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

Danilo Duarte Costa

A corpus-based investigation of the language of Chemistry and Physics Textbooks and its Implications for *English for Reading*

Belo Horizonte

2020

DANILO DUARTE COSTA

A corpus-based investigation of the language of Chemistry and Physics Textbooks and its Implications for *English for Reading*

Tese apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Doutor em Linguística Aplicada.

Área de Concentração: Linguística Aplicada
Linha de Pesquisa: Ensino/Aprendizagem de Línguas Estrangeiras

Orientadora: Profa. Dra. Deise Prina Dutra

Belo Horizonte
Faculdade de Letras da UFMG
2020

Ficha catalográfica elaborada pelos Bibliotecários da Biblioteca FALE/UFMG

C837c

Costa, Danilo Duarte.
A corpus-based investigation of the language of Chemistry and Phisycs textbooks and ist implications for english for reading [manuscrito] / Danilo Duarte Costa. – 2020.
150 f., enc. : il., grafs, tabs (color)

Orientadora: Deise Prina Dutra.

Área de concentração: Linguística Aplicada.

Linha de Pesquisa: Ensino/Aprendizagem de Línguas Estrangeiras.

Tese (doutorado) – Universidade Federal de Minas Gerais, Faculdade de Letras.

Bibliografia: f. 135-141.

Apêndices: f. 142-145.

Anexos: f. 150.

1. Língua inglesa – Estudo e ensino – Teses. 2. Linguística de corpus – Teses. 3. Livros didáticos – Avaliação – Teses. 4. Professores de inglês – Formação – Teses. 5. Língua inglesa – Métodos de ensino – Teses. 6. Língua inglesa – Compêndios para estrangeiros – Teses. I. Dutra, Deise Prina. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.

CDD: 418

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS

FOLHA DE APROVAÇÃO

A corpus-based investigation of the language of Chemistry and Physics Textbooks and its implications for English for Reading

DANILO DUARTE COSTA

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGUÍSTICOS, como requisito para obtenção do grau de Doutor em ESTUDOS LINGUÍSTICOS, área de concentração LINGUÍSTICA APLICADA, linha de pesquisa Ensino/Aprendizagem de Línguas Estrangeiras.

Aprovada em 31 de março de 2020, pela banca constituída pelos membros:

Prof(a). Deise Prina Dutra - Orientadora
UFMG

Prof(a). Randi Reppen
Northern Arizona University

Prof(a). Valdênia de Carvalho e Almeida
UFV

Prof(a). Leonardo Pereira Nunes
UFMG

Prof(a). Luciana de Oliveira Silva
UFMG

Belo Horizonte, 9 de abril de 2020.

ACKNOWLEDGMENTS

I would like to thank, first and foremost, my advisor, Dr. Deise P. Dutra, for her absolute patience and understanding, countless support, and careful advice throughout the writing of this dissertation. To Dr. Dutra I owe my profound gratitude for all the guidance in helping me become the researcher I am now starting to be.

I would also like to sincerely thank Dr. Randi Reppen, who so kindly advised me during my studies at Northern Arizona University (NAU), in Flagstaff-AZ, a wonderful place I will certainly never forget. Dr. Reppen is unquestionably one of the most inspiring people I have had the pleasure to meet.

I am also very grateful to all the inspiring professors I have had at UFMG and at NAU for sharing unlimited knowledge and experience.

A special thanks to the members of the *Grupo de Pesquisa em Estudos de Corpora Especializados e de Aprendizes* (CEGEA) for the valuable comments in this research.

I am also very appreciative to *CAPES* and the Brazilian Ministry of Education for funding my studies at NAU.

Reading the world always precedes reading the word.
and reading the word implies continually reading the world.

Paulo Freire

ABSTRACT

This is a descriptive study that investigates lexical and grammatical features of the language found in chemistry and physics textbooks used in higher education. Our ultimate goal is to map out common patters in this register so as to inform teachers and materials developers who are interested in the teaching of English for Specific Purposes; more specifically, English for Reading in the context of higher education science/engineering courses. For this reason, we also created sample activities to illustrate possible applications of our findings.

Most frequently, books aimed at teaching English for Reading (in Brazil, commonly referred to as *Inglês Instrumental*) make use of texts taken from newspapers and magazines as a means of input. Hence, in order to carry out our investigation, a corpus of news texts taken from the British National Corpus (BNC) is compared with two other corpora, namely TB-Chem (Corpus of Chemistry Textbooks) and TB-Phy (Corpus of Physics Textbooks), that were compiled following strict methodological procedures based on the concepts of *representativeness* (Biber, 1993) and *balance* (McEnery, Xiao, Tono, 2006). The three corpora contain approximately 500 thousand tokens each, and comparisons are made within and across corpora. Results show that lexical, grammatical and rhetorical patters differ considerably between chemistry and physics on one side, and news on the other.

For instance, the most common lexical items in chemistry and physics are dramatically different from those found in news. In addition, n-grams (sequences of words that commonly co-occur) present similar patterns in chemistry and physics textbooks, which contrast to those found in news. As regards verb tense/aspect, the present simple is preferred in chemistry and physics textbooks, while news presents a higher number of occurrences of the past simple. Among all other tense/aspect markers (e.g. present perfect, past progressive) news shows higher frequencies and different discourse functions than do TB-Chem and TB-Phy. Finally, concerning phrasal grammatical complexity, in chemistry and physics, most attributive adjectives modifying nouns were found to belong to the category of classifiers, whereas in news these adjectives are more evenly distributed in the categories of classifiers and descriptors. As for nouns being modified by other nouns, we have found that the number and types of nouns occupying the N1 and N2 slots in chemistry and physics sharply contrast to those found in news. These findings show that other sources of input rather than newspapers/magazines articles may better suit pedagogical materials aimed at the teaching of English for Reading in discipline-specific contexts.

Keywords: corpus linguistics; English for Specific Purposes; textbook language; chemistry textbooks; physics textbooks.

RESUMO

Este é um estudo descritivo que investiga características lexicais e gramaticais da linguagem encontrada nos livros didáticos de química e física utilizados no ensino superior. Nosso objetivo final é mapear padrões comuns neste registro para informar professores e desenvolvedores de materiais que estejam interessados no ensino de inglês para fins específicos; particularmente, inglês para leitura no contexto dos cursos de ciências duras/engenharia do ensino superior. Por esse motivo, também criamos amostras de atividades para ilustrar possíveis aplicações de nossas descobertas. Frequentemente, os livros destinados ao ensino de inglês para leitura (no Brasil, comumente chamado de inglês instrumental) fazem uso de textos retirados de jornais e revistas como um meio de insumo. Portanto, para realizar nossa investigação, um corpus de textos de notícias retirados do *British National Corpus* (BNC) é comparado com outros dois corpora, a saber, TB-Chem (Corpus de livros-textos de química) e TB-Phy Corpus de livros-textos de física), que foram compilados seguindo rigorosos procedimentos metodológicos baseados nos conceitos de representatividade (Biber, 1993) e equilíbrio (McEnery, Xiao, Tono, 2006). Os três corpora contêm aproximadamente 500 mil palavras cada, e são feitas comparações dentro e entre corpora. Os resultados mostram que os padrões lexicais, gramaticais e retóricos diferem consideravelmente entre química e física, de um lado, e notícias de jornais, de outro. Por exemplo, os itens lexicais mais comuns em química e física são dramaticamente diferentes daqueles encontrados nas notícias. Além disso, n-gramas (sequências de palavras que comumente coocorrem) apresentam padrões semelhantes nos livros didáticos de química e física, que contrastam com os encontrados no corpus de notícias. No que diz respeito aos tempos verbais, o presente simples é preferido nos livros didáticos de química e física, enquanto as notícias apresentam um número maior de ocorrências do passado simples. Entre todos os outros marcadores de tempo / aspecto (e.g. *present perfect* e *past progressive*) o corpus de notícias mostra frequências mais altas e funções discursivas diferentes do que TB-Chem e TB-Phy. Por fim, no que diz respeito à complexidade gramatical frasal, em química e física, verificou-se que a maioria dos substantivos modificados por adjetivos atributivos pertencem à categoria de classificadores, enquanto nas notícias esses adjetivos são distribuídos de maneira mais uniforme nas categorias de classificadores e descritores. Quanto aos substantivos modificados por outros substantivos, observamos que o número e os tipos de substantivos que ocupam os posições N1 e N2 na química e na física contrastam fortemente com os encontrados nas notícias. Os resultados apontam para a necessidade de que outras fontes de insumo, em vez de artigos de jornais/ revistas, sejam utilizadas nos materiais pedagógicos voltados para o ensino de inglês para leitura.

Palavras-chave: linguística de corpus; Inglês para fins específicos; linguagem de livros-textos; livros-textos de química; livros-textos de física.

List of Figures

List of Tables

Table of Contents

CHAPTER 1: INTRODUCTION

1.1. Background context

Teaching English as foreign language has been part of my life since the beginning of the 2000's when I got my first job as a teacher. In those times, I had to face many challenges regarding not only my language skills per se, but also the best pedagogical practices to employ, as well as other basic difficulties commonly faced by any novice professional. Despite that, what I had in my favor – without being fully aware of it at the time – was the fact that I always had excellent materials in hand to teach those classes, and hardly ever did I have to come up with activities that were external to the adopted textbook.

More recently, however, a new scenario has come into play, when I started teaching what in Portuguese we refer to as *Inglês Instrumental* (English for Reading) at the undergraduate level. The challenge this time lies in the fact that there is a paucity of materials available for the teaching of this subject. In fact, there are some recognized published books, such as those by Dias (2002) and Fiori-Souza et al. (2005), but those are not specifically aimed at academic English. Using these textbooks in the context of higher education, more specifically in the hard sciences, which is where I stand, has proven to be an unsuccessful practice for me.

At the college level, students need to know English in order to have access to the language used in the academic context, i.e. academic prose produced through textbooks and research articles. This type of language presents specific characteristics that distinguishes it from other texts, such as newspapers and magazines (Hyland, 2002; 2004; 2009). For example, when comparing the methodology section of research articles in the hard sciences with those produced in the area of Applied Linguistics, the latter presents more anaphoric references and lexical repetition, which points to more explicitness in discourse (Swales, 1990).

