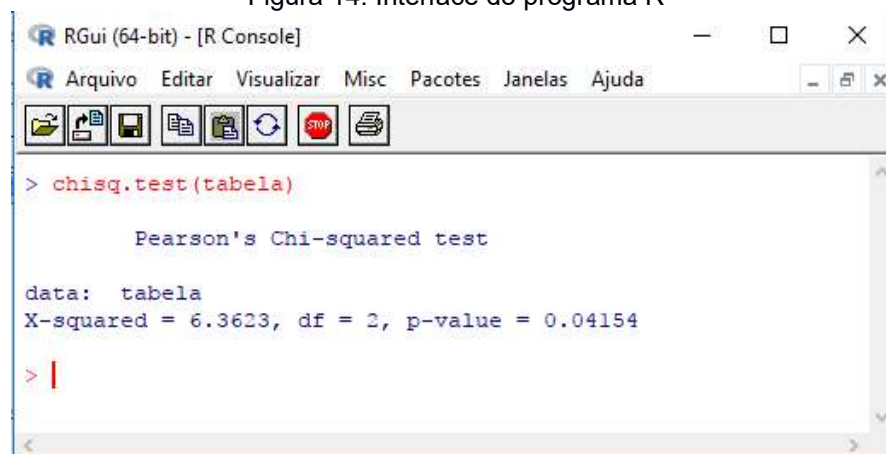
- sintagmas nominais (apositivo):

(80) *The Bible says that God made Adam, **the man**, before he made Eve, **the woman**.* (LOCNESS)

4.3.5 Teste estatístico e normalização dos dados

O presente estudo utiliza testes estatísticos de independência como um complemento para a descrição dos SNs produzidos pelos aprendizes brasileiros. Para fazer os cálculos estatísticos, utilizei o programa R (versão 3.3.3), um *software* livre utilizado para computações estatísticas e gráficos (SILVA, DINIZ, BORTOLUZZI, 2009).

Figura 14: Interface do programa R



O programa R pode ser adquirido via *download* no *site* do projeto <https://www.r-project.org>. O *software* possui uma interface parecida com a de um editor de texto, e comandos com a linguagem de programação R devem ser executados para se realizarem os cálculos, conforme exemplo na figura.

O teste estatístico aplicado no presente estudo foi o teste Qui-quadrado de Pearson. Esse teste indica, através do valor p , que um intervalo de dados possui diferenças significativas quando o valor de p é menor do que o escolhido como parâmetro. Defini o valor 0,05 para o nível de significância, o que significa que uma margem de erro de 5% é tolerada. Um valor de p menor do que o nível de significância (0,05) indica que os valores das linhas e colunas são dependentes. Dessa forma, pode-se concluir que há relação entre um tipo específico de modificador e quem o produziu (falante nativo ou aprendiz), o que significa que as diferenças podem ser consideradas estatisticamente significativas. Já um valor de p maior do que o nível de significância indica que os valores são independentes, o que significa que as diferenças não são consideradas estatisticamente significantes. Ou, seja, quanto menor o valor de p , mais próximo de 0 (zero), mais dependentes são os valores e mais significativas são as diferenças.

Para se calcularem as diferenças estatísticas entre as complexidades dos SNs produzidos pelos aprendizes e pelos nativos, os valores foram todos normalizados por 10 mil palavras. Como o número total de SNs dos *corpora* são desconhecidos, a partir de uma amostra de ensaios de cada *corpus*, estipulei a quantidade total de SNs realizados por um núcleo lexical para cada *corpus*.

Como trabalhei com dois *corpora* de tamanhos diferentes, para comparar a ocorrência de algum termo no *corpus* trabalhei com os dados normalizados (EVISOL, 2010). Por convenção, os dados são geralmente normalizados por mil, dez mil, cem mil ou um milhão de palavras. Optei por normalizar as ocorrências por 10 mil palavras. Essa normalização foi feita por média aritmética, seguindo a seguinte fórmula:

$$\frac{\text{Número de ocorrências total no corpus}}{\text{Tamanho do corpus}} \times 10.000 = \text{Ocorrência por 10 mil palavras}$$

Os resultados encontrados serão descritos no próximo capítulo.

DESCRIÇÃO E DISCUSSÃO DOS RESULTADOS

Neste capítulo, os resultados encontrados durante a análise dos dados serão descritos e discutidos. O capítulo é dividido em 6 partes. A primeira parte é reservada para os resultados quantitativos, de forma mais geral. Já a segunda contém os resultados encontrados no que se refere aos pré-modificadores. A terceira parte é dedicada à discussão dos dados encontrados no que se refere aos pós-modificadores. Já a quarta é reservada à descrição dos resultados referentes ao pronome. Na quinta parte são descritos os resultados obtidos na análise dos complementos de preposição. E, finalmente, na sexta parte, é feita uma discussão dos resultados descritos.

Como mencionado no capítulo inicial, o objetivo do trabalho consiste em descrever os sintagmas nominais produzidos por aprendizes brasileiros, em escrita acadêmica, sendo o objeto de estudo específico os modificadores de SN, através da análise qualitativa e quantitativa dos tipos de pré-modificadores e dos pós-modificadores de SN produzidos por falantes nativos e aprendizes. Isso foi feito através da investigação do tamanho e da variedade lexical, ou seja, da complexidade estrutural do SN.

A seguir, são descritos os resultados gerais encontrados neste trabalho.

5.1 Resultados gerais

Conforme descrito no capítulo 2, os núcleos dos sintagmas nominais analisados foram selecionados a partir da lista de palavras-chave da ferramenta *WordSmith Tools*, que dependia também da lista de palavras de ambos os *corpora* analisados. A lista de palavras já exibe algumas informações importantes sobre o

corpus: o número total de palavras e o número total de tipos de palavras. A primeira indica o total de ocorrências de todos os termos de um *corpus*, enquanto a segunda indica o total de palavras diferentes do *corpus*.

Tabela 2: Total de palavras dos corpora

	Número de palavras	Tipos de palavras
Aprendiz	161.395	9.544
Nativo	326.838	16.185

Para gerar a lista de palavras-chave, a ferramenta leva em consideração esses dois valores, além de efetuar cálculos estatísticos previamente selecionados. A partir da lista de palavras-chave, foram selecionadas as 10 palavras mais frequentes que ocorreram em ambas as listas de *keyword list* (BR-ICLE e LOCNESS). Além disso, foi selecionada uma amostra dos SNs cujo núcleo era realizado por um nome verbal. O somatório dos 10 substantivos realizados como núcleo mais os núcleos realizados por *-ing* deu um total de 4.189 SNs no *corpus* de aprendizes e 4.559 SNs no *corpus* de nativos, conforme tabela 3.

Tabela 3: Frequência dos termos pesquisados por dez mil palavras

		Corpus aprendizes			Corpus nativos		
		Freq. Bruta	%	Freq. Norm.	Freq. Bruta	%	Freq. Norm.
1	People	1403	0,88	85,83	1569	0,48	48,01
2	Money	495	0,31	30,28	400	0,12	12,24
3	Life	408	0,26	24,96	677	0,2	20,71
4	Television	385	0,24	23,55	154	0,05	4,71
5	Students	278	0,17	17,01	237	0,07	7,25
6	Things	275	0,17	16,82	165	0,05	5,05
7	Children	212	0,13	12,97	382	0,12	11,69
8	Job	199	0,13	12,17	141	0,04	4,31
9	Person	195	0,12	11,93	265	0,08	8,11
10	Problems	160	0,1	9,79	220	0,07	6,73
11	Núcleo -ing	179	0,11	10,95	349	0,11	10,68

Após a limpeza dos dados que resultou em apenas os núcleos de SNs realizados por essas palavras, restou um total de 4.151 SNs de aprendizes e 4.529

SNs de nativos. A tabela 3 mostra a frequência bruta de cada termo pesquisado, percentual da frequência correspondente em relação ao tamanho do *corpus* e a frequência normalizada. Desse total de SNs analisados em cada *corpus*, 1.150 SNs apresentaram algum tipo de modificação no *corpus* de aprendiz e 1.503 SNs apresentaram algum tipo de modificação no *corpus* de nativos.

Tabela 4: Total de SNs analisados e modificados

Corpus	Total de SNs analisados	Total de SNs modificados
Aprendiz	4.151	1.071
Nativo	4.529	1.447

Conforme descrito no capítulo 4, os SNs selecionados do *corpus* de aprendiz e do *corpus* comparativo receberam classificação de acordo com o tipo de complexidade (complexidade 0, sem modificação; complexidade 1, apenas com pré-modificação; complexidade 2, apenas com pós-modificação, e complexidade 3, com pré- e pós-modificação ao mesmo tempo). Essas categorias estão, a seguir, exemplificadas através de itens dos dois *corpora* que estão sendo comparados neste estudo.

- Complexidade 0:

(86) *some students after graduate, does not get **a job** on the grounds that little companies offer positions to this professional.* (BRICLE)

(87) *she has been on welfare for nine years and lost half of her 517 dollar welfare payment for refusing to look for **a job**.* (LOCNESS)

- Complexidade 1:

(88) *students spend their time, money and energy in a four, five or even six-year course in order to get **a good job**.* (BRICLE)

(89) *they constantly bring up the point that coal mining is **a very dangerous job**.* (LOCNESS).

- Complexidade 2:

(90) *harsh, biting positioning and attitudes of these women have been regarded as a step back for **the building of a relationship of equality among sexes**.* (BRICLE)

(91) *the government should do this by investing more money into public transport than into road improvements, such as **the widening of motorways**.* (LOCNESS)

- Complexidade 3:

(92) *there must be more questions about the other concepts, so as clarify **the true meaning of the statement***. (BRICLE)

(93) *the grounding of the Exxon Valdez was a catastrophe beyond belief, not only for the wildlife which calls the sound its home, but for **the native people of Alaska** as well*. (LOCNESS)

Os SNs de complexidade 0 apresentados, exemplos (86) e (87), não possuem modificação, possuindo apenas o núcleo *job* e o artigo indefinido *a* como determinante. Em seguida, temos dois exemplos de SN com complexidade 1, exemplos (88) e (89), que possuem como elementos o núcleo *job*, o determinante *a* e o pré-modificador *good*, no primeiro exemplo, e *very dangerous*, no segundo; dois exemplos de SN com complexidade 2 são mostrados em seguida: exemplos (90) e (91). O primeiro possui o núcleo *building*, o segundo *widening* e ambos são pós-modificados por um SPrep, sendo *of a relationship of equality among sexes* no primeiro exemplo e *of motorways* no segundo. Ambos os SNs são determinados pelo artigo definido *the*. Na sequência, exemplos (92) e (93), vêm dois exemplos de SN de complexidade 3, em que o núcleo *meaning* é pré-modificado pelo adjetivo participípio *true* e pós-modificado pela oração relativa *of the statement*, no exemplo (94); e o núcleo *people* é pré-modificado pelo adjetivo *native* e pós-modificado pelo SPrep *of Alaska*.

Como já era esperado, a complexidade 0 (sem modificação) foi a mais frequente em ambos os *corpora*, aproximadamente 70%. Os SNs com algum tipo de modificação representaram aproximadamente 30% do total de SNs produzidos.

Tabela 5: Frequência das complexidades por 10 mil palavras

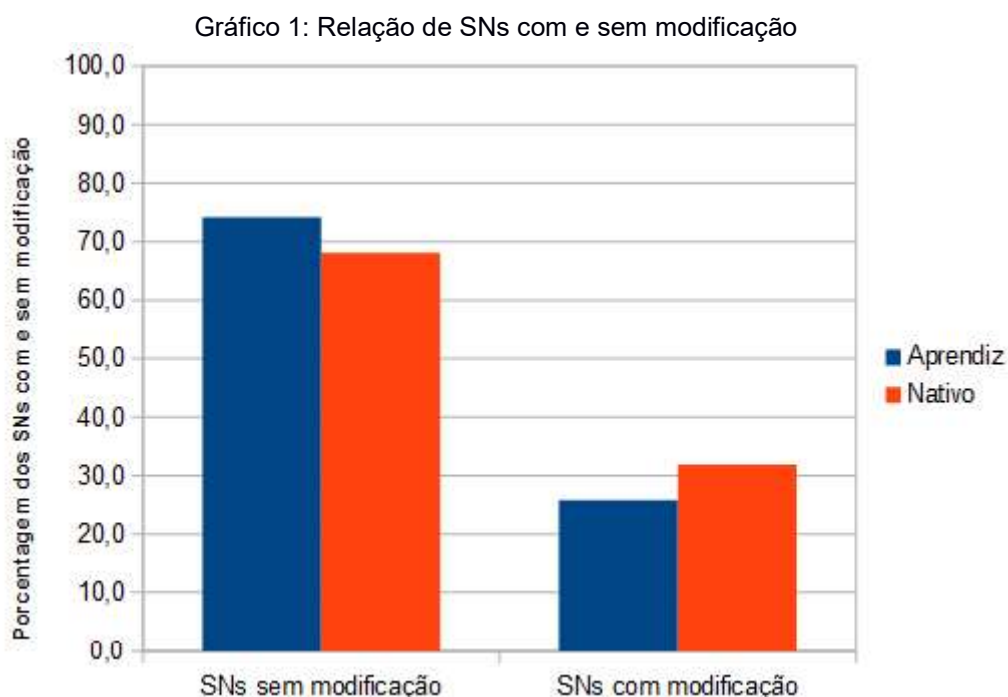
	Aprendiz		Nativo	
	Freq. Bruta	Freq. Norm.	Freq. Bruta	Freq. Norm.
Complexidade 0	3046	1.519	3029	1.605
Complexidade 1	539	269	539	286
Complexidade 2	486	242	906	480
Complexidade 3	80	40	94	50

Teste Qui-quadrado = 59,21 e valor $p < 0,00001$

O teste estatístico aplicado indica um valor de p menor que 0,00001. Esse valor indica que há relação entre os tipos de complexidade produzidos e o tipo de

corpus (nativo ou aprendiz). Isso significa que há diferença estatisticamente relevante entre as complexidades analisadas, pois o valor de p é menor que 0,05.

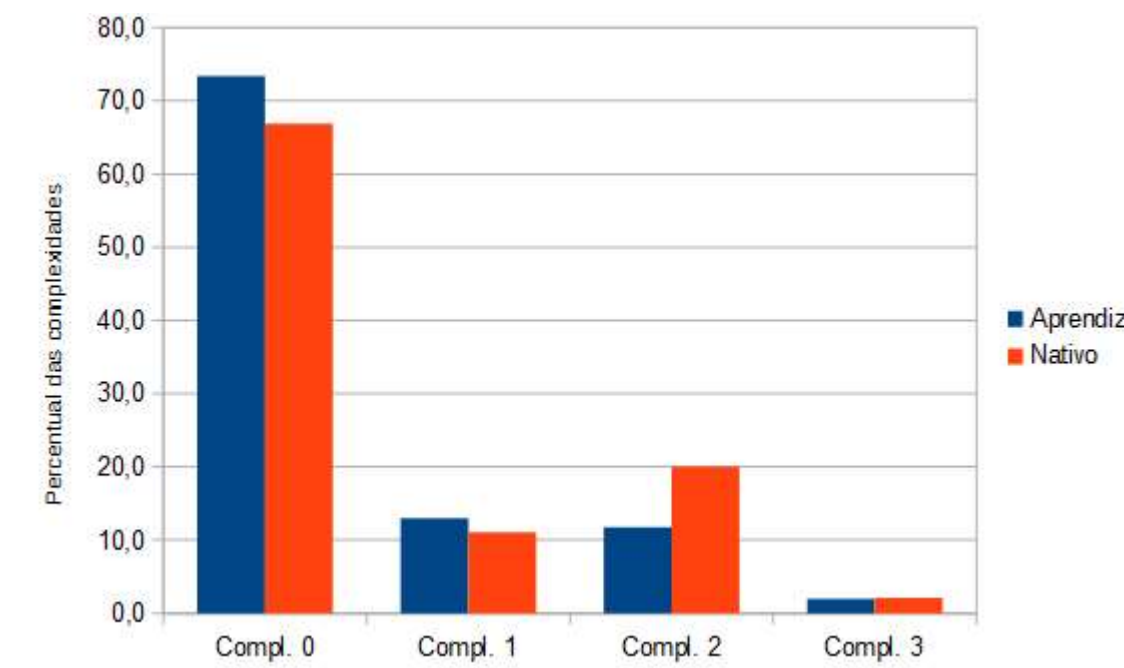
Em geral, foi constatado que aproximadamente 74,2% e 68,1% dos SNs, respectivamente produzidos por aprendizes e por nativos, não possuem modificação, de acordo com Gráfico 1:



O Gráfico 1 mostra o percentual de SNs com e sem algum tipo de modificação em relação ao número total de SNs analisados. Os aprendizes produziram menos SNs com modificação (25,8%, contra 31,9% produzidos por nativos) e mais SNs sem modificação (74,2%, contra 68,1% produzidos por nativos). O menor número de SNs modificados e o maior número de SNs de complexidade 0 no *corpus* de estudo 0 já é um indicativo de que os aprendizes brasileiros apresentam uma escrita mais simples.

O Gráfico 2 mostra a proporção dos SNs produzidos pelos nativos e pelos aprendizes, separados por tipo de complexidade.

Gráfico 2: Relação dos SNs e sua complexidade

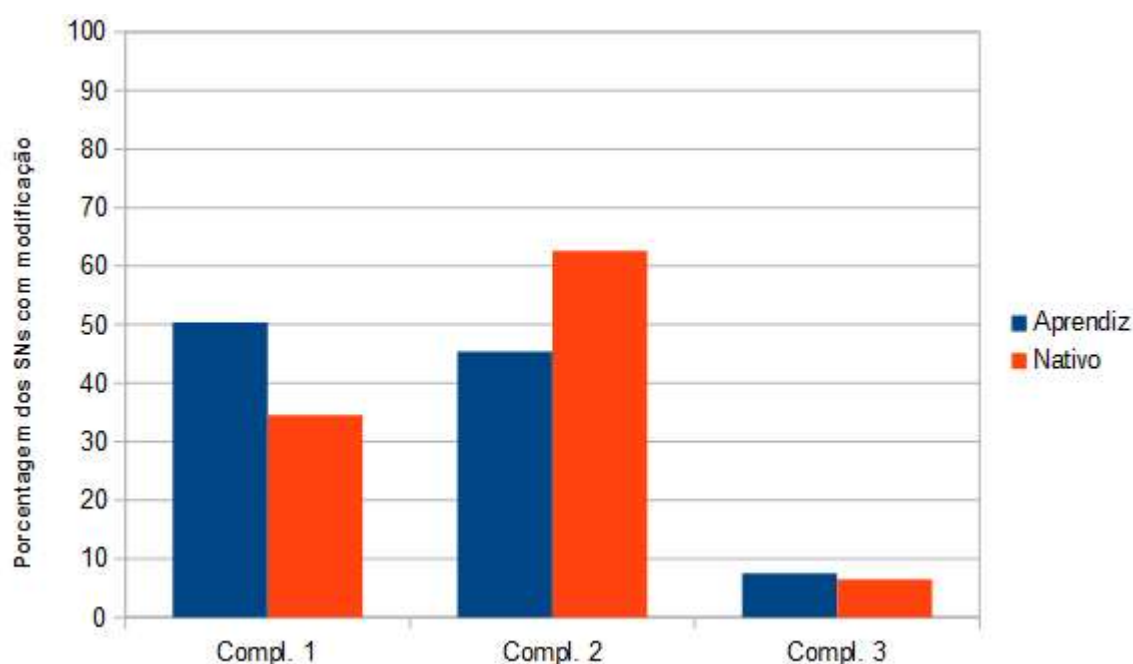


O Gráfico 2 mostra a relação de complexidades dos SNs com o número total de SNs analisados. Os resultados mostram que, quantitativamente, os aprendizes e os nativos estruturam os SNs de seus discursos com diferenças no nível de complexidade. Com base em um olhar mais geral no Gráfico 2, diferenças significativas se apresentaram nas complexidades 1 e 2. Proporcionalmente, os aprendizes produziram ligeiramente mais SNs com apenas pré-modificação (13% contra 11% produzidos pelos nativos) e menos SNs com apenas pós-modificação (11,7% contra 20% produzidos pelos nativos). Ou seja, comparando as complexidades 1 e 2 dos SNs produzidos por aprendizes e nativos, na produção dos aprendizes, houve maior frequência na complexidade 1, seguido pela complexidade 2. Já a produção dos nativos apresentou a frequência inversa, sendo a complexidade 2 mais frequente que a complexidade 1.