Another example would be the fact that prepositional phrase embedding, which is a distinguishing characteristic of academic prose, makes the discourse more structurally complex (Biber & Gray, 2010). From a very simple main sentence, such as "X was observed", the embedding of several prepositional phrases makes it structurally complex, making its processing more difficult for the low-proficient reader, as in the example below taken from a scientific article in the area of Industrial chemistry.[1]

High enantioselectivity ***was*** *also* ***observed*** [*for the asymmetric reduction* [*of activated a,b-unsaturated enones* [*catalyzed* [*by pentaerythritol tetranitrate reductase* [*for reaction product stereogenic centers* [*at the beta-carbon atom*]]]]].

The vast majority of English for Reading textbooks do not incorporate academic prose in their syllabus (discussed in section 2.1). On the contrary, the reading input given to students mostly comes in the form of news articles (newspapers, magazines, and the like). This means that students are often exposed to language that they might not encounter in the course of their studies, which, in turn, limits their learning. As stated by Anthony (1998, p. 3) “when there are no textbooks available for a particular discipline, [teachers] resolve to teaching from textbooks which may be quite unsuitable”.

1.2. Overview of this study

In view of what has just been exposed, we propose a descriptive study that investigates the language found within and across textbooks in the areas of chemistry and physics through corpora specifically designed for this research. We believe that by mapping out common characteristics of this type of language, we can inform teachers and material developers who are interested in the teaching of both academic reading and writing to students of the hard sciences

[1] WOHLGEMUTH, R. (2010). Biocatalysis - key to sustainable industrial chemistry. *Current Opinion in Biotechnology*, v. 21, Issue 6, p. 713-724.

and engineering. For such, we approach this description from a lexical and a syntactic point of view.

From a lexical perspective, we investigate the most common lexical items (both content and function words) as well as the most common 4-word n-grams found in the corpora. This information would be useful for the inclusion of relevant vocabulary to be used in future pedagogical materials. From a syntactic perspective, we investigate verb tense and aspect, which is a common topic taught in English for Reading materials, as well as phrasal grammatical complexity indices discussed in Biber & Gray (2010), typically found in academic discourse.

Three corpora are used for this investigation. Two of them, made up of texts from college level textbooks in the areas of Physics and Chemistry, and a third corpus comprising text from the register of news for comparison purposes. A first decision regarded the choice of the types of texts that would represent chemistry and physics. We opted for textbooks rather scientific articles because the former tends to present more accessible language to learners, bringing basic concepts regarding the topics being taught, while the latter are written to be read by professionals of a certain field of study, who are expected to be already familiar with the topic under discussion. Since we intended to use the corpora not only for linguistics analysis but also as a source for pedagogical activities, we believed textbooks would be a better choice. Below we bring two excerpts that discuss *peptides* to illustrate this difference: the first from a textbook[2] and the second from a scientific article[3] in the field of organic chemistry.

1. *When the a-amino group of one amino acid is linked to the carboxyl group of a second amino acid by an amide bond, the product is called a peptide.*

2. *A classical example of racemisation is encountered in peptide synthesis when the terminal acid peptide is activated, leading to the formation of the corresponding*

[2] Ouellette, R. J., & Rawn, J. D. (2015). Principles of organic chemistry. Elsevier Academic Press.

[3] Montalbetti, C. A., & Falque, V. (2005). Amide bond formation and peptide coupling. Tetrahedron, 61(46), 10827-10852

oxazolone. Under mild basic conditions, the oxazolone undergoes racemisation via the formation of conjugated anionic intermediate..

To establish comparisons with chemistry and physics textbooks we chose the news register, and this has two reasons. First, news is the most common text register used in textbooks for teaching English for Reading (see section 2.1 for discussion). Hence, we would like to better understand what input students are being given, as opposed to what they could be getting. Secondly, because if we want to know what linguistic features are typical of textbook language, we ought to compare it to a different register so that differences may surface. As stated by Biber & Conrad (2009, p.36) "It is virtually impossible to know what is distinctive about a particular register without comparing it to other registers".

Finally, at the end of each section in the results chapter, we bring sample activities based on our corpora to illustrate possible applications of the findings for the teaching of English for Reading in the context of science/technology courses in higher education.

1.3. Objectives and Research Questions

In order to map out different linguistic features encountered in our corpora, we have defined the following objectives for this study:

1) Investigate what the most common lexical items (both content and function words) are in chemistry and physics textbooks, as well as in news (for compassion purposes), so as to be able to better understand the types of words learners are most commonly exposed to when reading these types of texts.

2) Examine the common word combinations (4-word n-grams) within and across the disciplines so as to be able to better understand word combinations learners encounter when reading.

3) Understand how tense and aspect behave in the chemistry and physics textbooks in terms of frequency and situational/discourse functions.

4) Study the most common phrasal grammatical complexity indices (Biber & Gray, 2010), namely *attributive adjectives as nominal pre-modifiers* and *nouns as nominal pre-modifiers* in terms of frequency and situational/discourse functions.

5) Produce sample pedagogical activities to illustrate possible applications of our findings.

These objectives, therefore, unfold into the following research questions:

- What are the most frequent lexical items found across chemistry, physics and news?
- What are the most common word combinations (4-word n-grams) within and across the registers studied?
- How do verb tense and aspect behave in the corpora?
- How are the phrasal grammatical complexity indices (Biber & Gray, 2010) distributed in the corpora?

CHAPTER 2: THEORETICAL FRAMEWORK AND LITERATURE REVIEW

We begin this chapter by bringing an overview on the teaching of English for Specific Purposes (ESP) in Brazil, focusing particularly on the teaching of *English for Reading*. We, then, contextualize each of the topics that are treated in our results section by discussing the theoretical perspectives that have been used to support our further discussions. Thus, we discuss issues on vocabulary teaching and learning, characteristics of n-grams, verb tense and aspect, and grammatical complexity indices.

2.1. ESP and English for Reading in Brazil

English for Specific Purposes (ESP), according to Paltridge & Starfield (2013, p.2), refers to the ”teaching and learning of English as a second or foreign language where the goal of the learners is to use English in a particular domain”. In this sense, ESP encompasses many areas of research and practice, including English for Academic Purposes (EAP), Business English, Legal English, and English for Medical Purposes, for example, which is why students in ESP courses are usually adult learners. Although the ESP teaching practice and research dates back to the 1960s (Anthony, 1998; Johns, 2013), when this approach was still under construction, it was in the 1980’s that it broadened its scope through the introduction of new concepts. These concepts included, for instance, needs assessment, which investigated students’ needs in order to guide course design, the relationship between linguistic devices and their rhetorical purposes, as well as genre and rhetorical moves, which would contrast different text types and the specific linguistic features they possessed (Johns, 2013).

In Brazil specifically, according to Celani et al. (1988), ESP became more prominent in the late 1970’s, with the establishment of the *Brazilian National ESP Project*, under the coordination of Professor Maria Antonieta Celani with the cooperation of the British Overseas Development Administration (now Department for International Development) through the British Council, which supported the project by assigning three Key English Language Teaching academics to assist in the Brazilian initiative. The project was quite ambitious in that it was aimed at organizing and implementing ESP courses at universities in a country of continental proportions through teacher training, materials preparation and the establishment of a resource center at Pontifícia Universidade Católica de São Paulo (PUC-SP). However, unlike what we see in ESP today, where different language skills are taught, the focus of the Brazilian ESP project was “in the area of English for Academic Purposes and especially the skills of *reading and writing*” (Holmes & Celani, 2006, p.111, emphasis added), but the focus in the classes was always on reading skills (Celani et al., 1988; Holmes & Celani, 2006; Ramos, 2008).

Due to the emphasis on the teaching of reading skills, a myth was created in Brazil regarding ESP. As Ramos (2008, p. 10) puts it:

> The most prevailing myth associated to ESP in Brazil, and created because of the Brazilian ESP Project, is that "ESP is reading". (...) reading was the only skill that deserved special attention in the Project. Thus, on one hand, ESP is to be understood as synonymous with reading and, on the other hand, any reading course is to be understood as ESP.

This myth goes even further when we see that there is also confusion involving terminology regarding the definition of ESP. In the 1980's, Celani et al. (1988, p. 33) stated that "ESP in Brazil is called instrumental" but at present this term is still widely used in Brazil to refer to courses aimed at the teaching of any foreign language reading skills. The Brazilian ESP Project itself was, in Portuguese, named "*Projeto Nacional Ensino de Inglês Instrumental*". As a result, for many, Inglês Instrumental ("Instrumental English") should be called ESP; nevertheless, Inglês Instrumental tends to be a course on reading skills, while ESP does not necessarily has to focus on reading. In either case, Inglês Instrumental seems to be the most adopted term to refer to English for Reading courses at universities in Brazil, though the term 'Inglês para Leitura' (English for Reading) started to be used at some institutions, especially after the 2000s. In this scenario, we maintain that ESP and Inglês Instrumental are not synonymous, but instead, Inglês Instrumental makes up one of the many areas that are situated within ESP teaching/research.

Terminology controversies apart, another point of interest here regards materials development for the purposes of the teaching of reading skills in the context of ESP. In the context of the Brazilian ESP project, Celani et al. (1988) explain that one of the development parameters was that no central or national textbook would be produced. This was due to the fact that teachers across the country had different levels of development, besides the great number of academic disciplines that would have to be covered. On the other hand, efforts were made to train teachers to produce their own materials, one of the project's main objectives. Samples of materials called 'Resource

Packages' (illustrated in Appendix B) were developed as a means of demonstrating the principles underlying the concepts proposed by the project. As described by Ramos (2008), those resource packages included information such as the exploration of authentic texts with beginners; the use of reading strategies; the teaching of grammar; vocabulary acquisition and reading strategies; and testing reading comprehension.

Although these materials were meant to be used for the purpose of the development of academic reading, texts used in the materials were not taken from academic prose. Holmes & Celani (2006), when discussing the nonexistence of a unique textbook for all teachers in the project, state that "the local ESP teams would continue to feel an acute need for useful materials, including reading texts such as articles from magazines or newspapers" (p. 113). This choice was probably due to beliefs at those times that claimed specialized texts were linguistically very similar to non-specialized texts. As argued by Hutchinson & Waters (1987, p. 161), "There is no clear relationship between sentence grammar and specialisation of knowledge" adding that "there is no grammatical structure, function or discourse structure that can be identified specifically with (...) any particular subject"(p.165).

As a result, almost every English for Reading textbook produced throughout the years tended to adopt the same strategy as that of the Brazilian ESP Project; i.e. select texts belonging to various registers within the scope of the so-called general English, with great emphasis on newspaper/magazine articles. Table 2.1 below gives an overview of the types of texts used in English for Reading courses in Brazil[4].

[4] Syllabi from courses in different Brazilian universities were consulted in order to see what books were most commonly adopted in English for Reading courses.

Table 2.1: Types of texts and topics included in English for Reading materials in Brazil.

Evaristo et al. (1996)	Inglês instrumental: Estratégias de leitura	**Types:** Mostly news articles taken from sources such as American Health, Newsweek, News Scientist, Scientific American, Time. **Topics:** human heath (birth control and contraceptives in developing countries, safer baby food, antioxidant vitamins, exercise and longevity, antibodies from milk, cancer), environmental issues (the extinction of animal species, illegal forestry), human behavior (gender equality, suicide) exploration of Jupiter, economies of scales.
Munhoz (2000)	Ingles instrumental Estratégias de Leitura . Módulo I	**Types**: horoscope, cooking recipe, classified ads, operation manual, classified ads, sports events, trip ads, news, travel guide, health brochure, poem. **Topics:** zodiac signs, horoscope, cooking, operating a facsimile, tourism, electricity market, anxiety, Rio de Janeiro, HIV, search engines, plane crash, rainforest
Munhoz (2001)	Ingles instrumental Estratégias de Leitura . Módulo II	**Types**: letter, email, fax, memo, ads, classified ads, news article, brochure. **Topics**: nutrition awareness, crime, smoking, youngsters' alcohol use, Alcatraz, Martin Luther King, motivation/self-help, stereotyping, female soccer players.
Oliveira (2003)	Para ler em inglês – desenvolvimento da habilidade de leitura.	**Types**: encyclopedic texts (mostly taken from Microsoft Encarta Encyclopedia). **Topics**: planets, nutrition facts label, processed foods, professions and the zodiac, habits, Madonna, Pompei, ice cream, Japan, The Beatles, laughter, zodiac/horoscope; earthquakes, weather, floods, first aids, Walt Disney, Princess Diana, the city of São Paulo.
Torres (2007)	Inglês instrumental para profissionais da saúde	**Types**: none (texts about grammar explanations with illustrative sentences) **Topics**: medical
Lima (2013)	Upstream: Inglês Instrumental Petróleo e Gás	**Types:** research articles (edited/adapted), academic textbooks. **Topics:** petroleum geology, weathering, erosion, sedimentary basins, classification of traps, petroleum formation, presalt play, petroleum exploration, oil exploration

2.2. Studies on the teaching of vocabulary through reading

The way students learn and retain vocabulary is still being disputed after many years of fruitful research on language acquisition. Since the 1980's there have been many studies dedicated to the teaching and learning of L2 vocabulary, and, in this respect, Laufer (2009) reviews a collection of such studies and lists them in a very concise but comprehensive paper that will serve as a basis for the discussion we wish to make in the beginning of this section.

The first study to be addressed here is by Krashen (1989). In his paper, he defends that vocabulary in a second language is learned in basically the same way we learn any other component of language, i.e. through comprehensible input, which is explained by his Input Hypothesis. For the case of spelling and vocabulary, he believes that the best sources of input would be reading – in and outside of school – as well as aural input. He also states that when we learn, we do so subconsciously, that is, we do not know we are learning (incidental learning). Krashen argues that acquiring vocabulary through the input coming from extensive reading (especially reading for pleasure) has proven to be more efficient – and easier – than through formal instruction via drilling or via attempts to produce language (Skill-Building Hypothes and Output Hypotheses).

In order to test whether second language learners could indeed acquire vocabulary from incidental reading, Pitts, White, & Krashen (1989) replicated a study in which they asked English L2 students to read two chapters from the book *A Clockwork Orange.* This book contained slang words of Russian origin (*nadsat words*) which were the words they tested, as *nadsats* were not present in the subjects' first languages. The study had originally been carried out with English L1 speakers by Saragi, Nation, & Meister (1978) in which students successfully learned the nadsat words. In Pitts, White & Krashen (1989), results were also positive, showing a small but reliable increase in the knowledge of the words tested.

These findings, however, are not necessarily positive. A study conducted by Nation & Wang (1999) shows that graded readers (literary books that are adapted to meet learners' different proficiency levels) can be a good source for vocabulary development in L2. Nevertheless, for that practice to be effective, students would have to be exposed to a large amount of information by reading at least one book per week during their course of study. Clearly, this is not an easy task to accomplish since reading that many books would represent considerable extra work to students.

Other studies focus on the use of lists of words aimed at teaching vocabulary. For example, Laufer & Schmueli (1997) report that presenting lists of words to students is an effect strategy for both short and long term vocabulary retention. The authors argue that when students read texts for meaning, too much context (i.e. long texts) may distract students from learning individual words, including new words, as these could go unnoticed throughout the reading process, especially if they are not essential for text comprehension. Since focus and attention are fundamental in transforming input into intake, the authors argue that less context (i.e. less distraction) via shorter sentences or wordlists would only benefit students' vocabulary acquisition, because "maximum attention would be directed to the word" (p. 92), which in in direct conflict to Krashen's (1989) study.

Webb & Nation (2017, p. 228) claim that wordlists are "perhaps the most useful resources for beginners and students who are learning a language for specific purposes". They argue that lists of high frequency words are a useful tool for guiding teachers and learners to the words that are most used, which impacts language development to a great extent. The authors observe, however, that for the purposes of teaching, selecting semantically related items may not be a good idea, as grouping together semantically-related words results in an increase in the level of interference between words, which makes them more difficult to be learned.

The most effective way of obtaining lists of high frequency words is through the analyses of a principled collection of texts (i.e. a corpus). One could use one corpus specifically or

by the comparison of two corpora. If we use one single corpus, we can find the most common words in that specific corpus. Or, if we use two corpora (a general corpus vs. a specialized corpus), we will be looking at words that not only are frequent in that corpus, but also are specific to it, the so-called Keywords, defined by Baker (2004, p.346) as "words that appear statistically more often in one text than the other".

Keyword analyses have been used in different types of studies, especially those on lexis/vocabulary, and teaching. For example, Foley (2009) presents practical applications of keyword lists in English for Academic Purposes (EAP) classes that intend to use a communicative language teaching approach. Seale et al. (2006) use keyword analysis to make a comparison between the language of men and women with cancer (breast cancer and prostate cancer), and then used this quantitative information to carry out an analysis focusing on the meanings of word clusters associated with keywords. Paquot (2010) describes the extraction procedure to create the Academic Keyword List (AKL), a list containing "potential academic words" (p. 30), which she then uses to analyze EFL learners' use of lexical devices in academic writing.

2.3. Overview on Lexical bundles

Lexical bundles are, according to Biber ae al. (1999, p.990), "sequences of word forms that commonly go together in natural discourse", like *on the other hand*, and *it is difficult to,* knowing that "most lexical bundles do not represent a complete structural unit" (Biber et al., 2004, p. 377), such as *one of the*, and *at the end of*. Although the term lexical bundles has been widely used in phraseological studies, these commonly co-occurring word combinations are given different names in the scientific literature. For instance, Gries & Mukherjee (2010) employ the term *n-grams*, Liu (2012) and Wood & Appel (2014) prefer the term *multiword constructions*, while O'keeffe, McCarthy, & Carter (2007) refer to these word combinations as *chunks*. In this study, we use terms n-grams and lexical bundles interchangeably, as it is done in Csomay & Cortes (2010).

Despite the difference in terminology, what these studies have in common is the fact that the extraction/selection of these multi-word units are most often done according

to frequency, although frequency is not the only criterium. Studies like Biber, Conrad, & Cortes (2004), Dutra & Berber-Sardinha (2013) and Pan, Reppen, & Biber (2016) use a frequency-driven approach, in which lexical bundles are selected according to both frequency (i.e. they are commonly found in a corpus) and dispersion/range (i.e. they do not frequently occur in just a few texts), while Simpson-Vlach & Ellis (2010) adopt a combination of quantitative and qualitative criteria, including "corpus statistics and linguistic analyses, psycholinguistic processing metrics, and instructor insights" (p.490) as the selection criteria.

Frequency criteria takes into account the number of times a certain multi-word unit occurs in a given corpus as an inclusion criterium for a study, the so-called cut-off point. In general, cut-off points range from 20 to 40, depending on researchers' objectives. Cortes (2004), and Byrd & Coxhead (2010), for instance, adopt a cut-off point of 20 times per million words, while Chen & Baker (2010) prefer 25 times per million, and 40 times per million in used in Pan, Reppen, & Biber (2016).

Another issue concerning the inclusion criteria is the length of word combinations. Lexical bundles are usually studied in the form of 2 to 6-word bundles, such as the following: 2-word bundles (*go to*, *have a*), 3-word bundles (*in relation to*, *the use of*), 4-word bundles (*on the other hand*, *the value of the*), and the more rare 5 and 6-word bundles (*there are a number of, you know what I'm, on the other hand it is*). Studies on lexical bundles, especially those investigating academic registers (e.g. lectures, textbooks, student writing, professional writing), typically look at 4-word bundles (see, for instance, Biber, Conrad & Cortes, 2004; Nesi & Basturkmen, 2006; Hyland, 2008, 2012; Byrd & Coxhead, 2010; Chen & Baker, 2010, 2014). The preference for 4-word bundles occurs, according to Hyland (2012, p.151) "perhaps because they are over 10 times more frequent than five-word sequences and offer a wider variety of structures and functions to analyze". Hyland (2008, p.8) also argues that 4-word bundles "offer a clearer range of structures and functions than 3-word bundles".

Lexical bundles (or n-grams) can also be classified according to the function they perform in discourse. Biber, Conrad & Cortes (2004) developed a taxonomy for

classifying lexical bundles by looking at their extended context and then grouping them according to their meaning and use. For these authors, bundles can be grouped under three main functions: *stance expressions*, *discourse organizers*, and *referential expression*s. These categories are further also subdivided as shown in Table 2.2.

Table 2.2: Functional classification of lexical bundles according to Biber et al. (2004).