Em relação à complexidade 3, os resultados foram uma surpresa. Tanto os nativos quanto os aprendizes produziram um número baixíssimo de SNs com pré e pós-modificação ao mesmo tempo. Minha hipótese era que os aprendizes produziram poucos SNs de complexidade 3 e em menor quantidade que os nativos.

Os grupos, no entanto, apresentaram muito pouca diferença nessa produção (1,9% por parte dos aprendizes e 2,1% por parte dos nativos).

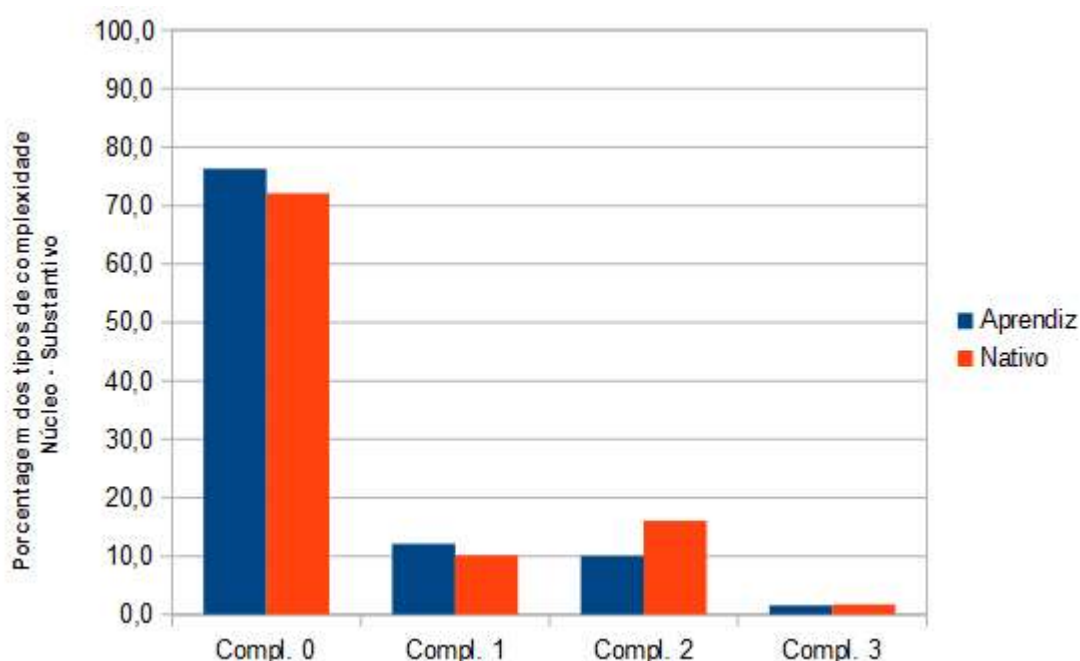
Gráfico 3: Relação das complexidades entre os SNs com modificação



O Gráfico 3 mostra a relação dos SNs modificados com o tipo de modificação que eles recebem (complexidades 1, 2 e 3), com base no número total de SNs que recebem algum tipo de modificação. Observa-se que os nativos produziram 63% de SNs de complexidade 2 e 35% de complexidade 1. Já os aprendizes não apresentam uma diferença muito grande entre a complexidade 1 (50%) e 2 (45%). De acordo com esse critério de classificação de complexidade, os aprendizes produziram levemente mais SNs pré-modificados, enquanto os nativos produziram mais SNs pós-modificados. Os SNs de complexidade 3 representam apenas em torno de 7% e 6% dos SNs modificados entre a produção dos nativos e dos aprendizes, respectivamente.

Ao se separar os modificadores por tipo de núcleo, diferenças ainda maiores foram encontradas. A seguir, temos o Gráfico 4, que mostra os SNs cujo núcleo é um substantivo, separados por tipo de complexidade.

Gráfico 4: Relação das complexidades dos SNs com núcleo Substantivo

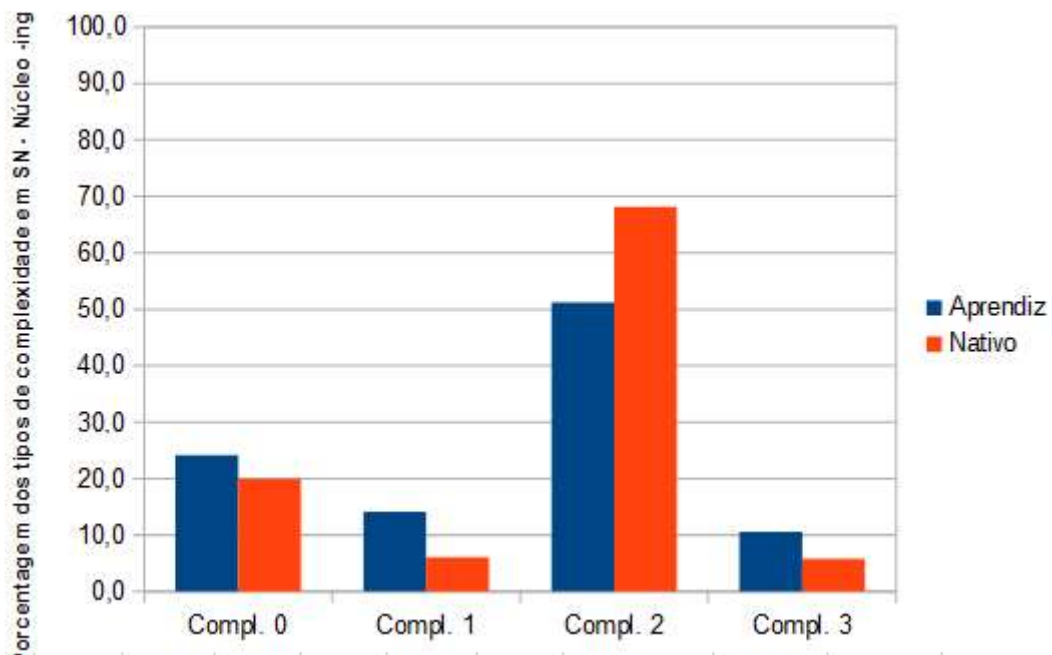


O Gráfico 4 mostra a porcentagem dos tipos de complexidade em SNs cujos núcleos são realizados por substantivos. Esse gráfico se assemelha ao Gráfico 2. Por isso, observa-se que a produção dos aprendizes revela maior proporção de SNs sem modificação (76,3%, contra 72,1% por parte dos nativos) e com pré-modificação (12,1%, contra 10,1% por parte dos nativos). Já os falantes nativos produziram maior quantidade de SNs pós-modificados (16%, contra 10% por parte dos aprendizes).

5.1.1 Modificação dos nomes verbais

A frequência de cada tipo de complexidade para os SNs cujo núcleo é um nome verbal difere dos SNs cujo núcleo é um substantivo. Essa diferença pode ser observada no gráfico 5.

Gráfico 5: Relação das complexidades dos SNs com núcleo *-ing*



O Gráfico 5 mostra a porcentagem dos tipos de complexidade em SNs cujo núcleo é realizado por nomes verbais. Em primeiro lugar, notei que, diferentemente dos SNs cujo núcleo é substantivo, os núcleos *-ing* são majoritariamente modificados, conforme exemplos.

- Complexidade 0: (94) *an understanding* (LOCNESS)
- Complexidade 1: (95) *this **excellent** reasoning* (LOCNESS)
- Complexidade 2: (96) *the arising **of new information*** (BRICLE).
- Complexidade 3: (97) *the **intense craving for the right clothes*** (LOCNESS)

Os SNs sem modificação correspondem a 20% (nativos) e 24% (aprendizes), enquanto os SNs com algum tipo de modificação correspondem a 80% (nativos) e 74% (aprendizes).

Apesar de os aprendizes e os nativos tenderem a pós-modificar mais SNs com esse tipo de núcleo, os aprendizes ainda pós-modificam menos (51%) do que os nativos (68%) e pré-modificam mais (14,1%) do que os nativos (6,1%).

Outro ponto a ser observado é em relação à variedade lexical dos SNs com núcleo *-ing*. Ao analisarmos cada um dos nomes verbais produzidos por aprendizes, percebemos que palavras mais comuns e frequentes, que quase não são percebidas

como de natureza verbal, como *beginning*, *feeling*, *(human) being* e *meaning*, correspondem a 50% dos núcleos dos SNs pesquisados do corpus de aprendizes, dentre os SNs com núcleo *-ing* por eles produzidos, conforme exemplos 98 a 101.

(98) *This seems like the **beginning** of a story.* (BRICLE)

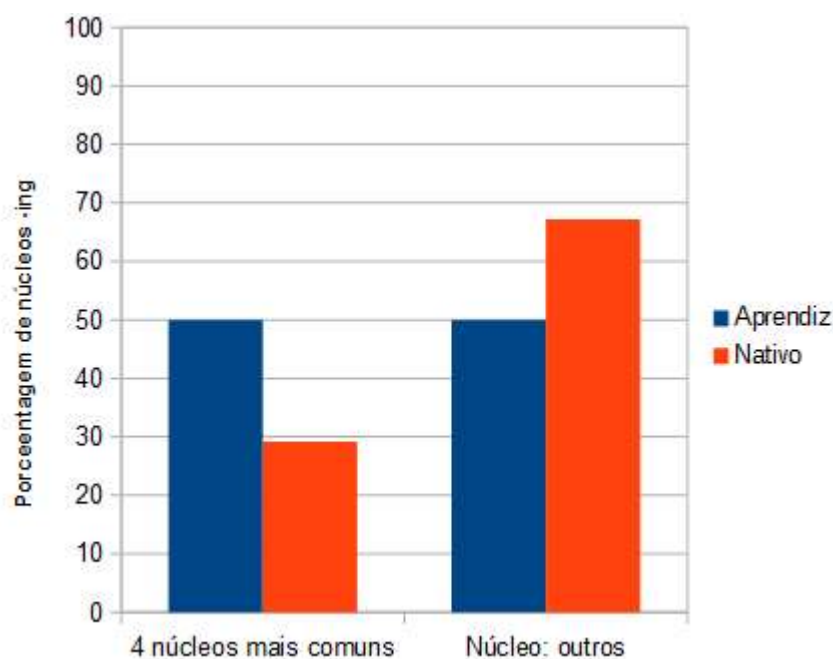
(99) (...) *just because of the greedy **feeling** of having more and more money.* (BRICLE)

(100) *According to the bible, the human **being** used to live longer.* (BRICLE)

(101) *for Americans the **meaning** of university is both shifted and broadened.* (BRICLE)

Em contrapartida, essas 4 palavras correspondem a 29% dos núcleos dos SNs pesquisados no *corpus* de nativos, dentre os SNs com núcleo *-ing* por eles produzidos, conforme o Gráfico 6:

Gráfico 6: Relação dos núcleos *-ing* realizados pelos 4 termos mais comuns



O Gráfico 6 mostra a porcentagem dos nomes verbais mais frequentes (*beginning*, *feeling*, *[human] being* e *meaning*) com função de núcleo de SN, em relação à quantidade total de SNs com núcleo *-ing*. Os dados mostram que os aprendizes apresentaram, em sua produção, menor variedade lexical. Eles produziram, além desses 4, outros 50 nomes verbais variados, o que corresponde a 29% em relação ao total de SNs produzidos com núcleo verbal. Os nativos, por outro

lado, apresentaram uma variedade lexical relativamente maior, produzindo 128 nomes verbais variados, além dos 4 mais comuns (*beginning*, *feeling*, *being* e *meaning*). Isso corresponde a 36% de variação em relação ao total de SNs produzidos com núcleo verbal. Dentre os núcleos produzidos pelos nativos, houve maior variedade de verbos, como *kidnapping* e *mapping*, exemplificados em 102 e 103, e verbos que integram estruturas do tipo *multi-word verbs*, exemplificados em 104 e 105.

(102) (...) *was sentenced to death in 1979 for the **kidnapping** and murder of two San Diego teenagers.* (LOCNESS)

(103) ...*starting with the discovery of DNA and the **mapping** and understanding of it's structure.* (LOCNESS)

(104) (...) *parodying the **taking away** of the people's guilt.* (LOCNESS)

(105) *The **breaking down** of barriers, etc, will be seen as (...)* (LOCNESS)

As principais diferenças encontradas entre a produção de SNs com núcleo *-ing* por parte dos nativos e dos aprendizes foram, então, em relação à variedade lexical. Essa variedade foi menor na produção dos aprendizes. A produção desse grupo também apresentou tendência a pré-modificar mais e pós-modificar menos seus SNs do que os nativos.

5.2 Os pré-modificadores

Conforme apontado por Biber, Poonpon e Gray (2011), a modificação por adjetivos atributivos e substantivos é uma característica do SN da escrita acadêmica. E os dados do presente estudo mostram essa mesma tendência. A tabela 6 mostra a frequência dos tipos de pré-modificadores produzidos por aprendizes e nativos por 10 mil palavras.

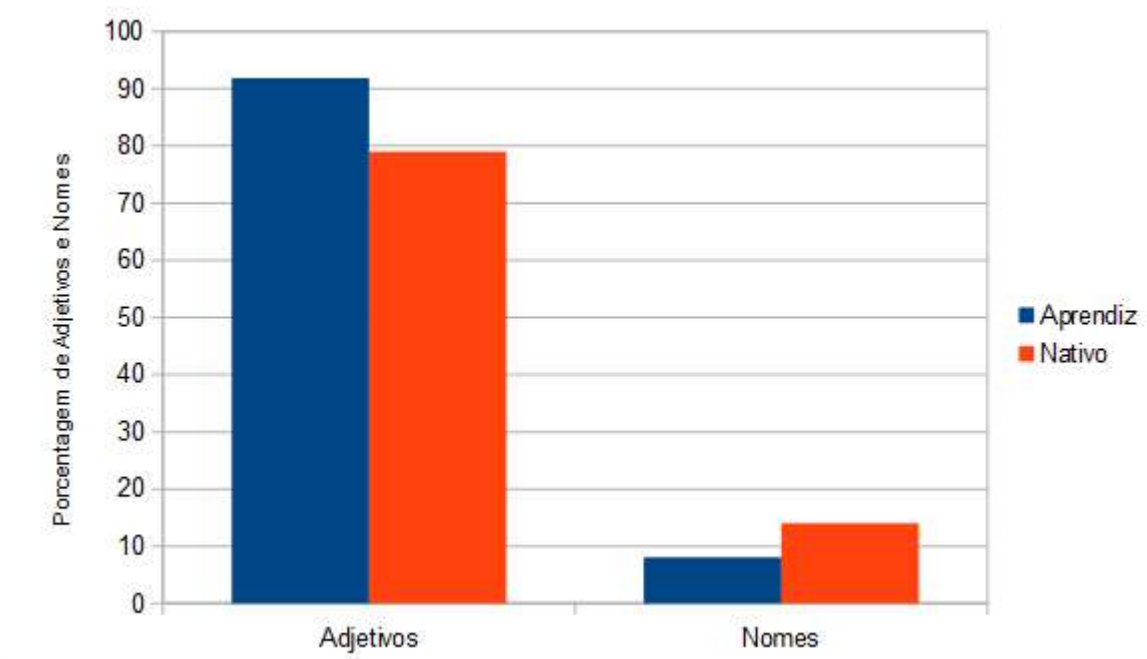
Tabela 6: Frequência dos tipos de pré-modificadores por 10 mil palavras

		Aprendiz		Nativo	
		Freq. Bruta	Freq. Norm.	Freq. Bruta	Freq. Norm.
Pré-Modificador	Adjetivo Atributivo	475	237	426	226
	Adjetivo Particípio	68	34	42	22
	Nome	48	24	74	39
Teste Qui-quadrado = 7,3623 e valor p = 0,04154					

O teste estatístico aplicado indica um valor de p igual a 0,04154. Isso indica que há diferença estatisticamente relevante entre os tipos de pré-modificadores produzidos por aprendizes e falantes nativos, pois o valor de p é menor que 0,05. Os resultados mostram que o aprendiz produz, em maior frequência, adjetivos atributivos ou participios e menos substantivos, como pré-modificadores do que os nativos.

Ao compararmos a proporção de pré-modificadores adjetivos e nominais em relação ao total de SNs pré-modificados, vemos que a maior parte dos SNs recebe modificação de adjetivos, conforme mostrado no Gráfico 7, a seguir.

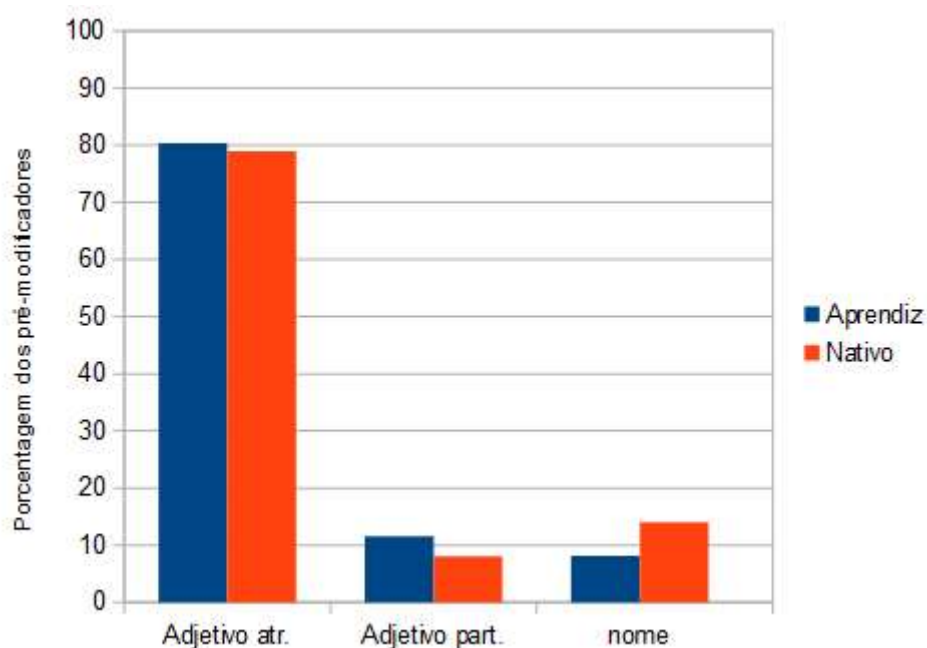
Gráfico 7: Relação dos pré-modificadores



Na produção dos aprendizes, mais de 90% dos SNs pré-modificados recebeu um adjetivo, seja atributivo ou participio, enquanto 79% dos SNs pré-modificados produzidos pelos nativos recebeu um adjetivo. Tanto os nativos quanto os aprendizes apresentaram baixa produção de modificação por substantivos. Os aprendizes apresentaram apenas 8% de SNs pré-modificados por substantivos, enquanto os nativos apresentaram a frequência dessa pré-modificação mais alta: 14%.

O Gráfico 8 mostra a porcentagem de cada tipo de pré-modificador em relação ao total de SNs pré-modificados. Os tipos de pré-modificadores encontrados foram adjetivos atributivos, adjetivos participio e substantivos.

Gráfico 8: Relação dos tipos de pré-modificadores



A seguir, exemplos de cada tipo de pré-modificador encontrado na produção dos aprendizes e nativos.

- SN pré-modificado por adjetivo atributivo:

(106) aprendiz: *To get **a good job**, it is fundamental to study hard.* (BRICLE)

(107) nativo: *Along with **the moral problems** there are also physical dangers.* (LOCNESS)

- SN pré-modificado por adjetivo participio:

(108) aprendiz: ***Extroverted children** tend to have ability to meet new friends.* (BRICLE)

(109) nativo: ***Impoverished people** were interpreted negatively as seen in the book.* (LOCNESS)

- SN pré-modificado por nome:

(110) aprendiz: *I have heard from **Law students** the same complaints: too much theory.* (BRICLE)

(111) nativo: *Consequently, cheating has become an aspect of **college life**.* (LOCNESS)

Em ambos os *corpora* estudados, a produção de SN pré-modificado por adjetivo atributivo foi mais frequente. Para os aprendizes, 80% das ocorrências de

pré-modificação foram com adjetivo atributivo; para os nativos, 78%. Os aprendizes produziram mais adjetivos participio (12%) do que nomes (8%) como pré-modificadores. Os nativos, por outro lado, produziram uma quantidade maior de nomes (14%) como pré-modificadores em comparação aos adjetivos participio (8%). Esse gráfico 8 mostra uma frequência levemente mais alta para a pré-modificação feita por adjetivos participio em detrimento da pré-modificação feita por substantivos, no *corpus* de aprendizes, em comparação com o *corpus* de nativos.

5.2.2 Estruturas atípicas encontradas na produção dos aprendizes referentes à pré-modificação de sintagma nominal

O processo manual de classificação dos tipos de modificadores de SN me permitiu identificar as estruturas atípicas da língua inglesa que ocorreram no *corpus* de aprendizes, no que se refere à pré-modificação. Utilizei o termo atípico para me referir às formas que fogem ao padrão culto da língua.

Os aprendizes apresentaram algumas dessas formas atípicas no que se refere ao uso de sufixos formadores de adjetivos e de substantivos. Na grande maioria dos casos, um sufixo de substantivo foi utilizado quando um nativo teria utilizado um sufixo formador de adjetivo. Os casos em questão se encontram destacados em negrito. Cada um desses casos, comentados nesta seção, foram conferidos nos *corpora* de referência e nos comparativos BAWE e LOCNESS, para que fosse verificado se, realmente, essas construções não seriam típicas na escrita de nativos.

(112) *Everyone has television in these days. In Rocinha slum, for example, 97% of the lodgers have TV, and 80% of the sets were changed for new ones since four years ago, so they use the little quantity of money they have with televisions despite of food or something essencial for a health life.. Many people have cable TV, too, and their favorite channels are Globo, Cartoon and Discovery.* (BRICLE)

No exemplo (112), o SN a *health life* possui um núcleo, *life*, que é modificado pelo termo *health*. A partir do contexto em que o SN foi escrito, a forma mais típica dessa construção seria com o modificador na forma adjetiva *healthy*, e não na forma nominal. Nesse trecho, além desse SN, observa-se que o SPrep *in these days* também apresenta uso atípico. Apesar dessa construção ser sintaticamente

possível, a expressão mais comum seria *these days* sem a preposição *in*, ou a expressão *nowadays*. Embora essa construção seja utilizada pelos nativos, o uso mais comum de *in these days* pelos nativos é na construção *in these days of*. É provável que a presença dessa preposição se dê devido à influência do português, visto que expressões de tempo como essa, no português, geralmente pedem uma preposição.

Outra forma atípica encontrada na produção dos aprendizes foi o uso de adjetivos derivados de verbos sem o sufixo formador de adjetivo apropriado.

(113) *It is common to hear some students saying when their courses at the university get completed they is going to be working in a big company and manage business, other ones have the dream to become an businessman, being a **succeed person**.* (BRICLE)

No exemplo 113, no SN *a succeed person*, o núcleo é modificado pela forma verbal *succeed* em vez da forma adjetiva *successful*.

Foram encontrados também formas atípicas de pré-modificadores adjetivos derivados de substantivos sem o sufixo (adjetivo escrito na forma nominal), conforme exemplo a seguir.

(114) *Nowadays these institutions offers a variety of courses in each specific area, being human or biological. They claim that after the course students is going to be prepared to face the real world, solve problems, being a **prominence person** in a company.* (BRICLE)

No exemplo 114, o SN *a prominence person* apresenta um pré-modificador substantivo *prominence*. A forma padrão utilizada seria *prominent*, a forma adjetiva do termo.

Foram encontrados, também, pré-modificadores realizados por adjetivos, porém na forma verbal infinitiva, em vez de na forma particípio, mostrado no exemplo 115:

(115) *Actually he wants to be like they are. His mother is a widow and a **very worry person**; also she is **an over protection person**, prejudicing his son's development.* (BRICLE)

O exemplo 115 mostra o trecho de um ensaio em que dois SNs foram produzidos com pré-modificador atípico. O primeiro SN destacado, *a very worry person*, possui um pré-modificador que, num registro típico de nativos, estaria na

forma de adjetivo particípio *worried*. Já *an over protection person* possui um pré-modificador na forma nominal, *protection*, em vez de na forma adjetiva *protective*. Além disso, temos um item de vocabulário inexistente na língua inglesa, *prejudicing* (em vez de *jeopardizing*), que é claramente uma palavra do português com flexão da língua inglesa.

Além dos casos de sufixação atípicas mencionados anteriormente, foram encontrados problemas referentes ao uso de genitivo, conforme exemplo (116):

(116) *In conclusion, televisions are essential to most families in the world and, today, it is capable of creating heroes and villains, winners and losers, experts and idiots. We can no longer accept other **people judgments** without questioning, we need to read books, newspapers and be **critic**.* (BRICLE)

No exemplo 116, o SN *other people judgments* possui um problema no uso da estrutura genitiva 's. O s da estrutura genitiva não foi colocado em *people*. Mais ao final do trecho, encontra-se sublinhada a palavra *critic*. O SN *critic*, pelo contexto, deveria ser um Sadj *critical*, no sentido de uma pessoa crítica, *a critical person*. O termo *critic* é mais utilizado no sentido da profissão, como *food critic*.

Em geral, os casos de pré-modificador atípicos encontrados na produção dos aprendizes são de dois tipos. O primeiro tem relação com a influência da língua materna, visto que muitos desses aprendizes não parecem estar em um estágio avançado de proficiência linguística. Dessa forma, ainda escrevem em inglês com base em estruturas e palavras do português. O segundo tem relação com o uso de sufixos, já que, em quase todos os casos, seja para substantivos ou adjetivos, o uso do sufixo não foi o ideal.

5.3 Os pós-modificadores

Conforme discutido anteriormente, os pós-modificadores são geralmente realizados por sintagmas preposicionados e por orações.

Tabela 7: Frequência dos tipos de pós-modificadores por 10.000 (dez mil) palavras

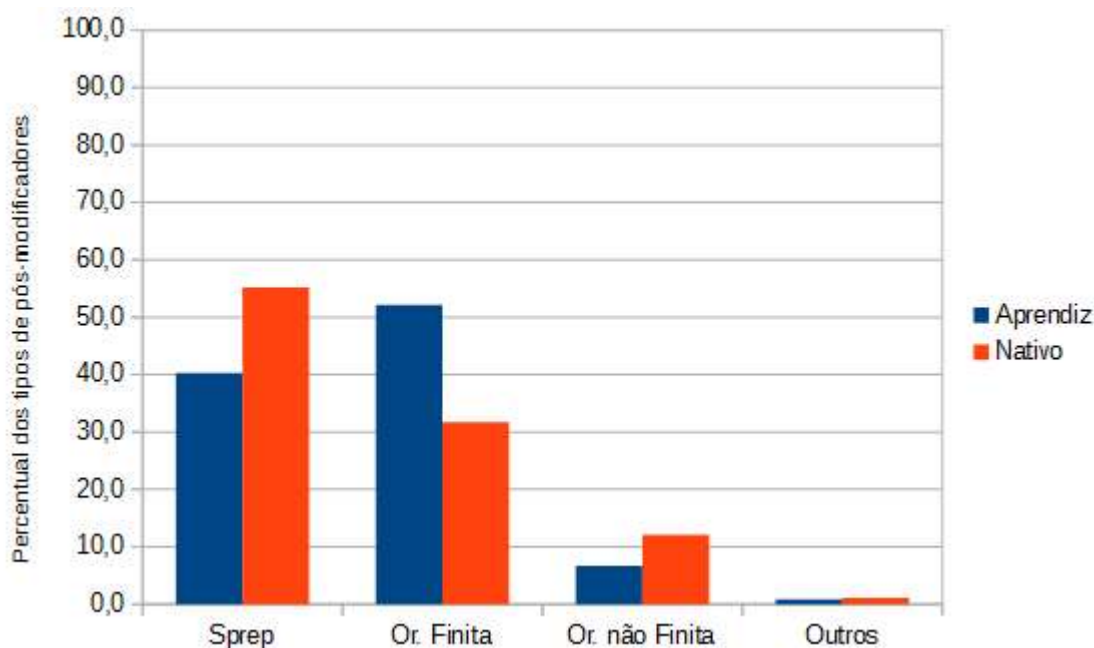
		Aprendiz		Nativo	
		Freq. Bruta	Freq. Norm.	Freq. Bruta	Freq. Norm.
Pós-Modificador	Sintagma Prep.	228	114	555	294
	Oração Finita	295	147	319	169
	Oração não-Finita	38	19	112	59
	Outros	5	2	14	7

Teste Qui-quadrado = 31,973 e valor $p < 0,00001$

O teste estatístico aplicado indica um valor de p menor que 0,00001. Isso significa que há diferença estatisticamente significativa entre as complexidades analisadas, pois o valor de p é menor que 0,05.

Em uma menor frequência, adjetivos, pronomes, dentre outros elementos, também podem realizar a função de pós-modificadores (classificados como outros na tabela 7). Ao avaliarmos o Gráfico 9, notamos que, de fato, a grande maioria dos pós-modificadores são realizados por sintagmas preposicionados e por orações. Os outros tipos de modificadores não configuram nem 2% dos casos, nos dois *corpora* analisados.

Gráfico 9: Relação dos tipos de pós modificadores



O Gráfico 9 mostra a porcentagem dos tipos de pós-modificadores em relação com o número total de SNs pós-modificados. A seguir, temos exemplos de cada tipo de pós-modificador. O SN se encontra sublinhado e os pós modificadores em negrito.

- Pós-modificador realizado por SPrep.:

(117) Aprendiz: *First of all, no one have the right to terminate **the life of another person**.* (BRICLE)

(118) Nativo: *Another value stated by people in favor of biological parents is the burden of guilt.* (LOCNESS)

- Pós-modificador realizado por oração finita:

(119) Aprendiz: *Many politicians believe that people who attend undergraduate courses are not qualified to work immediately.* (BRICLE)

(120) Nativo: *The number of people who enjoy the sport far outweigh those who dislike it.*

- Pós-modificador realizado por oração não finita:

(121) Aprendiz: (...) *due to the demand for such courses as well as the number of people willing to attend them.* (BRICLE)

(122) Nativo: *The defendant, and person to be executed, was Robert Alton Harris.* (LOCNESS)

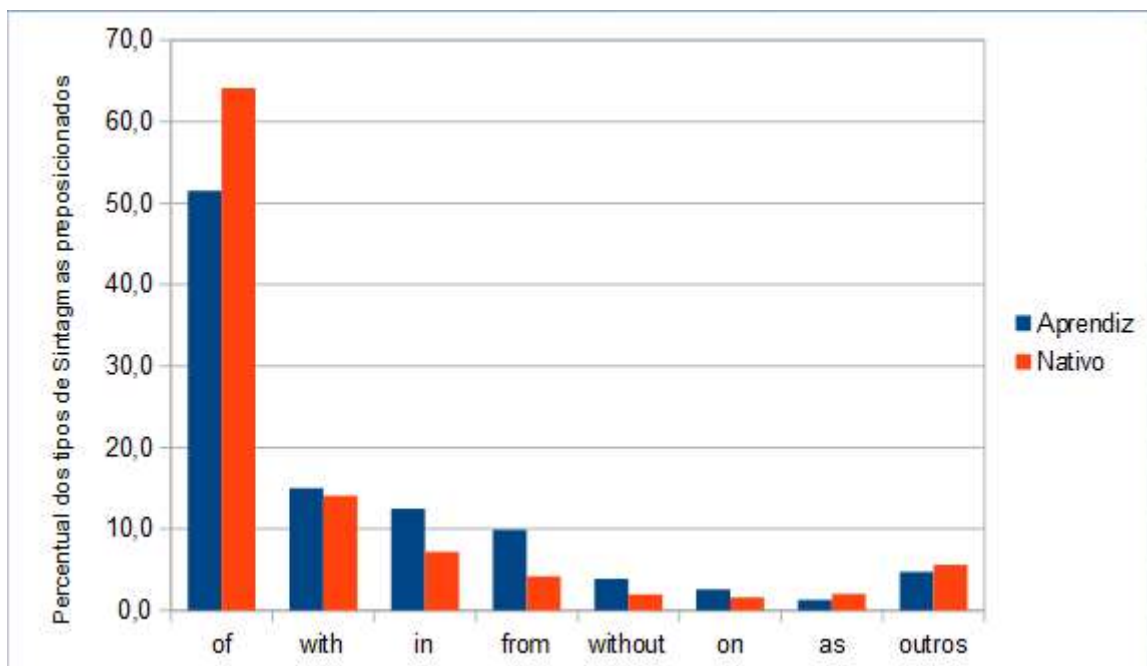
Observa-se, também, no gráfico 9, que os tipos mais frequentes de pós-modificação são o sintagma preposicionado e a oração finita. No entanto, os aprendizes produziram proporcionalmente mais orações finitas (52%, contra 32% por parte dos nativos), diferentemente dos nativos, que produziram proporcionalmente mais sintagmas preposicionados (55%, contra 40% por parte dos aprendizes). Tanto os nativos quanto os aprendizes produziram menos orações não finitas (12% e 8%, respectivamente) em relação aos modificadores realizados por SPreps e orações finitas. Contudo, os aprendizes apresentaram uma produção mais baixa de orações não finitas do que os nativos. Os outros tipos de modificadores, que não somaram 2% dos casos em ambos os *corpora*, foram realizados por adjetivos, nomes e pronomes reflexivos.

5.3.1 Pós-modificadores realizados por sintagmas preposicionados

Os pós-modificadores realizados por SPreps representam a maior parte dos pós-modificadores produzidos pelos nativos (55%), diferentemente dos pós-modificadores produzidos pelos aprendizes, que foram, na sua maioria, realizados por orações finitas. Porém, ao analisarmos a variedade de preposições utilizadas

pelos nativos e aprendizes, observamos que os dois grupos utilizam uma variedade semelhante, conforme o gráfico 10.

Gráfico 10: Relação dos tipos de pós-modificadores realizados por sintagmas preposicionados



O Gráfico 10 mostra a porcentagem dos tipos de sintagmas preposicionados em relação ao número total de SNs pós-modificados por sintagmas preposicionados. Embora tenham apresentado uma variedade de preposições parecida com a variedade apresentada pelos nativos, os aprendizes produziram poucos SNs pós-modificados por SPrep em comparação com os nativos. Os SNs se encontram sublinhados e os pós-modificadores realizados por SPreps estão em negrito.

- Pós-modificador realizado por SPrep com preposição *of*:

(123) Aprendiz: *rehabilitate penitentiary inmates by creating several jails in which they have to work in agriculture, industry or some job of similar kind*. (BRICLE)

Nativo: *Firstly the main problems of road transport will be identified and this shall be followed by suggested solutions to these problems*. (LOCNESS)

- Pós-modificador realizado por SPrep com preposição *with*:

(124) Aprendiz: *On countries that apply limitation laws, only children receive total medical and educational support, while children with siblings are not supported by the government in any means*. (BRICLE)

Nativo: *Blatant overpricing of many important drugs has drastically declined the quality of health care among many Americans, especially the elderly and people with fixed incomes*. (LOCNESS)

- Pós-modificador realizado por SPrep com preposição *in*:

(125) Aprendiz: *The money in the prison system is not well canalized*. (BRICLE)

Nativo: *The powerless people in society will try to overturn the powerful ruling class*. (LOCNESS)

- Pós-modificador realizado por SPrep com preposição *from*:

(126) Aprendiz: *People from several parts of the world have demanded a solution for this problem*. (BRICLE)

Nativo: *This system of education did not lead to much social mobility as the children from poorer social backgrounds left school after primary*. (LOCNESS)

- Pós-modificador realizado por SPrep com preposição *without*:

(127) Aprendiz: *working in a job without any challenge is very boring and tiring*. (BRICLE)

Nativo: *For instance, the majority of society relates negatively to people without homes*. (LOCNESS)

- Pós-modificador realizado por SPrep com preposição *on*:

(128) Aprendiz: *even if you don't need to compete with all this people on the job market*. (BRICLE)

Nativo: *Secondly, gambling is indulged in by a large number of people on lower incomes*. (LOCNESS)

- Pós-modificador realizado por SPrep com preposição *as*:

(129) Aprendiz: *After dropping out from school and getting a job as a waitress at a pub on a busy street in Saint Leonards (E. Sussex, UK), she left her four year old daughter with her mother and moved out*. (BRICLE)

Nativo: *Macy's tried to compensate by offering Mr Wookley a job as a Santa Supervisor*. (LOCNESS)

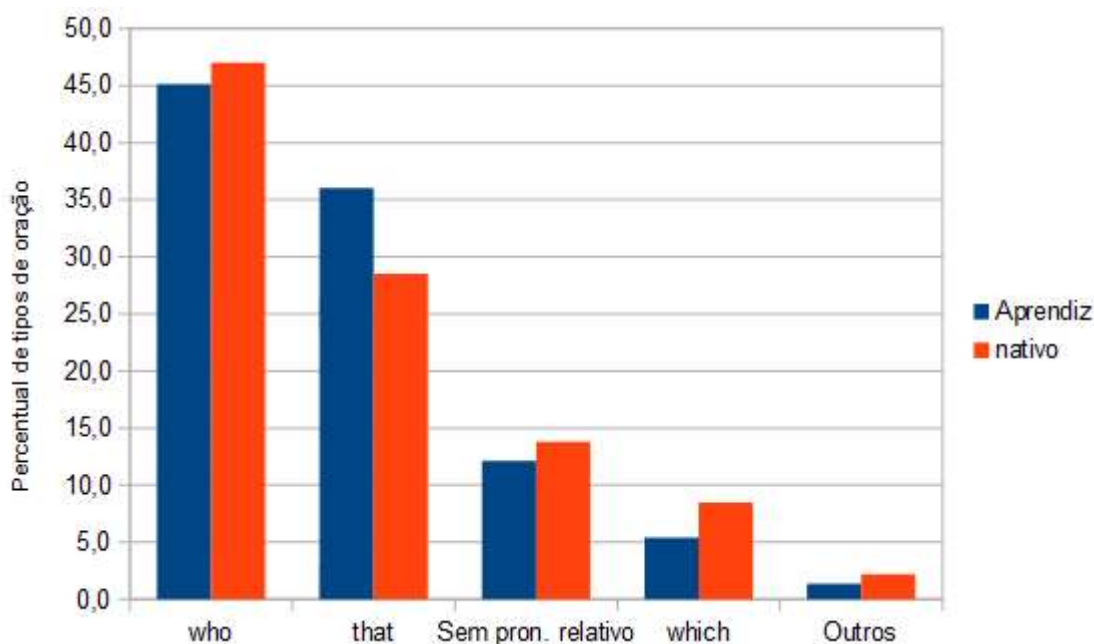
A preposição mais frequente para os modificadores realizados por SPreps é a preposição *of* (51,5% por parte dos aprendizes e 64% por parte dos nativos), seguida pelas preposições *with* (15% por parte dos aprendizes e 14,1% por parte dos nativos), *in* (12,4% por parte dos aprendizes e 7,2% por parte dos nativos), *from* (9,9% por parte dos aprendizes e 4,2% por parte dos nativos), *without* (3,9% por parte dos aprendizes e 2% por parte dos nativos), *on* (2,6% por parte dos aprendizes e 1,6% por parte dos nativos) e, por último, *as* (1,3% por parte dos aprendizes e 2% por parte dos nativos). Outras preposições, como *for*, *into*, *at*, *across* e *within* não chegaram à frequência de 1% dos SNs pós-modificados por SPreps e, por isso, foram contabilizadas como Outros.