Broad function	**Specific function**	**Subdivision of specific function**	**Example**
Stance expressions	Epistemic stance	-	*I don't know if, are more likely to*
	Attitudinal / modality stance	Desire	*if you want to, what do you want*
		Obligation/directive	*you don't have to, it is necessary to*
		Intention/prediction	*I'm not going to, are going to be*
		Ability	*to be able to, can be used to*
Discourse organizers	Topic introduction / focus	-	*if we look at, in this chapter we*
	Topic elaboration / clarification	-	*I mean you know, nothing to do with*
Referential Expressions	Identification / focus	-	*and this is a, those of you who*
	Imprecision	-	*or something like that*
	Specification of attributes	Quantity specification	*there's a lot of, a little bit more*
		Tangible framing attributes	*the size of the, in the form of*
		Intangible framing attributes	*the nature of the, in terms of the*
		Time/place/text reference	*in the United States, at the same time, as shown in figure*
		Multi-word reference	*the end of the, in the middle of*

Drawing on Biber et al.'s (2004) taxonomy, Simpson-Vlach & Ellis (2010) created a new classification for lexical bundles. The purpose of the study was to create a "pedagogically

useful list of formulaic sequences for academic speech and writing" (p. 487), which they called AFL (Academic Formulas List), a list of formulaic sequences for academic speech and writing, which is comparable to the AWL (Academic Word List) by Coxhead (2000). In this sense, this new classification maintained the three broad categories from the original study, but made modifications to them. For instance, for each category they identified the bundles that are core to the AFL, besides those that are primarily found in written or spoken discourse. In addition, they added a subcategory of contrast and comparison within the referential expressions group, and a subcategory of hedges and boosters, and evaluation within the stance expressions group. The authors also collapsed the categories of desire and intention/prediction into one, called volition/intention, and expanded and modified the discourse organizers group, thus making this a more extensive classification than Biber et al.'s, as seen in Table 2.3

.

Table 2.3: Functional classification of lexical bundles according to Simpson-Vlach & Ellis (2010)

Broad function	**Specific function**	**Subdivision of specific function**	**Example**
Referential expressions	Specification of attributes	Intangible framing attributes	*in the form of, (the) fact that (the)*
		Tangible framing attributes	*the amount of, the frequency of*
		Quantity specification	*is a number of, (a) wide range of (of)*
	Identification and focus	-	*a variety of, this means that*
	Contrast and comparison	-	*different from the, on the other hand*
	Deictics and locatives	-	*at the time of, (at) the end (of) (the)*
	Vagueness markers	-	*and so on, and so forth*
Stance expressions	Hedges	-	*there may be, are likely to*
	Epistemic stance	-	*according to the, be considered as*
	Obligation and directive	-	*take a look at, (it should) be noted*

	Expressions of ability and possibility	-	*can be used to, it is possible (that/to)*
	Evaluation	-	*it is important to, it is clear that*
	Intention/volition, prediction	-	*let me just, I just wanted to*
Discourse organizing functions	Metadiscourse and textual reference	-	*as shown in, in the next section*
	Topic introduction and focus	-	*first of all, for example (if/in/the)*
	Topic elaboration	non-clausal	*in more detail, any questions about*
		cause and effect	*as a result of, because it is*
	Discourse markers	Intangible framing attributes	*as well as, as a consequence*

Despite the variation in the number of categories, Biber et al. (2004) and Simpson-Vlach & Ellis (2010) also differ in their classification of some lexical bundles. For instance, while the former considers the bundle *the fact that the* as a stance expression, the latter classifies it as a referential expression. Other bundles also diverge in their classification, as we can see from Table 2.4 below.

Table 2.4: Differences in functional classification of lexical bundles in Biber et al. (2004) and Simpson-Vlach & Ellis (2010)

	Biber et al. (2004)	**Simpson-Vlach & Ellis (2010)**
(the) fact that (the)	**Stance expressions** (epistemic stance -impersonal)	**Referential expressions** (specification of attributes – intangible framing attributes)
(to) come up with	**Stance expressions** (ability - personal)	**Discourse organizing functions** (topic elaboration – non-causal)
a little bit about	**Discourse organizers** (topic introductions/focus)	**Stance expressions** (hedges)

has to do with / nothing to do with	**Discourse organizers** (topic elaboration)	**Referential expressions** (identification and focus)
shown in figure N / as shown in figure	**Referential expressions** (time/place/text reference – text deixis)	**Discourse organizing functions** (metadiscourse and textual reference)
on the other hand	**Discourse organizers** (topic elaboration/ clarification)	**Referential expressions** (contrast and comparison)

Besides being classified from a functional perspective, lexical bundles can also be grouped into structural categories. The most complete classification of lexical bundles according to their major structural patterns is found in Biber et al. (1999), where the authors group them into those bundles that are more widely used in conversation, and those more widely used in academic prose. Other studies either used Biber et al.'s classification or adapted their classification (see for instance, Biber et al., 2004; Hyland, 2008; Chen & Baker, 2010; Byrd & Coxhead, 2010). For the lexical bundles found in academic prose, Biber et al. (1999) distinguish among the following 12 major structural categories, shown in Table 2.5.

Table 2.5: classification of lexical bundles (structural categories) according to Biber et al. (1999).

Structural category	Examples
Noun phrase with of-phrase fragment	*the form of the, the length of the, the magnitude of the, the edge of the*
Noun phrase with other post-modifier fragments	*the way in which, the fact that the, an important part in, an increase in the*
Prepositional phrase with embedded of-phrase	*about the nature of, at the beginning of, for the purposes of, in the study of*
Other prepositional phrase (fragmente)	*at the same time, in such a way, in relation to the, in the next section.*
Anticipatory it + verb phrase/adjective phrase	*it is possible to, it is important that, it can be seen, it may be that*
Passive verb + prepositional phrase fragment	*is shown in figure, is referred to as, be used as a, can be seen as*

Copula be + noun phrase/adjective phrase	*is one of the, are a number of, is due to, is equal to the*
(Verb phrase +) that-clause fragment	*should be noted that, has been suggested that, can be seen that, does not mean that*
(Verb/adjective +) to-clause fragment	*are likely to be, will be able to, have been shown to, was found to be*
Adverbial clause fragmente	*as shown in figure, as we have seen, as we shall see, if there is a.*
Pronoun/noun phrase + be (+ ....)	*this is not the, there was no significant, there are a number, there has been a*
Other expressions	*as well as the, than that of the, may or may not, the presence or absence*

2.4. Overview on verb tense and aspect

In Portuguese, grammarians such as Bechara (1999) and Cegalla (2008) agree on the existence of three verb tenses in the indicative mode: present, past and future. Present tense does not undergo any subdivision, but past and future do. The past tense is subdivided into three: imperfect preterit, perfect preterit, and pluperfect preterit[5], while the future is subdivided into simple future and conditional future[6] (Bechara, 1999; Cegalla, 2008). In contrast, the English language system considers the existence of two tenses: present and past, but not future (Biber et al., 1999; Leech & Svartvik, 2003). This distinction is due to the fact that future is morphologically inflected in Portuguese, but not in English. In English, future time (not tense) is usually marked in the verb phrase by modal or semi-modal verbs such as *will*, *shall*, and *be going to*.

From a structural point of view, Portuguese presents various verb tenses while in English their correspondents are realized through a combination of verb tense and aspect. As Biber et al. (1999) put it, "whereas tense refers primarily to past and present time orientation, aspect relates to considerations such as the completion or lack of completion of events or states described by a verb" (p. 460). Thus, formally, tense and aspect do not have a one to one correspondence in Portuguese and English. Table 2.6

[5] In Portuguese: *perfeito, imperfeito,* and *mais-que-perfeito.*
[6] In Portuguese: *futuro do presente*, and *futuro do pretérito.*

below illustrates how the combination of tense and aspect in English is realized when translated into Brazilian Portuguese.

Table 2.6: Verb tense and aspect correspondence in English and Brazilian Portuguese.

English	example	Portuguese	example
Present simple	*I do it*	Present	*(Eu) faço*
Present progressive	*I am doing it*	Present + gerund	*(Eu) estou fazendo*
Present perfect	*I have done it*	Perfect preterit	*(Eu) fiz*
	I have done it since...	Present	*(Eu) faço (isso) desde...*
Past simple	*I did it*	Perfect preterit	*(Eu) fiz*
Past progressive	*I was doing it*	Perfect preterit + gerund	*(Eu) estava fazendo*
Past perfect	*I had done it*	Perfect preterit + participle (*or*) Pluperfect preterit	*(Eu) tinha feito / (Eu) fizera*

As we can see from Table 2.6 above, in addition to the two tenses (present and past), the English language possesses two aspects that combine with tense: progressive and perfect. Thus, we see the realization of verb forms that carry tense but not aspect (present simple and past simple), and those with both tense and aspect (present perfect, past perfect, present progressive, and past progressive), which are discussed in more detailed below. It is worth mentioning that in this section (and in this study), we do not examine the verb forms that take more than one aspect (i.e. present perfect progressive and past perfect progressive).

The present simple (or simple present) is unmarked morphologically, except in third person singular, where there is the addition of the suffix *(e)s*. According to Biber et al., (1999, p. 453) the present simple referring to the present time has two meanings: i) to describe a state existing at the present time (example 1), and ii) to describe present habitual behavior (example2). The authors add that the simple present can also report on an action ongoing at the time (example 3):

1. *I want a packet of crisps.*
2. *She's vegetarian but she eats chicken*.
3. *Here comes your mother.*

In addition to these uses, Leech & Svartvik (2003, p.47) show that the present simple can be used to refer to a present event, one that begun and ended at the very moment of speech (examples 4 and 5). The authors emphasize that this a rather specialized use, limited to formal declarations, sports commentaries, demonstrations, etc.

4. *Berner passes the ball to Lorimer*.
5. *I declare the meeting closed*.

The present simple can also be used to convey past and future meanings. For these uses, Greenbaum & Quirk (1990, p. 49) identify three meanings: i) the historic present, which refers to the past time and expresses the immediacy of an event happening at the time it is being narrated (example 6); ii) the present referring to the past with verbs of communication or reception to suggest that the information given is still valid (example 7), and iii) where the present simple occurs with time-position adverbials to suggest that a future event is certain to take place (example 8).