Os pós-modificadores realizados por sintagmas preposicionados com a preposição *of* representam a maior parte dos pós-modificadores de SN em escrita acadêmica profissional do inglês, seguidos pelas preposições *in*, *on* e *with* (BIBER, GRAY, POONPON, 2011). Os resultados apresentados no gráfico 10 mostram que os aprendizes e nativos produziram grande parte das preposições mais frequentes desse gênero.

5.3.2 Pós-modificadores realizados por orações

Conforme visto na seção 5.1, os aprendizes apresentaram maior proporção de pós-modificadores realizados por orações do que por sintagmas preposicionados.

Gráfico 11: Relação dos tipos de pronomes relativos introdutores de orações finitas como pós-modificadores.



O Gráfico 11 mostra a porcentagem dos tipos de pronomes relativos introdutores de orações finitas como pós-modificadores em relação ao número total de SNs pós-modificados por orações finitas. As orações relativas mais produzidas pelos aprendizes e nativos foram as introduzidas por *who*, com frequência de 45% por parte dos aprendizes e 46% por parte dos nativos. Além disso, nota-se que os aprendizes apresentaram uma frequência maior de orações relativas introduzidas por *that* do que os nativos (36% contra 28,5%). Os aprendizes e nativos

apresentaram uma variação de pronomes relativos semelhante, além de produzirem uma proporção semelhante de orações finitas sem pronome relativo, conforme exemplos (130) a (133).

- Pós-modificador realizado por oração finita encabeçada por pronome relativo *who*:

(130) Aprendiz: *Now people who control television control the information that gets in people's houses.* (BRICLE)

Nativo: *The target group was children who had become economically dependent because of the deaths of their fathers.* (LOCNESS)

- Pós-modificador realizado por oração finita encabeçada por pronome relativo *that*:

(131) Aprendiz: *The most recent case of a person that became famous from night to day on the Internet is from an old Scottish woman called Susan Boyle.* (BRICLE)

Nativo: *The students that were taught sex education in school are now personally benefiting from their knowledge of sexual education.* (LOCNESS)

- Pós-modificador realizado por oração finita sem pronome relativo:

(132) Aprendiz: *Rita discovered that she needed to acquire culture, to learn about many things she had never thought before.* (BRICLE)

Nativo: *On the other hand, its easy to do the things everyone expects of your gender.* (LOCNESS)

- Pós-modificador realizado por oração finita encabeçada por pronome relativo *which*:

(133) Aprendiz: *obesity, sleep deprivation, stunted sensory of development, are some of the problems which you may acquire by watching television excessively.* (BRICLE)

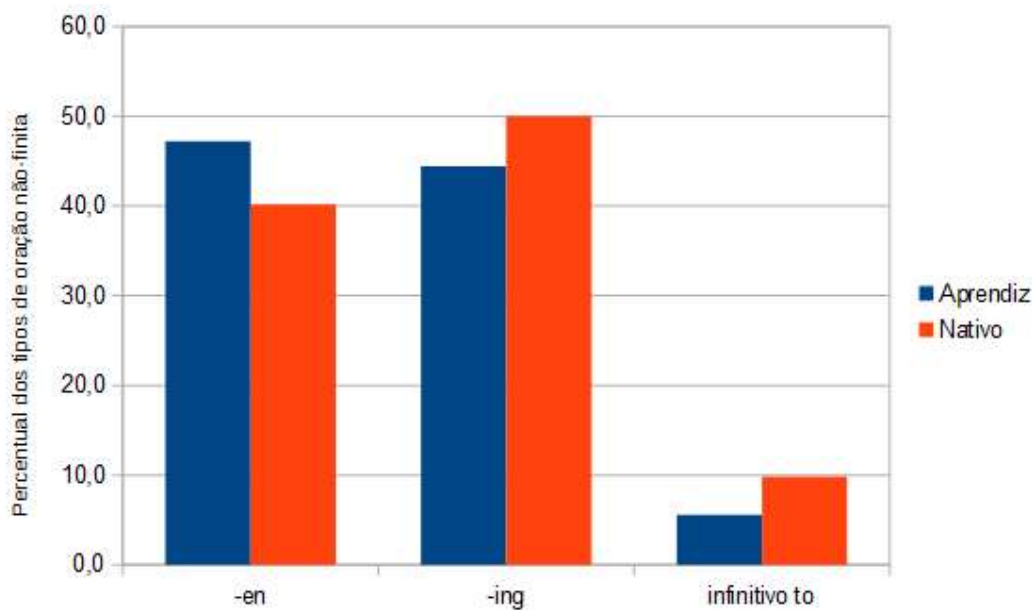
Nativo: *Problems which may arise for the parents are mainly the chances of the treatment being unsuccessful.* (LOCNESS)

Ao analisarmos os núcleos que acompanham essas orações relativas, notamos uma alta frequência de *that* acompanhando substantivos que se referem a pessoas (*people, person, children e students*). Essa frequência do pronome *that*, mesmo com entidades animadas, foi prevista por Biber e Gray (2011), que notaram um aumento no uso desse pronome relativo introduzindo orações relativas nos últimos dois século em textos jornalísticos. Embora seja um gênero diferente do pesquisado no presente estudo, essa mudança pode, também, ter se refletido no gênero acadêmico.

Além disso, observa-se que os aprendizes também produzem menos orações com o pronome relativo *which* do que os nativos (5,4%, contra 8,5%). Os outros tipos de pronomes relativos produzidos por aprendizes e nativos foram *where*, *whose* e *whom*, que apresentaram frequência abaixo de 1% cada.

A frequência de pós-modificadores realizados por orações não finitas, por outro lado, é 50% mais baixa por parte dos aprendizes em comparação com os nativos. Esse tipo de pós-modificador, no entanto, é menos frequente em relação aos outros dois tipos, sintagmas preposicionados e orações finitas. As orações não finitas representam aproximadamente 11% dos pós-modificadores produzidos por nativos e aproximadamente 6% dos pós-modificadores produzidos por aprendizes. Conforme gráfico 12.

Gráfico 12: Relação dos tipos de pós-modificadores realizados por orações não finitas



O Gráfico 12 mostra a porcentagem dos tipos orações não finitas como pós-modificadores, em relação ao número total de SNs pós-modificados por orações não finitas, conforme exemplos (134) a (136).

- Pós-modificador realizado por oração não finita com verbo na forma particípio passado (*en*):

(134) Aprendiz: *If there were such a law the money acquired by these billionaires should be used to diminish the social differences between the rich and the poor.* (BRICLE)

Nativo: *Generally children born into impoverished environments remain in that socioeconomic level their entire life.* (LOCNESS)

- Pós-modificador realizado por oração não finita com verbo na forma particípio presente (*ing*):

(135) Aprendiz: *So, according to IBGE São Paulo's capital has around 9.898.046 inhabitants which seems a large number of people competing on the job market.* (BRICLE)

Nativo: *there was never any concealment of the money going to the government.* (LOCNESS)

- Pós-modificador realizado por oração não finita com verbo na forma infinitiva (*to*):

(136) Aprendiz: *As far as you have got a television and the money to afford cable tv you have the whole world to see.* (BRICLE)

Nativo: *It must take into account the strengths and weaknesses of the people to be trained.* (LOCNESS)

Os tipos mais frequentes de oração não finita com função de pós-modificador produzidos por aprendizes e nativos foram as orações contendo particípio presente (*-ing*) e particípio passado (*-en*). Contudo, em geral, os aprendizes produziram menos orações não finitas que os nativos (conforme gráfico 9).

As orações realizadas por infinitivo *to* apresentaram frequência mais baixa nos dois *corpora* pesquisados. Além disso, os aprendizes produziram pós-modificação realizada por oração não finita com verbo na forma infinitiva em frequência bem mais baixa do que os nativos. Em relação aos nativos, 10% dos SNs pós-modificados por oração não finita são realizados por oração do tipo infinitivo *to*, enquanto, em relação aos aprendizes, 5% dos pós-modificadores de SNs pós-modificados por oração não finita são realizados por oração do tipo infinitivo *to*.

5.3.3 Estruturas atípicas encontradas na produção dos aprendizes referentes à pós-modificação de sintagma nominal

Conforme discutido, os tipos mais comuns de pós-modificadores são as orações e os sintagmas preposicionados. Em relação aos modificadores realizados por orações, os aprendizes apresentaram um uso fora da norma, referente ao pronome relativo *whom*. Esse pronome relativo ocorreu apenas 2 vezes e essas

duas produções apresentaram problemas: em ambos os casos, o pronome relativo *whom* foi utilizado em uma oração relativa com função de sujeito da oração relativa. As duas ocorrências foram encontradas no mesmo ensaio, o que indica que as orações foram escritas pela mesma pessoa, conforme exemplo (137).

(137) *Taking everything into account, every person have the right to use their money as they want, but never forgetting that there is always a person **whom** has no blanket or a home to sleep, or money enough for a medical treatment or eat. Being rich its not having a great count of money, but it is to be a given person **whom** knows how to help other people, with money, clothes, a smile or donating blood.* (BRICLE)

Observe que, no trecho acima, o pronome relativo mais apropriado para ser usado nas duas ocorrências seria *who*, já que esse pronome tem a função de sujeito nas orações relativas que introduzem, e não de objeto.

Esses dois casos de *whom* encontrados foram escritos pela mesma pessoa. Por esse motivo, selecionei todas as ocorrências do pronome relativo *whom* no *corpus* de aprendiz para verificar se esse uso seria um caso isolado. Em geral, o pronome *whom* foi pouco utilizado no *corpus* de aprendiz, totalizando 13 ocorrências. E dessas ocorrências, 5 foram usadas fora da norma, o que representa 38%.

Em relação aos modificadores realizados por preposições, os aprendizes produziram muito poucas estruturas atípicas. Elas se concentraram em preposições menos frequentes, descritas como “outros” no gráfico 10. A seguir, temos um exemplo de estrutura atípica de SN com pós-modificador realizado por SPrep, produzido por aprendizes. O SN encontra-se sublinhado e o modificador em negrito.

(138) *Third, as many adults did not have any opportunities of study during their early life, they are illiterate. Moreover, the government offers a reduced number of vacancies for them in public schools. All in all, Brazilian's problems **about equality in education** are not alone. The situation is even more irksome in Africa and Asia as well.* (BRICLE)

Observe que, no exemplo (138), o núcleo do SN *problems* é pós-modificado pelos SPrep *about equality in education*. O substantivo *problems*, quando pós-modificado por um SPrep, geralmente ocorre com as preposições *with*, *in* e *of* no *corpus* dos nativos, dependendo do contexto. No entanto, o uso dessa preposição também foi encontrado no *corpus* de nativos em frequência bem reduzida.

5.4 O USO DE PRONOMES

Na terceira parte deste trabalho, analisei quantitativamente os pronomes pessoais. Conforme discutido no segundo capítulo, além dos substantivos, pronomes também ocorrem como núcleo de SN e, geralmente, sem modificação. Os pronomes que são modificados com maior frequência são os indefinidos e são, geralmente, pós-modificados com orações relativas finitas, não finitas e adjetivos. Tendo em vista a não modificação dessas unidades, o SN com núcleo realizado por pronome pessoal foi, então, considerado de complexidade 0 no presente estudo.

Tabela 8: Frequência dos pronomes pessoais por 10 mil palavras

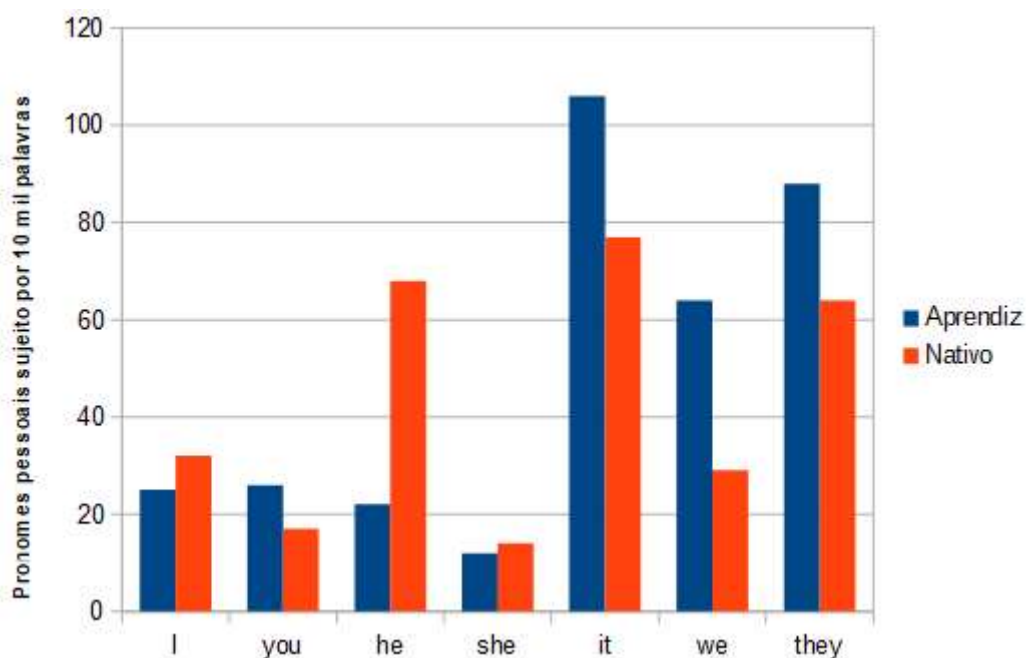
Pronome	Aprendiz		Nativo	
	Freq. Bruta	Freq. Norm.	Freq. Bruta	Freq. Norm.
Pron. pess. suj.	5616	344	9,816	300
Pron. pess. obj.	1446	88	2,799	86
Total	7062	432	12,615	386
Teste qui-quadrado = 0,33711 e valor p = 0,5615				

No total, os aprendizes produziram mais pronomes pessoais (344) por 10 mil palavras do que os nativos (300). Contudo, o teste qui-quadrado indica que não há diferenças estatisticamente relevantes quando comparamos o total de pronomes de sujeito e objeto entre aprendizes e nativos, já que o valor de p é igual a 0,5615. E, de fato, os aprendizes e nativos não apresentam diferenças no que se refere à frequência dos SNs realizados por pronomes pessoais de objeto. A maior diferença ocorreu entre os pronomes pessoais de sujeito, conforme mostrado na Tabela 8.

Quando a frequência de cada pronome pessoal de sujeito produzida pelos aprendizes é comparada com a frequência da produção dos nativos, é possível identificar quais pronomes são mais utilizados. Ao se analisarem esses pronomes mais frequentes a partir do contexto expandido de alguns ensaios produzidos por aprendizes, também é possível identificar uma provável sobre utilização desses pronomes.

O gráfico 13 mostra a frequência de cada pronome pessoal de sujeito por 10 mil palavras.

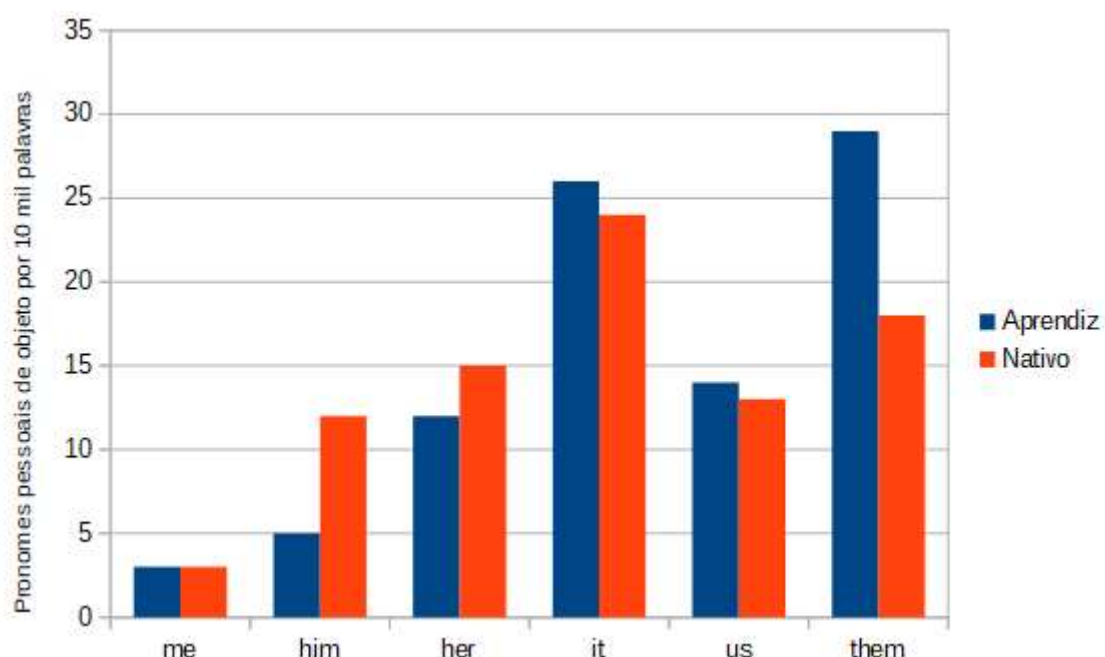
Gráfico 13: Relação dos pronomes pessoais de sujeito por 10 mil palavras



O teste qui-quadrado, aplicado com a frequência dos pronomes pessoais de sujeito, indica diferenças significativas, já que o valor de p é $p < 0,00001$. Conforme mostrado no Gráfico 13, os pronomes pessoais de sujeito produzidos em maior frequência pelos aprendizes (em comparação aos nativos) foram os pronomes *it* (106 por parte dos aprendizes e 77 por parte dos nativos), *they* (88 por parte dos aprendizes e 64 por parte dos nativos), *we* (64 por parte dos aprendizes e 29 por parte dos nativos) e *you* (26 por parte dos aprendizes e 17 por parte dos nativos).

Em relação aos pronomes pessoais de objeto, os aprendizes e nativos não apresentam muitas diferenças na frequência produzida. O teste qui-quadrado aplicado com a frequência dos pronomes pessoais de sujeito não indica diferenças significativas, já que o valor de p é $p = 0,3243$. O gráfico 14 mostra a frequência dos pronomes pessoais de objeto por 10 mil palavras.

Gráfico 14: Relação dos pronomes pessoais de objeto por 10 mil palavras



De acordo com a frequência dos pronomes pessoais de objeto mostrados no Gráfico 14, os aprendizes e nativos apresentam diferenças menores do que as diferenças apresentadas com os pronomes pessoais de sujeito. Em geral, os aprendizes produziram quantidade semelhante de pronomes pessoais de objeto (85 por 10 mil palavras) em comparação com os produzidos pelos nativos (89 por 10 mil palavras). O único pronome pessoal de objeto que parece ser mais sobre utilizado pelos aprendizes é *them*, com frequência de 29 casos por 10 mil palavras por parte dos aprendizes e 18 por 10 mil por parte dos nativos.

Em análise do contexto de algumas ocorrências dos pronomes utilizados pelos aprendizes, foi verificado que, em alguns ensaios, o pronome realmente é sobre utilizado por aprendizes. Esse é o caso do *it*, que, em alguns ensaios, foi utilizado tantas vezes que o referente acabou se perdendo, formando um texto confuso, como no trecho do exemplo (139).