6. *Just as we arrived, up comes Ben and slaps me on the back.*
7. *I hear that you need an assistant.*
8. *The plane leaves for Ankara at eight o'clock tonight*.

As for the past simple (or simple past), Biber et al. (1999) state that it is marked morphologically with the suffix *-ed* for regular verbs (e.g. want - wanted), and marked with different endings (or unmarked) for irregular verbs (e.g. build - built, break - broke, cut - cut). Greenbaum & Quirk (1990, p. 50) indicate that the past simple is used to refer to a situation set at a definite time in the past in three main cases: i) the event past, which is used with dynamic verbs to refer to a single definite event in the past (an event that took place over an extended period - example 9 - or at a specific point of time - example 10); ii) the habitual past, which is used with dynamic verb senses to refer to

past events that repeatedly occur (example 11); and iii) the state past, which is used with stative verbs to refer to a single unbroken state of affairs in the past (example 12).

9. *The Normans invaded England in 1066.*
10. *The plane left at 9 a.m.*
11. *We spent our holidays in Spain when we were children.*
12. *I once liked reading novels.*

The authors also identify three other special uses the of the simple past tense: i) in indirect speech or indirect thought, where the verb in the subordinate reported clause is backshifted into the simple past (example 13); ii) the attitudinal past, which is used to refer more tentatively/politely to a present state of mind (example 14), and iii) the hypothetical past, which is used in certain subordinate clauses - especially *if*-clauses - to convey what is contrary to the belief or expectation of the speaker (example 15).

13. *She said that she knew you*.
14. *Did you want to see me now*?
15. *If you knew him, you wouldn't say that*.

Leech and Svartvik (2003, p. 48-49) add that the simple past can also indicate a gap between the time referred to and the present moment (example 16), and that adverbials referring to a past point or period of time normally go with the past tense (example 17).

16. *His sister was an invalid all her life (i.e. she's now dead).*
17. *The discovery was made [in the sixteenth century]*.

As for the perfect aspect, Biber et al. (1999, p. 460) state that it "designates events or states taking place during a period leading up to the specified time". In this sense the present perfect is marked by the auxiliary verb have/has ('ve/'s) used in the present plus the *ed*-participle, and is used to refer to situations that began sometime in the past and

continues up to the present (example 18). In this example, *leaving for a holiday* describes a past action that creates a situation (result) that extends to the present.

18. *But now she's gone on holiday for a whole month.*

Indeed, as Biber et al. (1999) put it, both the present perfect and the simple past tense are used to refer to states or events in the past. The main difference between them is that the present perfect describes a situation that continues to exist up to the present (example 19), while the simple past refers to a situation that no longer exists (example 20) or an event that happened at a particular time in the past (example 21). This difference in meaning is often made clear by time adverbials; with the past tense, these adverbials describe when the event or state occurred, while with the present perfect they mark the beginning point or the duration of the period of time.

19. *We've had it [since last January]*.
20. *I saw him [yesterday].*
21. *[At that moment], Toby knocked*.

Greenbaum & Quirk (1990, p. 51-52) bring a more detailed account of the uses of the present perfect, including: i) the event present perfect, in which the event/events are reported as news and usually have occurred a short time before the present (example 22), or the event occurred at some more remote time in the past, but the time period still leads up to the present (example 23); ii) the habitual present perfect, which is used with dynamic verb senses to refer to past events that repeatedly occur up to and including the present (example 24).

22. *The Republicans have won the election*.
23. *All our children have had measles (and they are not likely to have it again).*
24. *Socrates has influenced many philosophers (till now).*

An interesting use of the present perfect in found in Biber et al (1999, p.465), where they state that the present perfect is used in academic prose to "imply the continuing validity of earlier findings or practices" (examples 22 and 23).

22. *Experiments have shown, however, that plants can obtain their nutrients at sufficient rates to maintain rapid growth.*

23. *It has become the usual practice to use only maintenance applications.*

Also according to Biber et al. (1999), compared with present perfect aspect, past perfect aspect has a more straightforward function, being used to refer to a time that is earlier than some specified past time (example 24). Or, more simply, as put by Leech & Svartvik (2003, p. 50), it indicates "past in the past"

24. *Widow Lilian Boyes, 70, had earlier pleaded with doctors to "finish her off," Winchester Crown Court heard.*

Leech & Svartvik (2003) also point out that when we talk about an event following another in the past, we can either indicate the earlier event by using he past perfect (example 25), or use the simple past tense for both events and use conjunctions (*after*, *before*, *when*) to show which event happened first (example 26).

25. *The teacher had left the room when the children started talking.*
26. *The children started talking after the teacher left the room.*

The last aspect, the progressive (also referred to as continuous), is described in Biber et al. (1999) as being marked by the auxiliary verb be + *ing* participle combined with either the present or past tense. In general terms, the present progressive aspect is used to describe events that are currently in progress (example 27) or are going to take place in the near future (example 28). Similarly, the past progressive aspect will describe "events that were in progress or about to take place at some earlier time" (p. 470), (examples 29 and 30).

27. *I'm looking for an employee of yours.*

28. *But she's coming back tomorrow.*

29. *I was looking at that one just now.*

30. *I was just coming back from Witham.*

In addition, the authors emphasize that although it is sometimes assumed that the progressive aspect occurs only dynamic verbs that describe activities or events, it can also be used with verbs that describe a static situation. In this case, the meaning expressed by the progressive is that of a temporary state that exists for a period of time (example 31).

31. *I was sitting in my office smoking one of James's cigarettes.*

Greenbaum & Quirk(1990, p. 55) note that the progressive aspect has some special uses: i) verbs denoting states of bodily sensation may be used more or less interchangeably in the progressive and non-progressive when referring to a temporary state (example 32); ii) in combination with indefinite frequency adverbs such as always and continually, the habitual progressive loses its temporary meaning and often conveys disapproval (example 33);

32. *My back aches/is aching*.

33. *Bill is always working late at the office.*

Finally, Leech & Svartvik (2003, p. 52) observe that when the progressive is linked to a non-progressive event verb, or to a point of time, the progressive normally indicates that at that point the activity or situation denoted by the verb is still in progress. In example 34 below, breakfast started before eight o'clock and continued after that time.

34. *At eight o'clock they were (already) eating breakfast*.

Having discussed verb tense and aspect in this section, we now turn to the next section, where we discuss grammatical complexity indices by showing how it used to be measured before and the more recent approach to measuring it.

2.5. Grammatical Complexity in academic prose

Different methods have been used to measure the grammatical complexity of texts written in English. Below we present the traditional (initial) measures and a new approach to measuring grammatical complexity. We would like to stress that in this study we use the new approach proposed in Biber and Gray (2010) and present the traditional approach simply for comparison purposes.

2.5.1. Traditional measures of grammatical complexity

Since the mid-1960s, the so-called T-unit has become a widely used parameter for measuring grammatical complexity in oral and written texts. Hirano (1988) explains that this was a term coined by Hunt (1965, 1970, 1977), and used as a way of measuring the syntactic maturity in the texts of L1 English speaking students and was defined as "a main sentence added to all subordinate clauses and non-clausal structures linked or attached to it" (Hunt, 1970, p.4). However, treating only the number of T-units in a sentence is not enough to measure the grammatical complexity of a given text, as can be seen in the examples below, taken from Gaies (1980, p.60):

1. I've known this woman for a long time. We can rely on her. I am confident of that. She'll get the job done.

2. This woman, whom I have known for a long time and whom I feel confident we can rely upon, will get the job done.

In Example 1, there are a total of four T-units, while example 2, explicitly more elaborate, has only two of these units. Therefore, instead of simply counting these units, the average number of words per T-unit is used as the index of complexity. In the examples above, this average corresponds to 6.25 words for the first and 24.0 for the second. Thus, the average number of words per T-unit was, according to O'Donnell (1976, p. 38), "the most useful and widely used syntactic development index".

Later, this index was improved and new parameters added to it so that it could be used in the context of L2 writing. Larsen-Freeman & Strom (1977) propose that, among several indicators, those that proved to be the most appropriate to serve as an index of writing development are, in addition to the extension of T-units (i.e. average words per T-unit in a text), also the total number of T-units without errors. By 'errors', the authors understand the incorrect use of elements such as morphology, syntax, prepositions, verb tense, aspect, articles, verbal agreement, case and negation.

More recently, Wolfe-Quintero et al. (1998) described several studies that deal with the development of L2 writing and emphasize T-unit sentences and the proportion of subordinate clauses by major clauses as the "best [...] measures of complexity so far" (p.118- 119), and perhaps because of this indication, studies in writing development since 1998 continue to use measures that are based on these units (BIBER et al., 2011).

2.5.2. New approach to measuring grammatical complexity

Although it has generated numerous studies, the use of T-units as a measure of grammatical complexity has recently been questioned, especially with regard to academic writing. In their publication, Biber et al. (2011, p.14) point out that measures based on subordinate clauses are insufficient to determine the syntactical complexity of a text, since subordination is a characteristic more predominantly found in oral texts (conversation), not in written texts.

To illustrate, the authors bring two sentences with T-units of the same size; the first from conversation and the second of an academic textbook.

1. Well, since he got so upset, I just did not think we would want to wait for Tina to come back.
2. This may be part of the reason for the statistical link between schizophrenia and membership in the lower socioeconomic classes.

By definition, both sentences constitute a single T-unit. However, the first sentence has four subordinate clauses per T-unit, while the second consists of the main sentence only, with no dependent clauses.

1. Well, [since he got so upset], I just didn't **think**
   [we would want
   [to wait
   [for Tina to come back]]].

2. This may **be** part of the reason for the statistical link between schizophrenia and membership in the lower socioeconomic classes.

Clearly, the second sentence seems more complex (from the point of view of processing the message) than the first. However, if we consider the average extension of T-units as a parameter, the two sentences would be equally complex, with 20 words each. On the other hand, if the parameter considered were the average of subordinate clauses per T-unit, the first sentence would be considered more complex, as it has four subordinations, as opposed to no subordination at all in the second sentence.

Given this incoherence, new indices of complexity were proposed and presented in Biber & Gray (2010), Biber, Gray & Poonpon (2011) and Biber & Gray (2016). These indexes take into account a dominant feature in the syntactic structures of written academic discourse: non-clausal elements embedded in nominal phrases (NPs). This includes, in particular, four highly productive elements in this type of register: adjectives as noun premodifiers (adj + n), nouns as a noun premodifier (n + n), prepositional phrase as noun phrases postmodifier (n + PP), and prepositional phrases with adverbial function (PP> adv), as can be seen in Fig. 2.1 below:

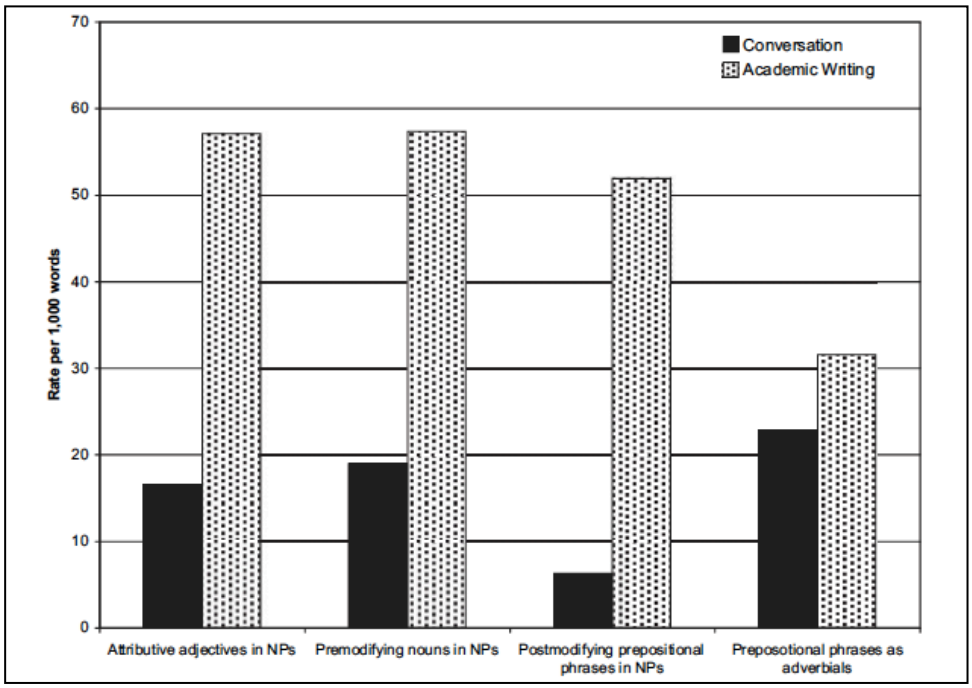


Figure 2.1: Common dependent phrase types. Biber & Gray (2010, p.8)

It can be seen, therefore, that while the predominant characteristics of academic discourse come from the phrasal level, in conversation they are found at the clausal level, especially with regard to the use of subordinate clauses, as can be seen in Fig. 2.2.

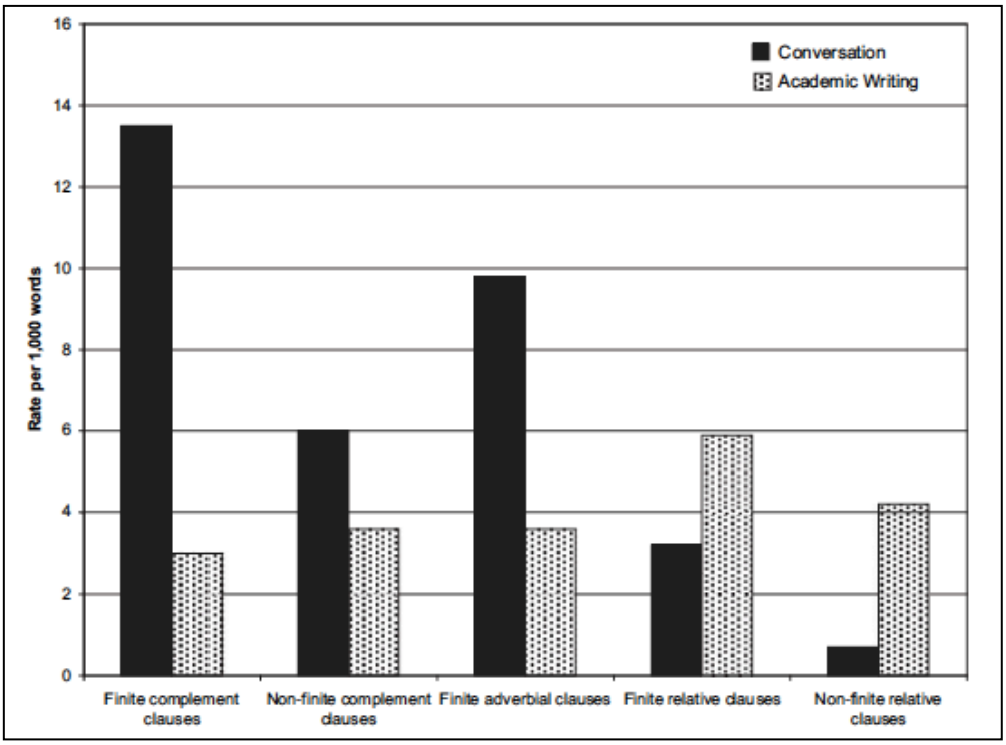


Figure 2.2: Common depend clause types. Biber & Gray (2010, p.8).

Subordinate clauses are highly productive in conversation, especially when compared to academic writing. Except for relative clauses (adjective subordinate clauses), which are more common in academic writing, the sentence component of the conversation is characterized by the transmission of messages using various clauses, especially through coordination and subordination. On the other hand, as observed by Biber et al. (2016), in academic writing, information is compressed into a single sentence by using components at the phrasal level, mainly through the insertion of modifying elements within the noun phrase. This is in accordance with what Halliday (2004, p. 171) suggests, when observing that "something that in spoken English would typically be expressed as a clause, is expressed instead [in writing] as a group of words centring on a noun".

CHAPTER 3: METHODS

3.1. Introduction

The following lines intend to describe the methods adopted for the development of this study. The procedures are based on the methodological approaches found by Corpus Linguistics (CL) and, in this respect, our study shares the views of those who, like Gries (2009), see CL as a methodology, not a theory. Below we describe the corpus compilation process as well as the techniques for data analyses.

3.2. The corpora used in this study

We have compiled three corpora for the development of this study. The first, TB-Chem (Corpus of Textbooks of Chemistry), has a total of 510.487 tokens divided in , with an average of 27.790 words per text, and the second, TB-Phy (Corpus of Textbooks of Physics), presents 501.559 tokens, with an average of 27.864 words per text. Both corpora are made up of textbooks published after the year 2000 and are used at American universities. In order to select which textbooks would be part of the corpora,

we consulted various course syllabi available online from different universities[7]. For Chemistry, 20 textbooks served as the source from which texts were extracted, whereas for physics we used 18 textbooks. A list containing all textbooks used is found in Appendix A.

The first step taken in the corpus compilation of TB-CHEM and TB-PHY was text selection. For a corpus to be valid for linguistic analyses, it needs to contain texts that will represent the language variety being studied, that is, the corpus needs to be representative. According to Biber (1993, p.243), representativeness is "the extent to which a sample includes the full range of variability in a population". This means that in order to compile a corpus, the text selection must follow a strict design. In our case, we consulted with specialists in order to know what disciplines most commonly represented the fields of Chemistry and Physics. By doing that, we would prevent the corpora from being skewed towards one or more areas. After the consultation, the following disciplines were chosen to represent the fields of Chemistry and Physics:

Table 3.1: Disciplines that make up the fields of Chemistry and Physics in TB-Chem and TB-Phy.

Chemistry	Physics
General Chemistry	General Physics
Organic Chemistry	Classical Mechanics
Inorganic Chemistry	Quantum Mechanics
Analytical Chemistry	Electrodynamics/Magnetism
Physical Chemistry	Acoustics
	Optics

[7] We have performed online searches for websites with the domain .edu so as to find available syllabi for specific disciplines. For instance, following two syllabi were one of the many consulted for organic chemistry

.

https://canvas.harvard.edu/courses/571/files/29587/download?verifier=KwZDg7B0LllRxeuCsQIZXypBjRjYysseyvM4i6ae

https://ocw.mit.edu/courses/chemistry/5-12-organic-chemistry-i-spring-2003/syllabus/

	Solid State Physics
	Wave Motion
	Relativity
	Thermodynamics
	Condensed Matter Physics

As shown in Table 3.1, we have included General Chemistry and General Physics among the branches that make up those areas of study. These disciplines are the ones taught in first year introductory courses, such as Chemistry 101 and Physics 101. These introductory textbooks generally include bits of information about all the other subfields listed in Table 3.1. Another very important aspect of the corpus compilation process is balance, in which "text categories are typically sampled proportionally" (McEnery, Xiao, Tono, 2006, p. 16). In this respect, we made sure to extract similar amounts of text from every subfield so that no discipline would be over or under represented in the corpora.

All texts were then separated, cleaned and converted into .txt format so that they would be ready for analyses. Cleaning texts means that unnecessary information was deleted, including tables, graphs, images, as well as unnecessary textual information, such as headers. Also, something very commonly found in chemistry and physics textbooks is equations. For this study, equations, formulas, chemical structures, and mathematical operations represented in the form of text were deleted from the corpora and replaced by the notation *formulae* so as to facilitate its readability. For example, for this study we rewrote sentence A below into sentence B.

Sentence A: *Fourier's theorem states that this function can be written as* $y(t) = \sum_n \ (A_n \ nt + B \ n \ cos \ cos \ 2 \ nt)$.

Sentence B: *Fourier's theorem states that this function can be written as formulae.*

The third corpus is a comparative corpus containing News texts. We have opted for BNC Baby[8], which is a subset of the original 100 million-word BNC (British National Corpus). The BNC Baby contains 4 million words divided into four registers of 1 million tokens each: conversation, fiction, academic, and news. According Burnard (2008), for the news register, BNC Baby followed the sampling methods of the original corpus compilers and is made up of a mix of national and local papers (British newspapers), containing a variety of topics and little duplication of dates. Approximately 60% of the newspaper data comes from five national papers, and the remaining 40% from regional newspapers. Texts were selected in a way that ensured dates of publication would not be the same so as to reduce the effects of seasonal or topical variation. In addition, the size of the texts was taken into consideration in order to guarantee a roughly equal distribution across different newspapers. The table below shows the number of words in each newspaper sampled for BNC Baby.