(139) *The better thing that can happen in the woman life is she has a baby, but if she does not agree with it, it becomes the worse thing, mainly when it is a rape.*

Note que, no exemplo (139), o pronome *it* foi usado três vezes, possivelmente referindo-se a elementos diferentes do trecho em destaque, tornando o período confuso.

Já no exemplo (140), temos um trecho de um texto mais bem estruturado, porém com um problema semelhante ao trecho (139) no que se refere ao uso do pronome *it*.

(140) *The changing of technology in a fast pace, however, has also its negative effects in human lives. Firstly, in what concerns to childhood, technology plays a very important role in the way **it** changes. Children do not use their imagination anymore.*

No trecho do exemplo 140, temos o pronome *it* em negrito na terceira linha. No início do trecho, o texto fala sobre a mudança da tecnologia, com o SN *the changing of technology in a fast pace*. Na oração seguinte, ele fala de infância, com o SN *childhood*. Em seguida, é afirmado que a tecnologia tem um papel importante na forma como algo (*it*) muda. Esse pronome *it* pode estar se referindo a qualquer um dos dois SNs mencionados (*the changing of technology* e *childhood*). Na próxima oração, quando o aprendiz menciona que as crianças não usam mais a imaginação, essa ambiguidade fica ainda mais evidente. Por isso, não se sabe o que mudou: a tecnologia ou a infância.

O mesmo ocorre com o pronome *we* nos ensaios dos aprendizes brasileiros. Em muitos deles, esse pronome parece ser sobre utilizado. No trecho (141), a seguir, nota-se que *we* foi utilizado 10 vezes.

(141) *The teacher besides not being valorized, is still a good example of how the learning university is rather theoretical. When **we** go out the school, **we** think that we are ready to teach our students. **We** think that it will be easy because **we** are professional, **we** have our certificate. But the frustration is big, when **we** begining our work. **We** are alone in front of the class with fourty different pupils. Each one with different personality, different knowledge. This moment, you realize that **we** don't know to do this. Which procedure take? The university don't teach us what to do in front of situations that will appear. Each situation is unique and **we** have to be resolved. This situation could happen in all jobs, because surprises always happen and especially when **we** work with humans.*

No trecho (141), o aprendiz escreve um ensaio sobre como a universidade é mais teórica que prática. Para ilustrar seu ponto de vista, o aprendiz dá como exemplo a formação acadêmica de um professor. Uma explicação para a alta frequência de *we* pode ser uma questão de estilística. Carter e McCarthy (2015,

2006) argumentam que o pronome pessoal *we* pode ser utilizado na escrita acadêmica como uma forma de conduzir o leitor em um argumento ou como uma forma de se criar um senso de comunidade dividida com todos os participantes do discurso.

Por fim, a frequência geral e o contexto em que os pronomes pessoais de sujeito se encontram também mostrou que o pronome *they* é utilizado em frequência alta. O trecho (142) é de um ensaio sobre imaginação no mundo moderno industrializado.

(142) *Following this further, we may think of scientists as the artists of reality. **They**'ve created lots of utilities to make our lives easier and longer, and **they** also disclose some of the world's mysteries that kept humanity awake for a long time thinking of it's meanings. These scientific idealizations, so this argument goes, do not only popped up in the air for **them** to catch **it** and use **it**, **they** had to think about **it** for a long time, to plan, to use their imagination. Furthermore, so many scientific theories came up from dreams or scientists imagination, before **they** could create the theory or the thesis to explain that. **They** had to think about other theories or formulas, **they** had to use their imagination to think about new ones. Moreover, theories are just dreams and imagination, because **they** can not be proved yet, and even when **they** are reality nobody can erase the fact **they** were a dream one time.*

No exemplo (142), o pronome pessoal *they* é utilizado 9 vezes no mesmo parágrafo, juntamente com vários outros pronomes pessoais, como *it* e *them*. O pronome *they* faz referência a SNs diferentes no desenvolver do parágrafo. Embora seja possível encontrar os referentes dos pronomes pelo contexto, a leitura do trecho fica prejudicada por causa de tantos pronomes utilizados ao mesmo tempo. Essa sobre utilização de pronomes acaba tornando o texto ambíguo.

5.5 Os complementos dos sintagmas preposicionados modificadores de SNs

Os sintagmas preposicionados são uma unidade linguística comumente encontrada modificando SNs. Conforme discutido no capítulo 2, esses sintagmas possuem dois elementos obrigatórios: a preposição e o complemento, que pode ser realizado por um SN ou por uma oração. Também foi discutida, nesse capítulo, a possibilidade de ocorrência de sintagmas encaixados sucessivamente, conforme exemplo a seguir.

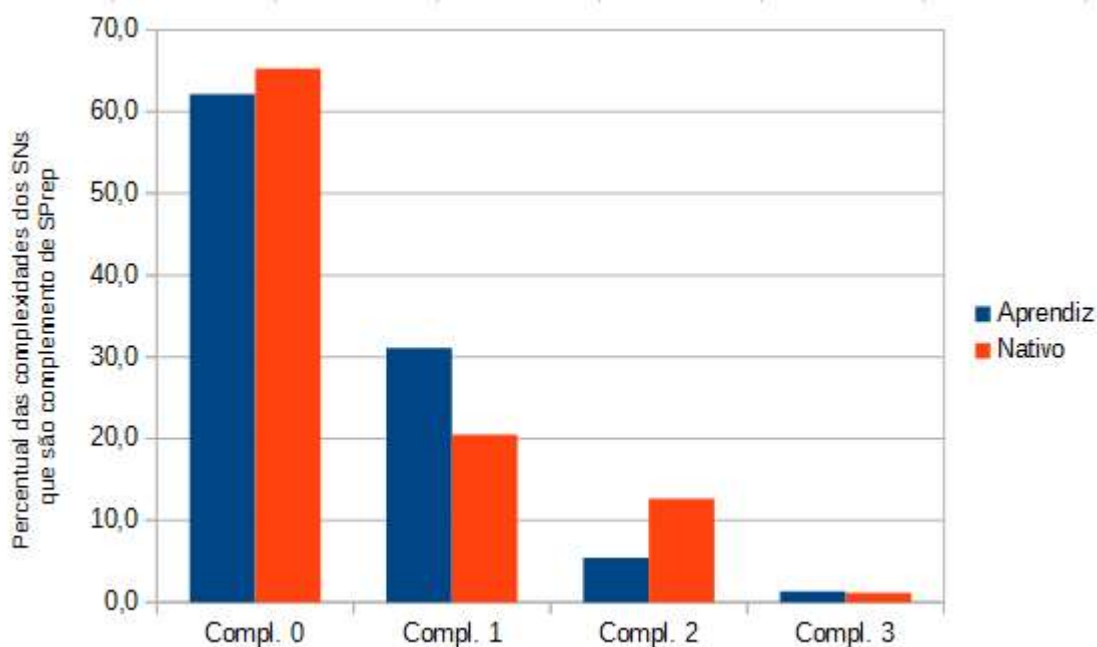
(143) *for one such report to be taken to heart it would need **the backing of the houses of parliament** (LOCNESS).*

Para ilustrar o encaixamento sucessivo de SPreps apresentado no exemplo (143), no exemplo (144) o SN em negrito do exemplo (143) foi dividido de acordo com cada novo SN. O núcleo de cada SN se encontra em negrito.

(144) *the **backing*** núcleo do SN1 (complexidade 2)
*of the **houses*** núcleo do SN2 (complexidade 2)
*of **parliament*** núcleo do SN3 (complexidade 0)

Tendo em vista essa possibilidade de sucessivos SPreps encaixados em outros SNs, os pós-modificadores dos SNs de complexidade 2 e 3 realizados por SPreps foram analisados nesta seção. Essa análise foi feita para verificar com qual frequência os aprendizes produziram SNs com sucessivos SPreps encaixados (SPrep seguido de SPrep), conforme exemplo. No gráfico 15, a maior parte dos complementos de preposição são realizados por SN de complexidade 0, seguido pelos SNs de complexidade 1, 2 e finalmente 3.

Gráfico 15 – Complexidade dos SNs com função de complemento de sintagma preposicionados, modificadores de SN



O gráfico 11 mostra os tipos de complexidade em relação ao total de complementos de SPreps modificadores de SNs. A seguir, exemplos de cada tipo de complexidade. O SN de complexidade 2 se encontra sublinhado, e o SN, complemento do SPrep, encontra-se em negrito.

- Complexidade 0:

(145) Aprendiz: *Since the beginning of our race, we have tried to make use of nature in order to live longer and more comfortably.*

(146) Nativo: *Obviously the people without homes have no chance.*

- Complexidade 1:

(147) Aprendiz: *Most of the problems of natural resources came as the time passed by.*

(148) Nativo: *The housing for low income families has dropped as well.*

- Complexidade 2:

(149) Aprendiz: *These changes had effects on society, economy and the way of life of people from entire world, starting in Britain.*

(150) Nativo: *The feeling from individuals in the black community was mutal.*

- Complexidade 3:

(151) Aprendiz: *Use of death penalty is a violating of the fundamental right of life.*

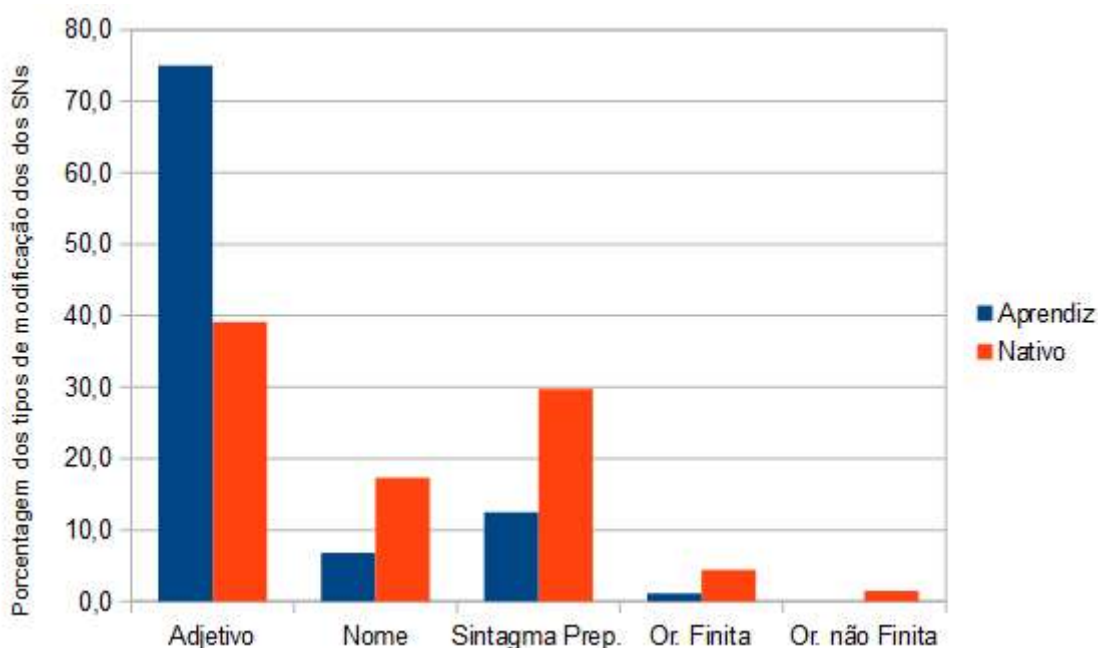
(152) Nativo: *(...) not only through Panglosses suffering but also the suffering of innocent people who are victims.*

A maior parte dos complementos de SPreps são realizados por SNs, representando 98,3% por parte dos aprendizes e 97,1% de SNs por parte dos nativos. As outras realizações de complemento de preposição foram realizadas por orações (1,7% por parte dos aprendizes e 2,9% por parte dos nativos).

Proporcionalmente, em relação aos SNs dos complementos de preposição, os aprendizes pós-modificaram os SNs muito menos (5,4%) do que os nativos (12,7%). E os aprendizes pré-modificaram mais SNs (31,1%) do que os nativos (20,5%). Mais uma vez, vemos que os aprendizes brasileiros desse *corpus* tendem a pré-modificar seus SNs mais do que os nativos (complexidade 1).

Além da tendência de pré-modificar mais os SNs dos complementos de preposição, os aprendizes e nativos também apresentam diferenças no que se refere ao tipo de elemento utilizado para fazer essa modificação, conforme mostrado no Gráfico 16.

Gráfico 16: Relação dos tipos de modificadores de SN com função de complemento de SPrep



O Gráfico 16 mostra a realização dos modificadores de SNs com função de complemento de SPrep. Os SNs que possuíam algum tipo de modificação, conforme mostrado anteriormente, representaram 37,1% dos SNs com função de complemento de preposição nos sintagmas produzidos por aprendizes e 33,3% nos produzidos por nativos.

Embora os aprendizes tenham apresentado maior frequência de modificação, grande parte dessa modificação foi feita por adjetivos, representando 75% das modificações (os nativos pré-modificaram 39% dos SNs com adjetivos). A pré-modificação realizada por substantivos representou 6,8% das modificações (os nativos pré-modificaram 17,3% dos SNs com substantivos). Já em relação à pós-modificação, os SPrep representaram 12,5% (os nativos pós-modificaram 29,7% dos SNs com SPreps), as orações finitas representaram apenas 1,1% (os nativos pós-modificaram 4,5% dos SNs com orações finitas) e nenhum caso de pós-modificação realizada por oração não finita foi encontrado (os nativos pós-modificaram 1,5% dos SNs com orações não finitas).

5.6 Discussão dos resultados obtidos

Quais são as características da escrita acadêmica dos aprendizes brasileiros de inglês no que se refere aos sintagmas nominais? Como a escrita desses aprendizes se compara com a escrita de falantes nativos do inglês? A análise de uma amostra de mais de 4.000 sintagmas nominais revelou que os aprendizes brasileiros produziram sintagmas nominais mais simples do que os nativos. E essa menor complexidade será explicada a seguir.

Ao classificar os SNs em níveis de complexidade (de 0 a 3) com base nos modificadores integrantes dessas estruturas linguísticas, verifiquei em quais aspectos os SNs produzidos por aprendizes são mais simples. A complexidade 0 se refere aos SNs sem modificação, a complexidade 1 aos SNs com pré-modificação, a complexidade 2 aos SNs com pós-modificação e a complexidade 3 aos SNs com pré- e pós-modificação ao mesmo tempo.

O primeiro indício dessa simplicidade ocorreu nos SNs produzidos com complexidade 0. Mais de 74% dos sintagmas nominais produzidos pelos aprendizes não possuem modificação (contra 68% dos sintagmas nominais produzidos pelos nativos). Isso significa que apenas 26% dos SNs produzidos pelos aprendizes possuíam complexidade 1, 2 ou 3 (os nativos modificaram 32% dos SNs produzidos). Os testes estatísticos aplicados também indicam uma diferença estatisticamente relevante entre a frequência das complexidades produzidas pelos aprendizes e pelos nativos. A maior frequência de SNs de complexidade 0 por parte dos aprendizes pode resultar em uma escrita mais genérica, conforme exemplos a seguir:

(153) Aprendiz: *to open a book and to imagine situations or to reflect about **the reading**.*

(154) Nativo: *been much more forceful, interventionist and political than **a reading of the consitution** would imply.*

O segundo indício de que os aprendizes produziram SNs mais simples foi em relação à diferença da frequência de SNs produzidos com complexidade 1 e 2. O *corpus* dos aprendizes apresentou maior frequência de pré-modificadores (complexidade 1) do que de pós-modificadores (complexidade 2) em todos os aspectos analisados. Os tipos de pré-modificadores também apresentaram diferenças estatisticamente significantes entre os aprendizes e os nativos. Em geral,

os aprendizes produziram mais SNs pré-modificados por adjetivos do que os nativos. Quando analisei os SNs de complexidade 2 para verificar a estrutura dos sintagmas preposicionados que modificavam esses SNs, percebi, novamente, que os aprendizes apresentaram menor frequência de incorporação de SNs mais complexos se comparados aos nativos.

Os aprendizes apresentaram baixa frequência de SNs pós-modificados realizados por SPreps. Outro tipo de pós-modificação menos utilizada pelos aprendizes foram as orações não finitas. Embora esse tipo de modificação tenha uma frequência mais baixa do que a modificação feita por SPreps e pelas orações relativas, os aprendizes apresentaram uma frequência de produção dessas estruturas abaixo de 50% em comparação com a produção dessas mesmas estruturas pelos nativos. O teste qui-quadrado também revelou diferenças estatisticamente relevantes com relação à frequência dos tipos de pós-modificadores produzidos por aprendizes e por nativos.

O tipo mais frequente de pós-modificação encontrado na produção dos aprendizes foi a pós-modificação por orações finitas (principalmente indicadas pelos pronomes relativos). Em estudo que buscou identificar as características da escrita acadêmica através dos sintagmas nominais, Biber, Gray e Poonpon (2011) concluíram que as orações relativas finitas são comuns tanto na língua oral quanto na língua escrita, e que a pós-modificação por SPreps e por orações não finitas é mais comum na escrita acadêmica. Os nativos apresentaram pós-modificação semelhante ao descrito pelos autores. Por outro lado, os aprendizes produziram mais orações finitas do que SPreps como pós-modificadores e apresentaram poucas orações não finitas como pós-modificadores em comparação com os nativos.

A análise dos complementos de preposição mostrou que os aprendizes produzem menos SPres encaixados do que os nativos e, como consequência, o número de SNs de complexidade maior (2 e 3) em um texto tende a ser menor. Além disso, foi encontrada uma prevalência de pré-modificação nesses elementos. Isso significa que os complementos de preposição produzidos pelos aprendizes são menos complexos, com prevalência das complexidades 0 e 1 relativas a esses complementos.

O presente estudo também mostrou que o aprendiz produz muito mais SNs com pré-modificação do que o nativo, mas, qualitativamente, esses pré-modificadores são, muitas vezes, estruturas fora da norma, formadas com adjetivos e substantivos. A construção formada por sintagma nominal com pré-modificação por substantivos, de fato, foi a que indicou a maior incidência dessas construções atípicas no *corpus* do aprendiz brasileiro. Esse fato foi notado por Parkison e Musgrave (2014) em estudo sobre estruturas N N produzidas por aprendizes de inglês (falantes nativos de outras línguas diferentes do português). Os autores notaram que os aprendizes que possuem língua materna que permite a pré-modificação de SN realizada por substantivos produziram mais estruturas N N do que aprendizes cuja língua materna não permite essa construção. No entanto, muitas dessas estruturas N N produzidas pelos aprendizes serão estruturas atípicas, fora da norma.

O presente estudo também mostrou que os aprendizes produziram menos SNs plenos e mais pronomes pessoais do que os nativos. O teste qui-quadrado mostrou que não há diferenças estatísticas no total de pronomes pessoais de sujeito e de objeto entre a produção dos aprendizes e nativos. O estudo mostrou, contudo, diferenças estatisticamente relevantes entre os tipos de pronomes pessoais de sujeito produzidos pelos aprendizes e pelos nativos. Analisando-se o contexto, os pronomes que mais aparentam ser sobre utilizados pelos aprendizes, gerando um texto ambíguo, são os pronomes pessoais *it* e *them*.

Todos esses resultados, em conjunto, indicam que, sob a ótica do SN e de suas complexidades, o aprendiz possui escrita menos complexa do que o nativo, com SNs menos densos. Os dados também mostraram que, em relação à pós-modificação, o aprendiz brasileiro depende mais de estruturas oracionais do que sintagmáticas em comparação com o nativo. Apesar disso, o aprendiz brasileiro é capaz de produzir todas as quatro complexidades, com todos os elementos que realizam tanto pré-modificadores quanto pós-modificadores. Ele apresentou uma ampla variedade lexical de pré-modificadores realizados por adjetivos atributivos e participio, apesar das construções atípicas produzidas.

Os pontos levantados por esses resultados serão levados em consideração para as sugestões de atividades didáticas para os aprendizes brasileiros de inglês apresentadas no próximo capítulo.

PROPOSTA DE ATIVIDADES DIDÁTICAS

Com base nos resultados encontrados no capítulo 5, proponho uma sequência de atividades que podem ser feitas em sala de aula com aprendizes de inglês. Essas atividades focam o sintagma nominal e suas complexidades e buscam trabalhar com o sintagma e seus modificadores, focando os pontos que os aprendizes apresentaram maiores diferenças em relação aos nativos.

Como o presente estudo analisou ensaios acadêmicos, busquei utilizar trechos de ensaios, ou de algum gênero científico, para as atividades propostas. Alguns trechos foram retirados de ensaios dos *corpora* utilizados no presente estudo (BAWE, LOCNESS e BR-ICLE), enquanto outros trechos foram extraídos de textos de jornais e revistas de língua inglesa que publicam textos científicos.

As atividades propostas visam ajudar os aprendizes a criarem consciência (*consciousness raising*) das estruturas sintáticas comuns à escrita acadêmica no que se refere às diferentes complexidades do SN. De acordo com Stewik (1980, p. 251 *apud* RUTHERFORD, 1987, p. 26-27), a conscientização “lança luz sobre os caminhos desconhecidos e os obstáculos arbitrários através dos quais [o aluno] deve, eventualmente, conseguir correr de um lado para o outro com os olhos fechados”.²⁹ Nos últimos anos, vários estudos têm apontado a importância da conscientização do aprendiz em relação a estruturas gramaticais da língua-alvo (RUTHERFORD, 1987; SCHMIDT, 1990; ELLIS, 1993; GILAKJANI, AHMADI, 2011). Esses autores defendem que o aprendizado por conscientização pode contribuir para a aquisição de segunda língua ou língua estrangeira.

29 No original: “*consciousness awareness* casts light on the unfamiliar pathways and the arbitrary obstacles through which [the student] must eventually be able to run back and forth with his eyes shut”.

6.1 Trabalhando com o conceito de modificação

A primeira sugestão nas atividades propostas é trabalhar o conceito de modificação de SN. Estudos como o de Mazgutova e Kormos (2015) mostram que, com apenas um mês de curso intensivo de leitura de textos acadêmicos e escrita acadêmica, a escrita dos alunos de inglês para fins acadêmicos melhorou consideravelmente. Os alunos desse curso produziram sintagmas nominais mais complexos ao seu final. Minha proposta, no entanto, é trabalhar diretamente com a identificação dos elementos do sintagma nominal para criar consciência da variedade de elementos e complexidades do SN.

As atividades que apresento nesta seção são uma mistura de atividades que fiz com alunos particulares de inglês, atividades apresentadas aos alunos no curso de Sintaxe do inglês, que ministrei na Universidade Federal de Minas Gerais em conjunto com outros professores, e também uma atividade sugerida por O'Maley (2013). O foco desse tipo de atividade é a identificação de SNs mais complexos e a análise de seus elementos. Para isso, é necessário trabalhar com os alunos a estrutura do sintagma nominal, as noções de núcleo, modificadores e determinantes. O mais importante nesse tipo de atividade não é a nomenclatura, mas, sim, a função que um determinado elemento tem no sintagma.

Sugiro ensinar a noção de modificação para os alunos, pois ampliar o conhecimento explícito que os aprendizes possuem acerca de uma língua e seu funcionamento pode ajudá-los a lidar melhor com textos acadêmicos (FANG, SCHLEPPEGRELL, COX, 2006). O objetivo é também fazer com que eles reflitam sobre o discurso acadêmico e suas características. E, para alcançar esses objetivos, o professor pode oferecer tarefas que engajem os alunos em “atividades baseadas na língua que tenham o objetivo de desenvolver sua consciência da estrutura e da lógica dos registros acadêmicos”³⁰ (FANG, SCHLEPPEGRELL, COX, 2006, p. 269). Para isso, é interessante que o aluno pratique a identificação, em um texto, de vários tipos de SNs, com elementos diferentes, abrangendo todas as complexidades. Conforme descrito nos capítulos anteriores, o SN possui como elementos núcleo,

30 No original: “language based tasks that aim at developing their awareness of the structure and logic of academic registers”.

modificadores e determinantes. E a complexidade do SN do presente estudo se baseia na presença ou não de modificadores. O presente estudo sugeriu uma classificação com 4 níveis diferentes de complexidade, a saber: 0 (sem modificação), 1 (com pré-modificação), 2 (com pós-modificação) e 3 (com pré- e pós-modificação).

O objetivo da atividade número 1 é levar à atenção do aluno a possibilidade dos tipos diferentes de complexidade do sintagma nominal. Essa atividade é feita a partir de um trecho de um ensaio extraído do *corpus* LOCNESS, escrito por um falante nativo de inglês.

Trecho para a atividade 1

The flames and rage that characterized racial relations in the late 1960's and early 70's in America seemed to have returned in the fall of 1992. Riots on college campuses were in the news again. Students at some southern schools insisted that black cultural centers be built at the heart of campus. They said there was a need for greater cultural awareness and complained that administrators built these centers in places where they would not get too much attention. The students felt that administrators were merely trying to quell their complaints without recognizing the true necessity for such centers.

Nessa atividade, proponho pedir aos alunos que encontrem um ou mais exemplos de sintagma nominal com as características destacadas, e que completem o quadro a seguir:

Característica	determinante	Pré-modificador	núcleo	Pós-modificador
Sem modificadores	<i>the</i>		<i>news</i>	
Com 1 pré-modificador	<i>some</i>	<i>southern</i>	<i>schools</i>	

Com 1 pós-modificador	<i>the</i>		<i>fall</i>	<i>of 1992</i>
Com 1 pré-modificador e 1 pós-modificador	<i>the</i>	<i>true</i>	<i>necessity</i>	<i>for such centers</i>

Esse tipo de atividade pode ser realizado com qualquer texto, já que todo texto apresenta certa variedade de SNs. Por isso, sugiro, também, que essa atividade seja feita com um texto produzido pelo próprio aluno. Dessa forma, o aprendiz pode tomar ciência da própria escrita, ver o quão complexos são os SNs produzidos por ele e compará-los com os SNs do texto de escolha do professor.

Uma outra forma de se trabalhar a identificação de SNs é pela pré-seleção de estruturas do texto.

Trecho para a atividade 2³¹

According to Haas, the philosophers personify their books. I find this interpretation intuitive, because the fresco was painted on a wall of the Stanza della Segnatura, then the pope's personal library. It also hints at a greater truth about the study of classic books. To portray the great thinkers, Raphael employed illusionism, a technique that makes the figures appear to be in the same room as the viewer. Thus, the viewer becomes not simply a passive observer of preserved knowledge, but a participant in ongoing conversations that span countries and centuries.

31 O trecho da atividade 2 é o primeiro parágrafo de um ensaio publicado no jornal *The Atlantic*. O jornal, juntamente com uma escola de Ensino Fundamental, criou uma competição internacional da qual alunos de mais de 44 países diferentes participaram. Os participantes desse concurso deveriam fazer uma leitura atenta de uma obra de arte que os inspirasse e escrever um ensaio sobre a obra escolhida. O ensaio vencedor seria publicado no jornal. O vencedor desse prêmio foi um aluno de Ensino Médio do Vietnã. Vale notar que este tipo de iniciativa serve para motivar alunos a escrever. O ensaio vencedor completo pode ser visto no anexo 3.

Nessa atividade, o professor pode escolher um texto de sua preferência, de acordo com o nível dos alunos. Eu sugiro que os alunos escrevam os SNs sublinhados nos quadros de acordo com o tipo de modificação que eles recebem.

Característica	determinante	Pré-modificador	núcleo	Pós-modificador
Sem modificadores	<i>the</i> <i>their</i>		<i>philosophers</i> <i>books</i> <i>it</i>	
Com 1 pré-modificador	<i>the</i> <i>the pope's</i>	<i>great</i> <i>personal</i>	<i>thinkers</i> <i>library</i>	
Com 1 pós-modificador	<i>a</i> <i>a</i>		<i>technique</i> <i>participant</i>	<i>that makes the figures appear to be in the same room as the viewer</i> <i>in ongoing conversations that span countries and centuries.</i>
Com 1 pré-modificador e 1 pós-modificador	<i>a</i> <i>a</i>	<i>greater</i> <i>passive</i>	<i>truth</i> <i>observer</i>	<i>about the study of classic books</i> <i>preserved knowledge</i>

Note que esse tipo de atividade pode ajudar o aluno a entender principalmente aqueles SNs muito grandes, com vários elementos, como pós-modificadores com sucessivos elementos encaixados.

Sugiro também uma terceira forma de se trabalhar com a identificação de SNs com diferentes complexidades em um texto.

Texto para atividade 3³²

Abstract

The rise of state dissidence has challenged the hegemony of Western liberalism on the international relations stage. Russia's ongoing involvement in the Ukraine crisis is a case in point. Russia's dissidence threatens not only the already fragile European order, but also the potency of liberalism as a system of international norms. Hence, a great deal of attention has been given to trying to determine the possible failures and solutions of global governance in dealing with Russia. In contrast, this article argues for the need to understand state resistance from the perspective of the dissenting state. By drawing upon Carl Schmitt's influential critique of globalizing liberalism, the article attempts to analyse what Russia's resistance reveals about the subtle mechanisms of global liberal governance. On the basis of Schmitt's theory, the article establishes that Russia's dissidence can be an attempt to preserve state sovereignty and its unique "way of life", as well as state pluralism on the global arena. In fact, to eradicate conflict, liberal governance attempts to suppress state pluralism as a potential cause of conflict. In the long run, however, this risks provoking radical resistance in response. The article then analyses the "hybrid" strategy of Russia's resistance employed in the Ukraine crisis, based on which it identifies the major weaknesses of liberal governance. The article concludes that the inadequacy of international law to deal with unconventional forms of warfare and refusal to acknowledge the possibility of animosity can significantly debilitate liberal governance. This article is published as part of a collection on global governance.

Na atividade 3, sugiro que os alunos, a partir de um texto de escolha do professor ou de escolha do próprio aluno, selecionem os SNs que desejem analisar e escrevam os SNs com seus modificadores nos quadros de acordo com o tipo de modificação que eles recebem.

32 Esse resumo faz parte de um artigo publicado na revista *Nature* e pode ser acessado no seguinte link: <<https://www.nature.com/articles/palcomms201696>>.

determinante	pré-modificador	núcleo	Pós-modificador

Nessa atividade, o aprendiz precisará identificar os SNs e seus elementos totalmente, já que o quadro não possuirá nenhum preenchimento e o texto não apresentará nenhuma marcação de SNs específicos.

6.2 Criando sintagmas nominais complexos a partir de textos autênticos

Nesta proposta de atividade, utilizei um texto do gênero divulgação científica publicado no periódico *Nature*. Selecionei um texto que possui grande quantidade de sintagmas nominais mais complexos (complexidades 2 e 3) com muitos elementos.

A proposta desta atividade é que os alunos identifiquem a incorporação de uma grande quantidade de informação em um único SN, baseando-se em SNs produzidos em textos autênticos. Para tanto, alguns SNs do artigo foram selecionados. A partir da informação contida nesses SNs, orações foram criadas para serem mostradas aos alunos. Dessa forma, o aluno, a partir de uma estrutura oracional, precisa trabalhar com as informações fornecidas para transformá-las em uma estrutura sintagmática. Ou seja, eles podem comprimir uma grande quantidade de informação em um único SN ou em um número mais reduzido de SNs.

Para a atividade, sugiro pedir que o aluno crie um SN mais complexo que contenha toda a informação do SN do texto original, conforme sugerido a seguir:

Atividade 4: Trecho para criação do SN 1

James Carpenter just needed some fake Moon dirt. Carpenter, _____, works on a drill designed to hunt for buried ice on the Moon. His team recently ordered half a tonne of powdery material to replicate the lunar surface from a commercial supplier

in the United States.

Crie um SN que contenha as seguintes informações:

1. Carpenter is a lunar-exploration expert. He works at the European Space Agency (ESA). This agency is in Noordwijk, the Netherlands.

A resposta sugerida nesta atividade é o SN original produzido no texto, embora outras estruturas também sejam possíveis, conforme trecho mostrado a seguir.³³

Sugestão de resposta para o SN 1

James Carpenter just needed some fake Moon dirt. **Carpenter, a lunar-exploration expert at the European Space Agency (ESA) in Noordwijk, the Netherlands,** works on a drill designed to hunt for buried ice on the Moon. His team recently ordered half a tonne of powdery material to replicate the lunar surface from a commercial supplier in the United States.

O trecho em negrito possui um SN *Carpenter* e um SN apostro *a lunar-exploration expert at the European Space Agency (ESA) in Noordwijk, the Netherlands*. Como pode ser observado, o apostro é composto por uma série de informações concentradas em apenas um SN. Dentre as informações contidas nesse sintagma, temos a área de especialidade do pesquisador, o nome da agência em que ele trabalha e o local (cidade e país) em que a agência se encontra.

Com esse exercício, o aluno poderá criar SNs com a complexidade que o professor definir. O SN 1, por exemplo, é o sujeito da oração e terá complexidade 3, além de possuir um pós-modificador composto por uma série de SPreps encaixados.

Atividade 5: Texto para criação do SN 2

³³ Artigo publicado na revista *Nature*, que pode ser acessado em <<http://www.nature.com/news/why-planetary-scientists-want-better-fake-space-dirt-1.22228>> e no anexo 4.

In Europe, Carpenter and his colleagues are still hunting for their perfect lunar soil — but they have given up on ordering it commercially. The researchers, who need 700 tonnes for _____, are looking much closer to home. They have decided to grind up rocks from the nearby basalt mines of the Eifel region.

Crie um SN que contenha as seguintes informações:

The researchers need 700 tonnes for a habitat. This habitat was planned. It also is lunar. It is at ESA's astronaut-training centre. The centre is in Cologne, Germany.

Nesse trecho, o SN 2 criado será de complexidade 2, e pode possuir uma oração relativa e pelo menos 3 sintagmas preposicionados.

No próximo exemplo, o SN 2 é de complexidade 2 e também contém muita informação concentrada no pós-modificador, conforme trecho a seguir.

Sugestão de resposta para o SN 2

In Europe, Carpenter and his colleagues are still hunting for their perfect lunar soil — but they have given up on ordering it commercially. The researchers, who need 700 tonnes for **a planned lunar habitat at ESA's astronaut-training centre in Cologne, Germany**, are looking much closer to home. They have decided to grind up rocks from the nearby basalt mines of the Eifel region.

O SN2 possui como núcleo *habitat*, que é pré- e pós-modificado. O pré-modificador informa o tipo de habitat. Já o pós-modificador do SN 2 informa o local da construção.

Como a escolha do texto fica a critério do professor, ele consegue controlar o gênero e nível do texto, assim como a complexidade dos SNs que ele deseja que os alunos construam. Além de atividades como as 5 e 6 propostas, o oposto também pode ser feito. A partir de um SN destacado, o aluno pode tentar identificar toda a informação contida naquele SN, conforme atividade 6, a seguir.

Atividade 6: Trecho com SN para ser analisado

His experience underscores a longstanding problem with artificial space soils, known as simulants: how to make them consistently and reliably. But now there is a fresh effort to bring the field into line. Last month, NASA established **a team of scientists from eight of its research centres** to analyse the physical properties and availability of existing simulants.

Identifique as informações contidas no SN destacado.

Sugestão de resposta: Nasa established a team of scientists. The scientists are from eight research centers. These centers belong to NASA.

Esse tipo de atividade também pode ser feito com SNs complexos que o próprio aluno selecionou a partir da leitura de um texto.

Neste capítulo, foram propostas 2 atividades diferentes. A primeira sugere o ensino da noção de modificação e das 4 diferentes complexidades de um SN. A segunda sugere a criação de SNs complexos a partir de SNs encontrados em textos autênticos, publicados por revistas e jornais de língua inglesa.

Espero que, com essas atividades sugeridas, o aluno não só tenha contato com as diferentes complexidades de um SN, mas também entenda como tanta informação pode ser armazenada em um único sintagma. Além disso, ele terá a chance de produzir SNs complexos, revisar o próprio texto e refletir sobre uma característica importante da escrita acadêmica: o SN e suas complexidades.

CONSIDERAÇÕES FINAIS

Foi mostrado, nos capítulos iniciais desta tese, que um dos critérios para se medir a complexidade da escrita acadêmica é a complexidade dos sintagmas nominais dessa escrita (BIBER, POONPON, GRAY, 2011; PARKINSON, MUSGRAVE, 2011). Vale lembrar que o que caracteriza a complexidade dos sintagmas nominais é sua estrutura, seus elementos componentes, principalmente os modificadores. Propus um critério de classificação das complexidades do SN e, a partir dessa classificação, descrevi os sintagmas nominais produzidos por aprendizes brasileiros. Essa análise foi feita em um contexto acadêmico, comparando a produção dessa unidade sintática em ensaios acadêmicos de brasileiros com a produção de ensaios acadêmicos produzidos por nativos.

Para a realização da pesquisa, foram utilizados recursos tecnológicos próprios da linguística de *corpus* (BIBER, 1999; GRANGER, 2002; SARDINHA, 2004). Na metodologia deste trabalho, foram usadas ferramentas como o *WordSmith Tools* e o COCA para a seleção, análise e comparação dos sintagmas nominais produzidos por aprendizes brasileiros. As estruturas gramaticais consideradas em nossa análise foram descritas em gramáticas da língua inglesa, baseadas em *corpus* ou na língua em uso (QUIRK, 1998; DOWNING, 2015; CARTHER, MCCARTHY, 2015). Os trabalhos que também utilizaram *corpus* de aprendizes para a descrição da complexidade dos sintagmas nominais e que também nortearam esta pesquisa foram Pastor (2008), Biber, Gray e Poonpon (2011), O'Maley (2013), Parkinson e Musgrave (2014), Parkinson e Musgrave (2015), dentre outros.

O trabalho aqui descrito teve como objetivo analisar quantitativa e qualitativamente os tipos de pré e pós-modificadores de SN produzidos por aprendizes brasileiros de inglês e compará-los aos SNs produzidos por falantes nativos da língua inglesa. A análise quantitativa se deu a partir da identificação do tamanho dos SNs, ou seja, sua complexidade estrutural. Já a análise qualitativa se

deu com a ajuda das linhas de concordância e do contexto estendido das ocorrências dos SNs, através da observação das estruturas atípicas, ou fora da norma, e da variedade lexical. Esses objetivos visavam responder às seguintes perguntas:

1. Os aprendizes brasileiros produzem SNs complexos em seus textos na língua inglesa?
2. Como as estruturas dos SNs produzidos pelos aprendizes se comparam com as estruturas dos SNs produzidos por falantes nativos?

Em relação à primeira pergunta, verifiquei que os aprendizes brasileiros produzem, sim, sintagmas nominais mais complexos. Na primeira parte do estudo, dividi as complexidades em 4 tipos: complexidade 0 para SNs sem modificação, complexidade 1 para SNs apenas com pré-modificação, complexidade 2 para SNs apenas com pós-modificação e complexidade 3 para SNs com pré- e pós-modificação ao mesmo tempo.

Os aprendizes produziram SNs em todos os tipos de complexidade, prevalecendo o tipo 0, seguido das complexidades 1, 2 e, por último, 3. Contudo, a hipótese de que os aprendizes produziriam menos SNs de complexidade 3 do que os nativos não se confirmou, visto que a produção dos dois grupos foi muito próxima. Além disso, descobri um fato inesperado: o SN de complexidade 3 possui frequência baixíssima nos dois *corpora* analisados. Isso se deve, provavelmente, ao fato de ambos os *corpora* serem de produção de estudantes e não de profissionais. Verifiquei, também, que os aprendizes brasileiros produziram mais estruturas fora da norma, relacionadas à pré-modificação de SNs do que à pós-modificação de SNs, principalmente com adjetivos atributivos derivados de verbos e substantivos.