Table 3.2: Number of words sampled from each newspaper for BNC Baby (Burnard, 2008).

Newspaper	words	%
Daily Mirror	124251	12%
Daily Telegraph	128794	13%
Guardian	129598	13%
Independent	131205	13%
Today	91238	9%
Belfast Telegraph	43006	4%
East Anglian Daily Times	43674	4%
Liverpool Daily Post and Echo	85441	9%
Northern Echo	68887	7%
The Alton Herald	56316	6%
The East Anglian	15814	2%
The Scotsman	66709	7%
Ulster Newsletter	16888	2%
Total component	1001821	100%

For this study we did not use the entire news folder containing 1 million tokens, but rather we chose to use half of its total, i.e. around 500k. We chose to use this amount of text so that the news corpus would be more of a similar size in comparison to TB-

[8] Available for download free of charge at:
https://ota.bodleian.ox.ac.uk/repository/xmlui/handle/20.500.12024/2553

Chem and TB-Phy. In our selection, we made sure to sample equivalent amounts of text from each file in order to maintain its balance. Each one of the files (originally in .xml format) were opened in a web browser, cleaned (i.e. metadata were deleted), and then transformed into .txt format.

3.3. Steps for Corpus Analyses

The first research question we answered is about the most common lexical items found in our corpora. For this purpose, we used the software AntConc 3.5.7 (Anthony, 2018)[9]. After having found the number of occurrences for each lexical item in both corpora, their frequencies were normalized[10]. Normalizing frequencies is a fundamental step for working with corpora of different sizes, as otherwise results could be entirely biased. Results from the searches were normalized per 1,000 words, following the Biber, Conrad & Reppen (1998, p.264), where they state that "frequency counts should be normed to the typical text length in a corpus". In the case of textbooks and news texts, looking at the data distribution on a 1000-word base makes it easier to interpret results.

In addition to general frequency lists, we chose to work with keyword lists, defined by Hunston, (2002, p. 68) as "words which are significantly more frequent in one corpus than another". One advantage of working with keywords is that the resulting output list shows those words that are especially typical of a given corpus. For ranking the most frequent keywords, we have opted for their classification according to keyness. As put by Coxhead (2013, p. 40), "if a word has a high level of keyness, the occurrence is probably not by chance".

In order to generate a keyword list, one needs to select another corpus for comparison – usually a general reference corpus. This means that our corpora were compared to a general reference corpus and, through the comparison, a list of keywords could be generated. The reference corpus used here was the MASC (Manually Annotated Sub-

[9] AntConc can be download free of charge at http://www.laurenceanthony.net/software/antconc/.
[10] All of the data in the results section were normalized.

Corpus), which is a balanced subset of 500K words of written texts and transcribed speech drawn primarily from the Open American National Corpus (OANC), including around 25K words from each of 19 different registers. The texts are distributed as follows:

Table 3.3: Distribution of registers in the MASC corpus.

Register	No. of files	No. of words	% corpus
Court transcript	2	30052	6%
Debate transcript	2	32325	6%
Email	78	27642	6%
Essay	7	25590	5%
Fiction	5	31518	6%
Gov't documents	5	24578	5%
Journal	10	25635	5%
Letters	40	23325	5%
Newspaper	41	23545	5%
Non-fiction	4	25182	5%
Spoken	11	25783	5%
Technical	8	27895	6%
Travel guides	7	26708	5%
Twitter	2	24180	5%
Blog	21	28199	6%
Ficlets	5	26299	5%
Movie script	2	28240	6%
Spam	110	23490	5%
Jokes	16	26582	5%
TOTAL	**376**	**506768**	**100%**

It is important to mention that our analyses concerning the most frequent lexical items in the corpora were not limited to content words. We also looked at function words (prepositions, determiners, and modal verbs) that revealed interesting findings in the comparison of the corpora.

Our second research question aimed at investigating n-grams in our corpora. In order to generate a list of n-grams in Antconc, we first need to decide on a cut-ff point for inclusion, that is, the number of times a certain word sequence should appear in a corpus in order to be considered for analyses. Although some previous studies adopted a cut-off point of n-grams occurring 20 times per million words (Hyland, 2008a, 2008b; Cortes, 2004), we have opted for a more conservative cut-off point of 40 times per million words, following Biber, Conrad & Cortes (2004). This conservative approach seems to be more in line with the present study, as these last authors also dealt with a corpus of textbooks in their study. In addition, for an n-gram to be included in our analysis, it should be present in at least 50% of the texts. This is an arbitrary decision that was taken so as to make sure the word sequences found in the corpus would represent the most productive word sequences to be included in future pedagogical materials.

Our third research question addresses verb tense and aspect in our corpora. We investigated six different verb tense/aspect markers, namely: present simple, past simple, present perfect, past perfect, present progressive, and past progressive. Before starting the searches, our corpora were tagged with Claws 5 Tagger, which means that every word in the corpus was assigned a part-of-speech tag. Nouns, for example, were tagged as NN0, NN1, NN2, or NP0, where NN0 means singular noun neutral for number (aircraft, data), NN1 means singular noun (pencil, goose), NN2 means plural noun (pencils, geese), and NP0 means plural noun (London, Michael). Below is an example of a tagged sentence from our chemistry corpus:

[01] **Look**_VVB **again**_AV0 **at**_PRP **the**_AT0 **reaction**_NN1 **between**_PRP **sodium**_NN1 **and**_CJC **chlorine**_NN1 **introduced**_VVN **in**_PRP **Section**_NN1 **3.2**_CRD **.**_SENT [CHEM 01]

For each verb tense/aspect, different criteria were established for the search, since Claws 5 assigns different tags specifically for the inflected forms of verbs. Thus, it differentiates among the forms of *be (am is are), do (do, does), have (have, has),* and other verbs as a whole such as take (take, takes). In this sense, if we wanted to count all instances of the present simple, we should look for each tag separately and then add the totals up. However, the tagger does not distinguish among main verbs and auxiliary verbs, which poses a challenge to the searches. For example, it assigns the tag VHZ to the verb *has* being used as both a main verb (example 02) and as an auxiliary verb (example 03).

[02] *An average-size adult human male **has** a mass of about 70 kilograms.* [CHEM 03]

[03] *(...) the Fahrenheit scale **has** since been modified so that (...).* [CHEM 03]

Thus, to get to the counts of the present simple, searches were made based on the tags and then subtracted from the totals where the verbs have/has, and do/does were being used as auxiliary verbs. In addition, we had to subtract counts of the verb *be* used for the present progressive so as to differentiate between uses such as "the molecules are moving" and "hot and cold are relative", because the verb form *are* is tagged as VBB in both cases. This step was taken for all other verb tenses/aspects, as there was always overlapping cases that had to be taken into consideration.

Another point of concern had to do with the notion of precision, i.e. to what extent are all of the KWIC lines matching the query actually occurrences of the linguistic item searched? In our case, Claws 5 presented low precision for the tagging of the past simple. In a random sample of 300 concordance lines, 32% of the tagged words in chemistry were not correct for the past simple, while in physics this number reached 17%. News, on the other hand, was tagged correctly for the past simple in 100% of the cases. The wrong tags in chemistry and physics were mainly related to a word being tagged as VVD (past tense form of lexical verb), when it should have been tagged as VVN (participle form of lexical verb), as shown in example 05. In addition, a few of the wrong

tags were due to a word being considered VVD (past tense form of lexical verb), when it should have been tagged as AJ0 (adjective) - example 06.

[05] However_AV0 ,_PUN in_PRP October_NP0 1983_CRD ,_PUN the_AT0 meter_NN1 (_PUL m_ZZ0)_PUR was_VBD redefined_VVN as_PRP the_AT0 distance_NN1 **traveled_VVD** by_PRP light_NN1 in_PRP vacuum_NN1 during_PRP a_AT0 time_NN1 of_PRF 1/299_CRD 792_CRD 458_CRD second_NN1 [PHY 12]

[06] Wintergreen_NP0 (_PUL methyl_NN1 salicylate_NN1)_PUR **flavored_VVD** candies_NN2 ,_PUN on_PRP the_AT0 other_AJ0 hand_NN1 ,_PUN visibly_AV0 glow_VVB when_CJS crushed_VVN [CHEM 12]

Finally, our fourth research question addressed two of the phrasal grammatical complexity indices discussed in Biber & Gray (2010), namely: attributive adjective (adjective as noun pre-modifier) - example 07 - and premodifying nouns (noun as noun pre-modifier), example 08:

[07] *Alkyl groups may also act as bridges in **inorganic** compounds, a function rarely encountered in **organic** chemistry*. [CHEM 16]

[08] *The energy of the **electron beam** can be decreased from the typical 70 electron volts (eV) to 20–25 eV, where much less fragmentation occurs.* [CHEM 17]

These two grammatical features were chosen for this study because among the four features investigated in Biber & Gray (2010) - including prepositional phrase as noun post-modifier and prepositional phrase as adverbial - those were the most frequent in academic writing. For the searches concerning phrasal complexity, we followed search patterns proposed by Gray, Staples & Egbert (2018), using the Regex feature of AntConc 3.5.7. For each search, a different regular expression was used, as follows:

- Attributive adjectives (adj + n): \w+_AJ\w*\s+\w+_NN\w*
- Pre-modifying nouns (n + n): \w+_NN\w*\s+\w+_NN\w*

- Pre-modifying nouns (n + n + n) \w+_NN\w*\s+(?=\w+_NN\w*)

After doing each one of the searches, we evaluated the accuracy of the query by analyzing the data in terms of precision. In a random sample of 300 concordance lines, we found 94% precision in chemistry, 96.5% in physics and 98% in news. Thus, the total counts were subtracted accordingly so that we could come to more accurate results. In the case of noun + noun sequences, wrong counts occurred mainly due to a word being incorrectly tagged as a noun when it should have been tagged as a verb (example 09), or when a noun + noun sequences did not in fact represent a noun modifying another noun (example 10).