E quanto à segunda pergunta, ou seja, como os sintagmas nominais produzidos pelos brasileiros se comparam aos sintagmas nominais produzidos pelos falantes nativos, minha hipótese inicial foi confirmada. Em vários aspectos, a escrita dos aprendizes se mostrou mais simples. Além disso, diferenças significativas foram confirmadas com o teste estatístico Qui-quadrado.

O objetivo do presente estudo foi descrever o sintagma nominal produzido pelo aprendiz brasileiro, mais especificamente na complexidade dos SNs no que se refere aos modificadores. As principais características encontradas estão descritas no quadro 7.

Quadro 7: Resumo das características do sintagma nominal produzido pelos aprendizes

	Descrição
Pronome	Alta frequência de pronomes pessoais, principalmente <i>it</i> e <i>they</i> .
Complexidade 0	Alta frequência de SNs sem modificação.
Complexidade 1	Mais SNs pré-modificados do que pós-modificados.
	SNs majoritariamente pré-modificados por adjetivos.
	Baixa frequência de pré-modificação substantiva.
	Sufixação atípica nos pré-modificadores, principalmente com os pré-modificadores substantivos.
Complexidade 2	Baixa frequência de pós-modificação em relação ao total de SNs.
	Alta frequência de pós-modificação realizada por orações finitas.
	Baixa frequência de pós-modificação realizada por SPreps.
	Baixa frequência de pós-modificação realizada por orações não finitas, principalmente <i>-ing</i> e infinitivo <i>to</i> .
	Uso fora da norma do pronome relativo <i>whom</i> .
Complexidade 3	Frequência baixa de uso da complexidade 3, semelhante à dos nativos.

O presente estudo permitiu descrever a densidade do SN produzido pelo aprendiz brasileiro. Os resultados deste trabalho apontaram uma alta produção das complexidades menores (0 e 1, menor densidade) e uma baixa produção das complexidades maiores (2 e 3, maior densidade) por parte dos aprendizes brasileiros. A partir dessas características descritas, o professor pode trabalhar para

desenvolver, nos aprendizes, a capacidade de condensar informações em apenas um SN, o que resultaria no fato de esse SN ser de maior complexidade.

Os resultados da pesquisa não foram suficientes para encontrar uma resposta para o porquê da baixa frequência da complexidade 3. Uma hipótese para essa baixa frequência de complexidade 3 é que nenhum dos *corpora* utilizados neste trabalho é de escritores acadêmicos profissionais, mas, sim, de estudantes universitários. Outra hipótese é que no gênero ensaios acadêmicos, a frequência do tipo 3 é mais baixa que em outros gêneros acadêmicos. Considero que esse tema pode ser mais aprofundado em trabalhos futuros.

O presente estudo também analisou o uso de pronomes pelos aprendizes e nativos como um complemento aos objetivos deste trabalho. Como o pronome raramente recebe algum tipo de modificação (CARTER, MCCARTHY, 2015), ele geralmente realiza SNs de complexidade 0. Por isso, uma alta frequência dos pronomes seria mais um indício de que os aprendizes tendem a produzir SNs menos complexos. Os testes estatísticos não confirmaram uma diferença relevante entre o uso de pronomes pessoais de sujeito e de objeto entre nativos e aprendizes, mas mostrou diferenças estatisticamente relevantes entre os tipos de pronome pessoal de sujeito utilizados. Considero que esse tema pode ser mais aprofundado em trabalhos futuros.

Outro complemento aos objetivos deste trabalho foi a análise da densidade lexical de modificador realizado por SPrep. Conforme discutido anteriormente, quanto mais sintagmas e orações são encaixados em uma unidade linguística, maior é a sua complexidade. E os aprendizes produziram muito menos SPrep com complemento de complexidade 2 e mais complementos de complexidade 1 do que os nativos. Esse foi mais um indício de que o SN produzido pelo aprendiz é menos complexo.

A partir das atividades propostas no capítulo 5, pretendi mostrar a professores maneiras de se trabalhar com a complexidade do SN. As atividades visam criar uma conscientização nos alunos das complexidades que a estrutura do SN pode apresentar. E, a partir do reconhecimento de SNs complexos, ter, também, a oportunidade de produzi-los.

Como o SN não se restringe apenas ao texto acadêmico, um campo ainda a ser explorado é a comparação da complexidade do SN entre gêneros diferentes da escrita. Essa comparação pode ser feita envolvendo gêneros dentro da escrita acadêmica, como resumo, artigos científicos, relatos de pesquisa, resenhas, dentre outros. E comparações entre outros gêneros, como textos jornalísticos, crônicas, artigos de opinião, quadrinhos, dentre outros.

Os resultados apresentados no presente estudo são relevantes em várias perspectivas. Eles têm aplicação pedagógica, visto que a descrição das características do aprendiz brasileiro referentes à modificação do SN pode ser utilizada para auxiliar a criação de materiais para o ensino do inglês para aprendizes brasileiros. Além disso, os resultados têm relevância prática, visto que a pesquisa apresenta atividades que podem ser utilizadas em sala de aula com o intuito de desenvolver a habilidade do aprendiz de produzir SNs de complexidades diversas. Por fim, este trabalho tem relevância teórico-descritiva, visto que é uma descrição detalhada de como um aprendiz brasileiro de inglês estrutura o SN e seus modificadores, contribuindo para a descrição da língua do aprendiz de inglês.

REFERÊNCIAS

- ALMEIDA, V. C. **Investigando colocações em um corpus de aprendiz**. 2014. Tese (Doutorado em Estudos Linguísticos). Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2014.
- BARDOVI-HARLING, K. A second look at T-unit analysis: Reconsidering the sentence. **TESOL Quarterly**, v. 26, p. 390-395, 1992.
- BENASSI, C. M; FLORES, S. P; SOBRERO, M. E; STEFANUK, M. L; BENASSI, M. C; MAYOL, C. E. The impact of the complex noun phrase in reading comprehension of English Scientific texts. **Rev. Cienc. Tecnol**, n. 16, p. 13-20, 2011.
- BERBER SARDINHA, T. Linguística de *Corpus*: Histórico e Problemática. **DELTA**, São Paulo, v. 16, n. 2, 323-367, 2000.
- BERBER SARDINHA, T. **Pesquisa em Linguística de Corpus com Wordsmith Tools**. Mercado De Letras, Campinas. 2004.
- BÉRTOLI, P. P; SHEPHERD, T. Escrita Acadêmica: Um estudo exploratório de quadrigramas. **ESpecialist**, 2015, v. 36, n. 2, p. 241-262, 2015.
- BIBER, D. A Corpus Driven Approach to formulaic Language in English. Multi-word Patterns in Speech and Writing. **International Journal of Corpus Linguistics**, v. 14, n. 3, p. 235-311, 2009.
- BIBER, D., GRAY, B., & POONPON, K. Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? **TESOL Quarterly**, v. 45, n. 1, 2011, p. 5-35, 2011.
- BIBER, D; CLARK, V. Historical shifts in modification patterns with complex noun phrase structures. How long can you go without a verb? **Current Issues in Linguistic Theory**, v. 1, 2002, p. 43-66.
- BIBER, D; GRAY, B. Grammatical change in the noun phrase: the influence of written language use. **English Language and Linguistics**, v. 15, n. 2, p. 223-259, 2011.
- BIBER, D; GRIEVE, J; IBERRI-SHEA, G. Noun phrase modification. In: ROHDENBURG, G; SCHLÜTER, J. (Eds.). **One language, two grammars? Differences between British and American English**. Cambridge: Cambridge University Press, 2009, p. 183-193.

BIBER, D; JOHANSSON, S; LEECH, G; CONRAD, S; FINEGAN, E. **Longman Grammar of Spoken and Written English**. Harlow: Pearson Education Limited, 1999.

BONELLI, E. T.; Theoretical overview of the evolution of corpus linguistics. In: OKEEFFE, A; MCCARTHY, M. (Ed.). **The Routledge handbook of Corpus Linguistics**. Abingdon: Routledge, 2010, p. 14-28.

CARTER, R.; MCCARTHY, M. **Cambridge Grammar of English: A Comprehensive Guide-Spoken and Written English Grammar and Usage**. 8. Ed. Cambridge: Cambridge University Press, 2015.

COOPER, T. Measuring Written Syntactic Patterns of Second Language Learners of German. **The journal of educational research**, v. 69, p. 176-183, 1976.

COSTA, D. D. **Linking adverbials in applied linguistics research articles: a corpus-based study**. Dissertação (Mestrado em Estudos Linguísticos). 2015. Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

DELEGÁ-LÚCIO, D. **A variação entre textos argumentativos e o material didático de inglês**: aplicações de análise multidimensional e do Corpus Internacional de Aprendizes de Inglês (ICLE). 2013. Tese (Doutorado em Estudos Linguísticos), Faculdade de Letras, Pontifícia Universidade Católica de São Paulo, São Paulo, 2013.

DOWNING, A.; LOCKE, P. **English Grammar: A university Course**. 2. Ed. Londres: Taylor & Francis e-Library, 2006.

DOWNING, A. **English Grammar: A university Course**. 3. Ed. Londres: Taylor & Francis e-Library, 2015.

DUTRA, D. P; SILERO, R. P. Descobertas linguísticas para pesquisadores e aprendizes: a Linguística de Corpus e o ensino de gramática. **Revista Brasileira de Linguística Aplicada**, Belo Horizonte, v. 10, n. 4, p. 909-930, 2010.

ELLIS, R. **Task-based language learning and teaching**. Oxford: Oxford University Press, 2003.

ELLIS, R; YUAN, F. The effects of planning on fluency, complexity, and accuracy in second language narrative writing. **Studies in Second Language Acquisition**, ELSEVIER, v. 26, p. 59-84, 2004.

EVISON, J. What are the basics of analysing a corpus? In: OKEEFFE, A; MCCARTHY, M. (Eds.). **The Routledge handbook of Corpus Linguistics**. Abingdon: Routledge, 2010, p. 122-135.

FANG, Z; SCHLEPPEGRELL, M. J; COX, B. E. Understanding the Language Demands of Schooling: Nouns in academic registers. **Journal of literacy research**, v. 38, n. 3, p. 247-273, 2006.

FERRIS, D. R. Lexical and Syntactic Features of ESL Writing by Students at Different Levels of L2 Proficiency. **TESOL Quarterly**, v. 28, n. 2., p. 414-420, 1994.

FLAHIVE, D; SNOW, B. Measures of syntactic complexity in evaluating ESL compositions. In: OLLER, J. W; PERKINS, K. (Eds.). **Research in language testing**. Rowley: Newbury House, 1980, p. 171-176.

GEA-VALOR, M. L; REY-ROCHA, J; MORENO, A. I. Publishing research in the international context: An analysis of Spanish scholars' academic writing needs in the social sciences. **English for specific purposes**, v. 36, p. 47-59, 2014.

GILAKJANI, A. P; AHMADI, S. M. Role of Consciousness in Second Language Acquisition. **Theory and Practice in Language Studies**, v. 1, n. 5, p. 435-442, 2011.

GIPPS, C; EWEN, E. Scoring written work in English as a second language. **Educational Research**, n.16, p. 121-125, 1974.

GRANGER, S. The computer learner corpus: a versatile new source of data for SLA research. In: **Learner English on Computer**. Amsterdam: John Benjamins Publishing Company, 1998, p. 3-18.

GRANGER, S. A Bird's-eye View of Learner Corpora Research, In: GRANGER, S; HUNG, J; PETCH-TYSON, S. (Eds.). **Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching**. Amsterdam: John Benjamins, 2002, p. 3-36.

GRANT, L; GINTHER, A. Using Computer-Tagged Linguistic Features to Describe L2 Writing Differences. **Journal of Second Language Writing**, v. 9, n. 2, p. 123-145, 2000.

GREENBAUM, S; QUIRK, R. **A Student's Grammar of the English Language**. Essex: Longman, 2004.

HALLIDAY, M. A. K. Language as system and language as instance: the corpus as a theoretical construct. In: SVARTIK, J. (Org.) **Directions in corpus linguistics: Proceedings of Nobel Symposium 82**. Berlim/New York: De Gruyter, 1992, p. 61-78.

HALLIDAY, M. A. K. Some grammatical problems in scientific english. In: HALLIDAY, M. A. K; MARTIN, J. R. (Eds.). **Writing Science**. London: The Falmer Press, 1993, p. 159-180.

HALLIDAY, M. A. K. **Computational and Quantitative Studies**. 1. ed. Londres: Continuum, 2005.

HALLIDAY, MATTHUSSEN. **Halliday's introduction to functional grammar**. 4. ed. Londres: Routledge, 2014.

HIATT, M. P. The feminine style: Theory and fact. **College Composition and Communication**, n. 29, p. 222-226, 1978.

HUDDLESTON, R. **Introduction to the grammar of English**. Cambridge: Cambridge University Press, 1984.

HUDDLESTON, R; PULLUM, G. K. **A Student's Intro to Grammar**. 3. ed. Cambridge: Cambridge University Press, 2015.

HUNDT, M; DENISON, D; SCHNEIDERR, G. Relative complexity in scientific discourse. **English Language and Linguistics**, v. 16, n. 2, p. 209-240, 2012.

HUNSTON, S. **Corpora in applied linguistics**. Cambridge: Cambridge University Press, 2002.

HUNSTON, S. How can a corpus be used to explore patterns? In: OKEEFFE, A; McCARTHY, M. (Eds.). **The Routledge handbook of Corpus Linguistics**. Abingdon: Routledge, 2010, p. 152-166.

HYLAND, K.; MILTON, J. Qualifications and certainty in L1 and L2 students' writing. **Journal of Second Language Writing**, v. 6, n. 2, p. 183-205, 1997.

ISHIKAWA, S. Objective measurement of low-proficiency EFL narrative writing. **Journal of Second Language Writing**, v. 4, n. 1, p. 51-69, 1995.

KOESTER, A. Building small specialised corpora. In: OKEEFFE, A; McCARTHY, M. (Eds.). **The Routledge handbook of Corpus Linguistics**. Abingdon: Routledge, 2010, p. 66-79.

LARSEN-FREEMAN, D. The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. **Applied Linguistics**, v. 27, p. 590-619, 2006.

LEECH, G. The state of the art in corpus linguistics. In: ALJMER, K.; ALTENBERG, B. (Org.). **English corpus linguistics: Studies in honour of Jan Svartvik**. Londres: Longman, 1991, 8-29

LEECH, G; SVARTVIK, J. **A Communicative Grammar of English**. 3. Ed. Londres: Pearson Education Limited, 2002.

LU, X. A Corpus-Based Evaluation of Syntactic Complexity Measures as Indices of College-Level ESL Writers' Language Development. **TESOL QUARTERLY**, v. 45, n. 1, p. 36-62, 2011.

McCARTHY, M; O'KEEFFE, A. Historical Perspective: What are corpora and how have they evolved? In: O'KEEFFE, A; McCARTHY, M. (Eds.). **The Routledge handbook of Corpus Linguistics**. Abingdon: Routledge, 2010, p. 3-13.

NELSON, N; VAN METER, A. Measuring written language ability in narrative samples. **Reading and Writing Quarterly**, v. 23, p. 287-309, 2007.

O'KEEFE, A. The fourth Corpus Linguistics Conference, 2007, Birmingham. **The Pragmatics of Corpus Linguistics**. (Anais)

OLIVEIRA, A. L. A. M.; ORFANÓ, B.; TENUTA, A. Epistemic modality through the use of adverbs: a corpus-based study on learners' written discourse. **Letras & Letras (Online)**, v. 30, p. 104-121, 2014.

OLIVEIRA, J. H. **O uso de advérbios intensificadores na escrita de aprendizes brasileiros de inglês como L2: um estudo baseado em corpus**. Tese (Doutorado em Estudos Linguísticos). 2015, Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

OLIVEIRA, F. S. **Noun groups, their elements and their syntactic function in the clause: an analysis of abstracts of scientific texts in English**. 30 f. Monografia (Linguística Teórica e Descritiva). 2014. Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2014.

O'MALLEY, L. Grammatical complexity of noun phrases found in GED. Dissertação (Mestrado em Estudos Linguísticos). 2013. Faculdade de Artes, Departamento de Inglês como Segunda Língua, Hamline University, Saint Paul, 2013.

ORENHA-OTTAIANO, A. English collocations extracted from a corpus of university learners and its contribution to a language teaching pedagogy. **Acta Scientiarum. Language and Culture** (UEM), v. 34, n. 2, p. 241-25, 2012.

ORTEGA, L. **Understanding second language aquisition**. Londres: Hodder Education, 2009.

ORTEGA, L. **Second Language Aquisition**. Abingdon: Routledge, 2011.

PANA, F; REPPEN, R; BIBER, D. Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles. **Journal of English for Academic Purposes**, ELSEVIER. v. 21, p. 60-71, 2016.

PARKINSIN, J; MUSGRAVE, J. Development of noun phrase complexity in the writing of English for Academic Purposes students. **Journal of English for Academic Purposes**, ELSEVIER, v. 14, p. 48-59, 2011.

PARKINSON, J. Noun–noun collocations in learner writing. **Journal of English for Academic Purposes**, ELSEVIER, v. 20, p.103-113, 2015.

PARKINSON, J., MUSGRAVE, Development of noun phrase complexity in the writing of English for Academic Purposes students. **Journal of English for Academic Purposes**, Elsevier, n. 14, p. 48-59, 2014.

PASTOR, M. L. C. English complex noun phrase interpretation by Spanish learners. **RESLA**, v. 21, PÁGINAS, 2008.

PÉREZ-GUERRA. **English Historical Syntax and Morphology**. Amsterdam: John Benjamins Publishing, 2002.

RAMOS, L. R. “I think this question is very relevant nowadays...”: investigando categorias e padrões de modalidade na escrita de aprendizes brasileiros de língua inglesa. **Trab. Ling. Aplic.**, Campinas, v. 44 n. 1, p. 19-35, 2005.

RECKSKI, L. J. “It’s *Really Ultimately Very* Cruel”: contrastando colocações de advérbios de intensidade em inglês entre a escrita de aprendizes de EFL e o discurso acadêmico oral de falantes nativos. **DELTA**, São Paulo, v. 20 n. 2, p. 211-234, 2004

REPPEN, R. Building a corpus. What are the key considerations? In: O’KEEFFE, A; MCCARTHY, M. (Eds.). **The Routledge handbook of Corpus Linguistics**. Abingdon: Routledge, 2010, p. 31-37.

RUTHERFORD, W. E. **Second Language grammar: learning and teaching**. Londres: Longman, 1997.

SALIMI, A; DADASPOUR, S; ASADOLLAHFAM, H. The Effect of Task Complexity on EFL learners’ Written performance. **Procedia - Social and Behavioral Sciences**, v. 29, p. 1390-1399, 2011.

SINCLAIR, J. Corpus and Text: Basic Principles. In: WYNNE, M. (Eds.). **Developing linguistic corpora: a guide to good practice**. Oxford: Oxbow Books, 2005, p. 1-16.

SELINKER, L. Interlanguage. **International Review of Applied Linguistics**, v.10, p. 209-231, 1972.

TEIXEIRA, E. D. **A linguística de Corpus a serviço do tradutor**: proposta de um dicionário de culinária voltado para a produção textual. Tese (Doutorado em Estudos Linguísticos). 2008. Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2008.

TENUTA, A. M.; OLIVEIRA, A. L. A. M.; ORFANO, B. M. How Brazilian learners express modality in their writing: A corpus-based study on lexical bundles. **Intercâmbio**, v. 26, p. 1-15, 2012.

TENUTA, A. M.; OLIVEIRA, A. L. A. M.; ORFANO, B. M. How Brazilian learners express modality through verbs and adverbs in their writing: a corpus-based study on n-grams. **Documentação de Estudos em Linguística Teórica e Aplicada**, v. 31, p. 333-357, 2015.

VIANA, v. Modals in Brazilian Advanced EFL Learners' Compositions: A Corpus-Based Investigation. **Teachers' Professional Development**, v. 7, n. 1, p. 77-86, 2006.

WOLFE-QUINTERO, K; INAGAKI, H. K. **Second language development in writing**: Measures of fluency, accuracy, & complexity. Manoa: University of Hawaii Press, 1998.

WRIGHT, V. C. O. **Os sintagmas nominais do inglês**: Um estudo comparativo de corpus entre o uso de determinantes na produção escrita de falantes nativos do inglês e aprendizes brasileiros. Dissertação (Mestrado em Estudos Linguísticos). 2014. Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2014.

Anexo 1 - Títulos sugeridos pelo projeto ICLE

1	Crime does not pay
2	The prison system is outdated. No civilized society should punish its criminals: it should rehabilitate them
3	Most university degrees are theoretical and do not prepare students for the real world. They are therefore of very little value
4	A man/woman's financial reward should be commensurate with their contribution to the society they live in.
5	The role of censorship in Western society
6	Marx once said that religion was the opium of the masses. If he was alive at the end of the 20th century, he would replace religion with television
7	All armies should consist entirely of professional soldiers: there is no value in a system of military service
8	The Gulf War has shown us that it is still a great thing to fight for one's country
9	Feminists have done more harm to the cause of women than good
10	In his novel Animal Farm, George Orwell wrote "All men are equal: but some are more equal than others" How true is this today?
11	In the words of the old song "Money is the root of all evil"
12	Europe
13	In the 19th century, Victor Hugo said: "How sad it is to think that nature is calling out but humanity refuses to pay heed. "Do you think it is still true nowadays?
14	Some people say that in our modern world, dominated by science technology and industrialization, there is no longer a place for dreaming and imagination. What is your opinion?

Anexo 2 - Títulos sugeridos projeto LOCNESS para elaboração dos ensaios

Euthanasia		
Does affirmative action work?	Yoga	Nuclear power
Values and consequences of school interaction	Pride or segregation	Surrogate motherhood
Can we afford wellness	Prozac: the wonder drug	Homosexuality
Animal testing	Prayer in schools	Praying for a miracle
Sex equality	Teenagers	Aids
Orphanages	Profit: good or evil	Freedom of the press
Sex in schools	Welfare reforms needs a return to family values	The cost of grass
Abortion	Ethics	O.J. Simpson
Would anyone care for a drink	Cheating in colleges	Suicide
Feminists have done more harm to the cause of women than good	A single Europe: A loss of sovereignty for Britain	A man / woman's financial reward should be commensurate with their contribution to the society in which they live
Money is the root of all evil	Adolescent suicide	Water pollution
Legalization of marijuana	Homelessness	The welfare system
Divorce	Gender roles in our society	Crime does not pay

Anexo 3: Os nomes verbais analisados no corpus BR-ICLE

	Nome Verbal	Freq. Bruta		Nome Verbal	Freq. Bruta
1	beginning	42	28	covering of	1
2	feeling	22	29	distancing	1
3	being	11	30	dreaming	1
4	meaning	9	31	drinking	1
5	warning	5	32	filling	1
6	learning	5	33	functioning	1
7	saying	5	34	heating	1
8	living	4	35	helping	1
9	building	3	36	highlighting	1
10	calling	3	37	hiring	1
11	changing	3	38	measuring	1
12	following	3	39	positioning	1
13	marketing	3	40	questioning	1
14	programming	3	41	reading	1
15	spreading	3	42	reasoning	1
16	breaking	2	43	recovering	1
17	increasing	2	44	reshaping	1
18	misunderstanding	2	45	returning	1
19	misusing	2	46	rising	1
20	soothing	2	47	seasoning	1
21	teaching	2	48	setting	1
22	arising	1	49	thinking	1
23	brushing	1	50	violating	1
24	chasing	1	51	warning	1
25	coming	1	52	weelbeing	1
26	controlling	1	53	working	1
27	cooking	1	54	writing	1

Anexo 4: Os nomes verbais analisados no corpus LOCNESS

Nome Verbal	F. Bruta	Nome Verbal	F. Bruta	Nome Verbal	F. Bruta
1 beginning	49	46 televising	2	91 individualizing	1
2 feeling	28	47 thinking	2	92 kidnapping	1
3 meaning	22	48 uprising	2	93 longing	1
4 reasoning	12	49 abandoning	1	94 manufacturing	1
5 living	9	50 aching	1	95 mapping	1
6 killing	7	51 advancing	1	96 misunderstanding	1
7 passing	7	52 ageing	1	97 murdering	1
8 building	6	53 appointing	1	98 overcrowding	1
9 setting	6	54 beating	1	99 overruling	1
10 suffering	6	55 blessing	1	100 padding	1
11 testing	6	56 boxing	1	101 partaking	1
12 banning	5	57 breaking down	1	102 protesting	1
13 ending	5	58 buying	1	103 publishing	1
14 making	5	59 churc hing	1	104 rating	1
15 rulling	5	60 c lothing	1	105 Re-ordering	1
16 saying	5	61 c lustering	1	106 rearing	1
17 taking	5	62 c oming	1	107 renounc ing	1
18 reading	4	63 c ondemning	1	108 retelling	1
19 running	4	64 c ontrolling	1	109 revolving	1
20 understanding	4	65 c ooking	1	110 rioting	1
21 being	3	66 c raving	1	111 sacrificing	1
22 breeding	3	67 dismantling	1	112 selling	1
23 c hanging	3	68 diversify ing	1	113 shaping	1
24 funding	3	69 drafting	1	114 shooting	1
25 giving	3	70 drawing up	1	115 shopping	1
26 meeting	3	71 dropping	1	116 signing	1
27 painting	3	72 dumping	1	117 skyrocketing	1
28 pounding	3	73 dwelling	1	118 smashing	1
29 warning	3	74 dying	1	119 stealing	1
30 altering	2	75 easing	1	120 sterling	1
31 awarding	2	76 eating	1	121 streaming	1
32 backing	2	77 fettering	1	122 studying	1
33 drowning	2	78 fighting	1	123 suing	1
34 gathering	2	79 filming	1	124 surrendering	1
35 grounding	2	80 flogging	1	125 sweling	1
36 healing	2	81 foming	1	126 switching	1
37 housing	2	82 founding	1	127 taking awaly	1
38 lifting	2	83 getting	1	128 teaching	1
39 lowering	2	84 giving up	1	129 thawing	1
40 moving	2	85 gouging	1	130 tracking	1
41 opening	2	86 grouping	1	131 trying	1
42 performing	2	87 hacking	1	132 unfolding	1
43 planning	2	88 handling	1	133 uniting	1
44 programming	2	89 Hanging	1	134 widening	1
45 spending	2	90 hoarding	1	135 writring	1

Anexo 5: Ensaio utilizado nas atividades sugeridas do capítulo 6.



Reading Raphael in Hanoi

Critical inquiry is discouraged in modern-day Vietnamese schools. But, one student argues, there's an encouraging trend of young people seeking out alternative viewpoints from around the world.

THANH T. NGUYEN SEPTEMBER 2016 ISSUE EDUCATION

The second annual student Writing Prize, a competition organized by The Atlantic and the College Board, challenged high-school students to do a close reading of a work of art that inspires them. More than 2,000 students from 44 countries entered. Two dozen professors of composition and art history narrowed the field to 20 semifinalists, and a panel of judges from the College Board and The Atlantic made the final selection. The winner, Thanh T. Nguyen, lives in Hanoi, Vietnam, and will attend Duke University this fall.

Aristotle is earthly; Plato, otherworldly. The thinkers are secular, but the architecture is Christian. The greatest High Renaissance artists—Michelangelo, Leonardo, Bramante—merge with the greatest minds of ancient Greece. Light floods the fresco, the light of an age of classical-antiquity revival, of the artist-humanist, of spiritual and intellectual reconciliation. Initially, I believed *The School of Athens* could be understood only as the glorious rediscovery of the ancient Western tradition by Renaissance Europeans. What value has it for a Vietnamese like me, brought up in a fundamentally different time, culture, and education system? Yet closer study revealed that the painting represents an ideal of learning as universal as it is inspiring: relentless, unhindered critical inquiry that crosses physical, cultural, and disciplinary boundaries. Resonating from ancient Athens through Renaissance Italy to modern-day Hanoi, where I live, the ideal depicted in Raphael's *School* salvages the hope that, despite cultural and political restrictions, love of knowledge will prevail.

The title *The School of Athens* is a misnomer that conceals Raphael's broad vision of learning. Banners over the fresco, which was painted between 1509 and 1511 under the commission of Pope Julius II, read *Causarum Cognitio*, meaning "Seek Knowledge of Causes." With that, Raphael seems to suggest that what unites the figures is neither an institution nor prestige, but a common quest for knowledge and wisdom ("knowing why," as Aristotle put it). The work is not so much a glorification of the great thinkers of antiquity as an extolment of the process of learning itself.

Raphael conveys the life of the mind through physical activities. Nowadays, we tend to take this conception of intellectual engagement for granted; back in the early 16th century, however, it was revolutionary. According to Glenn W. Most in "Reading Raphael: *The School of Athens* and Its Pre-Text," in the tradition of personifying the seven liberal arts in painting, philosophy had usually been portrayed as an idealized woman flanked by male philosophers. Though alluding to this tradition by painting an enthroned lady on a tondo above the fresco, Raphael separated her completely from the central scene, allowing Plato and

Aristotle to dominate. While scholars have focused on the contrast and balance between Plato and Aristotle—everything from their hand gestures to the colors of their clothing—what impresses me the most is Raphael's decision to emphasize not an idealized personification of knowledge, but ordinary mortals immersed in the process of seeking knowledge through demonstrating and debating.

In fact, while having fun identifying each figure, I was surprised at the pluralism and inclusiveness of Raphael's vision of learning. By some counts, only a third of the figures hail from ancient Athens; others lived far beyond that time or place. Raphael even included thinkers from outside the Western canon, including Zoroaster, a prophet from ancient Persia, and Averroës, a medieval Islamic scholar and an interpreter of Aristotle's philosophy.

These two figures also challenge the common understanding that the various thinkers can be neatly separated along the idealism-realism axis between Plato and Aristotle. Averroës was a brilliant physician, physicist, and astronomer, yet he is placed on the side of the idealist Plato, and is consulting Pythagoras, who is often associated with mysticism. Similarly, Zoroaster, who resembles an astrologer holding a celestial globe, represents the practice of studying the stars and planets in order to divine human affairs—thus blending heavenly and earthly studies. All of this seems to recognize the spirit of inquiry as being fully inclusive, crossing temporal, geographical, cultural, and disciplinary borders.

Inspiring as that vision of learning may be, I still wondered where the viewer fits in. Should we simply gaze upon these great figures with admiration for their relentless pursuit of knowledge? Robert Haas, in *Raphael's School of Athens: A Theorem in a Painting?*, offers an edifying answer: "Raphael is here painting not a historical gathering, but a Renaissance library collection." According to Haas, the philosophers personify their books. I find this interpretation intuitive, because the fresco was painted on a wall of the *Stanza della Segnatura*, then the pope's personal library. It also hints at a greater truth about the study of classic books. To portray the great thinkers, Raphael employed illusionism, a technique that makes the figures appear to be in the same room as the viewer. Thus, the viewer

becomes not simply a passive observer of preserved knowledge, but a participant in ongoing conversations that span countries and centuries.

With this interpretation in mind, I've come to view Raphael's fresco as an invitation to engage with the great thinkers of the past—to learn from or question their ideas. Unfortunately, such an invitation has been largely withheld from students in Vietnam. Here, ideological constraints in the modern education system along with a tradition of rote learning have resulted in limited exploration of a narrow range of ideas. For instance, in mandatory high-school civics courses, “philosophy” is introduced only to advance Marxism-Leninism. Ideas seem to be taught not so much for critical study as for nation building, through the promotion of a single worldview.

But rather than being pessimistic, I am inspired by the notion that *The School of Athens* represents a personal library, for it suggests that if we recognize the need to unlearn the habits of our formal schooling, and strive to be independent thinkers and engage with books, we can still attain true wisdom. Moreover, the painting's historical context gives me hope that in the future, the pursuit of knowledge in Vietnam might flourish. Just as prosperity and leisure time enabled ancient Athenians to contemplate high-minded ideas, and a surge of wealth due to trade ushered in the Renaissance in Italy, Vietnam's rapid economic growth might be changing its intellectual scene. More individuals are pursuing education and employment abroad, then bringing back new ideas and methods of learning for younger generations. An increasing number of foreign-language learners means that works by a wider range of thinkers can be translated, examined, and discussed. Though we have far more work ahead, it is promising to see debate organizations and experimental-learning communities sprouting up. In Hanoi, for instance, an informal organization called the Integrity Network has brought together a group of leading theorists and students in the humanities and social sciences to form Reading Circle Vietnam, whose participants discuss books and debate with experts on a variety of topics. The questions posed, on topics ranging from democracy to liberty to feminism to the validity of Vietnamese

nationalism, keep getting bolder and encourage contrasting perspectives. Although still small in scope, these initiatives seem to be pioneering a trend of which I am happy to be a part.

This encouraging sign proves that rather than mindlessly absorbing a single perspective, more and more Vietnamese youths are interested in seeking alternative viewpoints from around the world, especially in the realms of history and politics. Yet we can still do better in extending Raphael's ideal of critical inquiry into our intellectual endeavors. For example, recently, members of a club of high-school and college students interested in international relations, including me, began reading the works of Henry Kissinger. We admired his effective realpolitik and noble-minded interpretation of American diplomacy. Indeed, there is something to learn from Kissinger's efforts to end the Vietnam War and normalize U.S. relations with China. However, upon further research, I find it troubling that there was no debate or even acknowledgment among us of Kissinger's actions (some critics have called them war crimes) in supporting secret bombing in Cambodia, massacres in what was then East Pakistan, and oppressive regimes in Latin America. Perhaps due to a lingering tendency to accept the authoritativeness of "great ideas," instead of carefully examining them, we simply treated his thoughts with reverence. Our approach was far from the one immortalized in Raphael's painting.

The most insightful lesson I derived from *The School of Athens* is that the great thinkers are there not for us to respect unquestioningly, but rather for us to question respectfully. For me, the work is powerful ultimately because it reminds me that even under my country's circumstances, individuals who love learning can work toward the ideal it depicts. The universal spirit of intellectual freedom and critical inquiry cannot be subdued, and while I look forward to seeing the painting in person someday, I will always strive to establish a personal School of Athens wherever my quest for knowledge takes me.

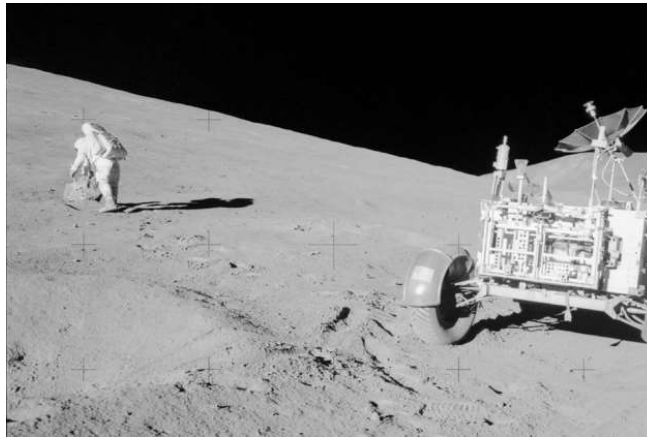
Anexo 6: Artigo utilizado nas atividades sugeridas do capítulo 6

Nature | News

Why planetary scientists want better fake space dirt

Artificial soils that mimic the surfaces of the Moon, Mars and asteroids are hard to make — and often miss the mark.

Alexandra Witze 07 July 2017



NASA - An Apollo 15 astronaut collects a soil sample on the Moon in 1971.

James Carpenter just needed some fake Moon dirt. Carpenter, a lunar-exploration expert at the European Space Agency (ESA) in Noordwijk, the Netherlands, works on a drill designed to hunt for buried ice on the Moon. His team recently ordered half a tonne of powdery material [to replicate the lunar surface](#) from a commercial supplier in the United States. But what showed up was not what the team was expecting. “The physical properties were visibly different,” says Carpenter.

His experience underscores a longstanding problem with artificial space soils, known as simulants: how to make them consistently and reliably. But now there is a fresh effort to bring the field into line. Last month, NASA established a team of scientists from eight of its research centres to analyse the physical properties and availability of existing simulants. And, for the first time, an asteroid-mining company in Florida is making scientifically accurate powders meant to represent

the surfaces of four classes of asteroid. It delivered its second batch to NASA on 28 June.

“NASA is trying to conquer the Wild West of simulants,” says Philip Metzger, a planetary scientist at the University of Central Florida in Orlando.

Such materials are meant to mimic the mix of dust and broken rock that covers the surfaces of planets and asteroids. Engineers use the artificial soils to test [space-exploration technologies such as drills and rovers](#), and to determine whether astronauts could make structures by feeding space dirt [into 3D printers](#) or by compressing it¹. Scientists use simulants to explore geological processes such as how rocks weather in space.

The dirt on dirt

Over the years, space agencies and research groups have tended to make their own artificial soils as needed from mixtures of ash and grit, sand and crushed bricks, and even glass beads. This has led to a wild proliferation of soils; there are more than 30 lunar simulants alone². “There are a lot of people out there creating their own simulant with no geology or materials-processing background,” says Jennifer Edmunson, a geologist at NASA’s Marshall Space Flight Center in Huntsville, Alabama.

But no artificial soil can re-create all the physical and chemical properties of a planet’s surface. A mixture that was developed for engineers to drive rovers in would probably be terrible for studying the geochemical properties of the Moon.

Researchers do not always pay attention to those limitations, says Clive Neal, a lunar scientist at the University of Notre Dame in Indiana. “We have no accreditation in terms of what this can be used for and what it can’t be used for,” he says. “If you use it for the wrong thing you end up with misleading results.”



Courtesy of ESA

The European Space Agency produced this lunar simulant at its European Astronaut Centre.

In 2010, a panel of lunar scientists recommended that NASA develop a database that researchers could use to compare the characteristics of different simulants and pick the best one for each use. But the agency had no money to support such a project. The new working group aims to outline how much it would cost to produce a database covering simulants for all types of planetary bodies. “Hopefully, we’ll be able to develop this repository,” says Brad Bailey, associate director of NASA’s Solar System Exploration Research Virtual Institute, who is based in Washington DC.

The database would include the four new asteroid simulants being made by the Orlando office of Deep Space Industries, an asteroid-mining company. NASA has ordered five tonnes over the next two years. Each simulant is based on a different class of meteorites called carbonaceous chondrites, which are thought to be chunks of asteroids.

Secret recipe

To make fake asteroid dirt, technicians mix various minerals — including bronzite, which is sourced from jewellery suppliers as polished stones — compress them into bricks and then pulverize them. “We have to do something that is basically equivalent to hitting a solid rock with thousands of meteorites over a long period of time,” says Stephen Covey, the company’s director of research and development.

Deep Space Industries delivered 512 kilograms of the first simulant to NASA in March, and 532 kilograms of the second type in June. The agency plans to use it in work on missions such as [OSIRIS-REx, a spacecraft that is making its way to an asteroid](#) to collect a sample and bring it back to Earth.

In Europe, Carpenter and his colleagues are still hunting for their perfect lunar soil — but they have given up on ordering it commercially. The researchers, who need 700 tonnes for a planned lunar habitat at ESA's astronaut-training centre in Cologne, Germany, are looking much closer to home. They have decided to grind up rocks from the nearby basalt mines of the Eifel region.

Journal name: Nature Volume: 547, Pages: 146–147 Date published: (13 July 2017) DOI: doi:10.1038/nature.2017.22228

Referências

1. Chow, B. J., Chen, T., Zhong, Y. & Qiao, Y. Sci. Rep. 7, 1151 (2017).
2. Taylor, L. A., Pieters, C. M. & Britt, D. Planet. Space Sci. 126, 1–7 (2016).