[09] The_AT0 motion_NN1 **results_NN2** in_PRP collisions_NN2 between_PRP the_AT0 particles_NN2 and_CJC the_AT0 surfaces_NN2 around_PRP them_PNP ._SENT [CHEM 02]

[10] This_DT0 long_AJ0 delay_NN1 is_VBZ probably_AV0 due_PRP to_PRP the_AT0 **fact_NN1 engineers_NN2** were_VBD suddenly_AV0 provided_VVN with_PRP l ight_AJ0 beams_NN2 [PHY 06]

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Introduction

In this section we set out to discuss the research questions proposed in this study. For each subsection, we show and analyze results from both a quantitative and qualitative point of view, followed by a brief discussion of possible applications of these results in pedagogical materials aimed at teaching English for Reading in the context of English as a foreign language. We begin by addressing the most common lexical items in our corpora through simple wordlists as well as keyword lists. In this context, lexical items include not only content words, but also function words such as prepositions and modal verbs. Following that, we move on to show what n-grams frequently occur in our

corpora and how they can be interpreted in terms of their discourse functions and syntactic patterns. Then, we go on to address how verb tense and aspect behave in discourse found in chemistry and physics textbooks. Finally, we examine two of the most common phrasal grammatical complexity indices found in academic prose, namely attributive adjectives and premodifying nouns.

4.2. Most common lexical items

Identifying the most frequent words in a corpus helps researchers, teachers, and materials developers to have a good general picture of the type of register they are dealing with. A wordlist containing the most common words in a corpus can be obtained rather simply, as it can be done automatically through software, such as AntConc (Anthony, 2018) and Wordsmith Tools (Scott, 2019). However, it is important to remember that for any given corpus of either written or spoken language, the most frequent words are always going to be function words.

We begin our analyses with TB-Chem and, not surprisingly, out of the 100 most common words in the corpus, 57 are function words, as opposed to 43 content words shown in the list below (content words are underlined).

1. the
2. of
3. a
4. in
5. and
6. is
7. to
8. are
9. that
10. as
11. for
12. be
13. with
14. by
15. an
16. this
17. we
18. it
19. or
20. can
21. from
22. at
23. on
24. energy
25. which
26. one
27. have
28. not
29. two
30. atoms
31. has
32. more
33. electrons
34. than
35. when
36. water

37. molecules
38. because
39. these
40. electron
41. other
42. if
43. atom
44. will
45. so
46. figure
47. but
48. carbon
49. used
50. its
51. between
52. they
53. each
54. example
55. such
56. bond
57. also
58. reaction
59. orbitals
60. all
61. molecule
62. hydrogen
63. only
64. number
65. temperature
66. same
67. into
68. acid
69. ions
70. solution
71. called
72. some
73. compounds
74. ion
75. may
76. form
77. you
78. group
79. their
80. there
81. bonds
82. gas
83. most
84. many
85. pressure
86. different
87. liquid
88. point
89. orbital
90. about
91. chemical
92. molecular
93. however
94. see
95. reactions
96. sample
97. then
98. state
99. structure
100. shown

Most nouns (*energy, electron, carbon*) and adjectives (*liquid*, *molecular*) are words that belong to the domain of Chemistry. Verbs, on the other hand, do not display the same characteristic. The first two most frequent verbs (*have* and *has*) are in fact inflated counts, as their frequency account for instances of both auxiliary and main verbs (examples 11 and 12, respectively).

[11] *Understanding the structure of the atom **has** been a fundamental challenge for centuries* [CHEM 16]

[12] *Bioinorganic chemistry bridges biochemistry and inorganic chemistry and **has** an important focus on medical applications*. [CHEM 16]

Other verbs (*see*, *shown*) rank in the list for being very commonly found in textbook language to signal a reference to some other part of the text, as shown in examples 13 and 14.

[13] (…) *Infrared and Raman spectroscopies (**see** also Section 4.7) are branches of vibrational spectroscopy*. [CHEM 08]

[14] *The phase diagram of carbon dioxide is **shown** in Fig. 3.6*. [CHEM 07].

As for Physics, TB-Phy reveals that out of the 100 most frequent words, 39 are content words (underlined below), whereas 61 are function words. Interestingly, Physics and Chemistry share 31.5% of the most frequent words but some of the function words appear only in Physics (*would, where, very, no, any).*

1. the
2. of
3. a
4. is
5. in
6. to
7. and
8. that
9. as
10. are
11. this
12. we
13. for
14. be
15. it
16. by
17. with
18. an
19. on
20. at
21. can
22. energy
23. from
24. or
25. which
26. one
27. not
28. if
29. force
30. two
31. figure
32. have
33. time
34. will
35. when
36. wave
37. its
38. but
39. has
40. other
41. between
42. there

43. all
44. field
45. light
46. was
47. such
48. these
49. only
50. same
51. object
52. motion
53. then
54. so
55. point
56. they
57. mass
58. where
59. also
60. than
61. you
62. formula
63. because
64. some
65. more
66. system
67. theory
68. called
69. any
70. electrons
71. see
72. would
73. different
74. direction
75. each
76. particle
77. velocity
78. first
79. no
80. quantum
81. electron
82. potential
83. forces
84. their
85. given
86. case
87. very
88. must
89. into
90. magnetic
91. surface
92. through
93. example
94. sound
95. equation
96. speed
97. constant
98. vector
99. distance
100. however

One remarkable fact lies in the use of *where* in adverbial dependent clauses, in which *where* specifically refers back to some element of an equation. This is particularly common in Physics, though Chemistry also shows a number of occurrences of this pattern, shown in examples 15 and 16 below:

[15] *Spherical waves are described by gM(t,r) = a(r) f(t ± r/V), where r is the distance to the center of the wave.* [PHY 06]

[16] (...) *or, in equation form, that E = nhv, where E is the energy of an oscillator, n is an integer, h is a proportionality constant (…).* [CHEM 19].

We can also approach frequent words in a corpus through the so-called *keywords*, defined by Hunston, (2002, p. 68) as "words which are significantly more frequent in one corpus than another". We begin with TB-Chem and, for the purposes of this analyses, we have considered only the 100 most frequent content words.

1. energy
2. atoms
3. electrons
4. electron
5. molecules
6. atom
7. carbon
8. orbitals
9. bond
10. reaction
11. hydrogen
12. molecule
13. ions
14. water
15. compounds
16. ion
17. bonds
18. temperature
19. acid
20. orbital
21. gas
22. figure
23. solution
24. liquid
25. molecular
26. reactions
27. pressure
28. sample
29. used
30. atomic
31. chemical
32. example
33. mass
34. compound
35. oxygen
36. metal
37. volume
38. each
39. particles
40. bonding
41. structure
42. charge
43. solid
44. equation
45. phase
46. elements
47. form
48. number
49. equilibrium
50. fig.
51. formula
52. shown
53. concentration
54. sodium
55. potential
56. radiation
57. chemistry
58. solvent
59. heat
60. called
61. metals
62. properties
63. surface
64. energies
65. organic
66. group
67. titration
68. symmetry
69. magnetic
70. electrode
71. quantum
72. constant
73. table
74. ionic
75. chloride
76. boiling
77. spectrum
78. absorption
79. analyte
80. acids
81. nucleus

82. mixture
83. can
84. negative
85. which
86. transition
87. base
88. particle
89. dipole
90. chapter
91. bonded
92. thus
93. point
94. substance
95. wavelength
96. containing
97. structures
98. proton
99. gases
100. increase

As it can be seen from the list, the keyword list shows a much better picture of the words used in Chemistry textbooks. Notably, Chemistry and Physics are related fields of study, which explains why they display a number of similar frequent words. From the TB-Phy keywords list below, we see that 32% of the words coincide between Physics and Chemistry, a very similar figure to the one found when we compared the corpora through wordlists above (31.5%).

1. Energy
2. Force
3. Wave
4. Figure
5. Field
6. formula
7. motion
8. electrons
9. mass
10. object
11. electron
12. particle
13. velocity
14. quantum
15. light
16. magnetic
17. direction
18. particles
19. momentum
20. vector
21. equation
22. theory
23. waves
24. potential
25. constant
26. atoms
27. surface
28. forces
29. lattice
30. point
31. speed
32. temperature
33. acceleration
34. distance
35. electric
36. objects
37. Newton
38. Frequency
39. Between
40. Space
41. Sound
42. Physics
43. Kinetic
44. Relativity
45. Equal

46. Mechanics
47. Thus
48. Which
49. Frame
50. Crystal
51. Atom
52. Phase
53. Called
54. Given
55. Equations
56. Earth
57. Axis
58. Density
59. universe
60. plane
61. radiation
62. gravity
63. shown
64. same
65. electromagnetic
66. scattering
67. can
68. wavelength
69. moving
70. friction
71. length
72. angular
73. gravitational
74. equilibrium
75. material
76. charge
77. properties
78. coordinates
79. spin
80. zero
81. angle
82. Einstein
83. dimensional
84. different
85. solid
86. pressure
87. function
88. rays
89. consider
90. symmetry
91. principle
92. inertial
93. matter
94. model
95. vectors
96. beam
97. mechanical
98. diffraction
99. collision
100. system

This similarity noticed between chemistry and physics, however, is not seen when we compare them to the news corpus. As we can see from the list below, keywords found in news mainly include proper names (Peter, Thatcher, Kinnock), reference to cities, countries and nationalities (London, France, German), as well as words related to finances (shares, profits, dividend), and politics (MP, party, election).

1. yesterday
2. Mr
3. has
4. year
5. cent
6. police
7. labour
8. British
9. last
10. Britain
11. London
12. profits
13. said
14. shares
15. council

16. centre
17. minister
18. john
19. UK
20. party
21. Essex
22. spokesman
23. government
24. bid
25. south
26. group
27. France
28. market
29. chancellor
30. Kinnock
31. Mrs
32. England
33. election
34. ft.
35. royal
36. bn
37. cup
38. defence
39. figures
40. half
41. Peter
42. tory
43. prime
44. Thatcher
45. Germany
46. chairman
47. Colchester
48. hospital
49. Clowes
50. Ireland
51. former
52. appeal
53. wales
54. Barlow
55. European
56. programme
57. organisation
58. week
59. west
60. he
61. correspondent
62. MP
63. scheme
64. political
65. company
66. staff
67. David
68. property
69. major
70. night
71. month
72. German
73. aged
74. dividend
75. soviet
76. Lamont
77. Liverpool
78. car
79. Scotland
80. months
81. Scottish
82. commons
83. Sir
84. ambulance
85. expected
86. Anthony
87. home
88. sales
89. Europe
90. shareholders
91. investors
92. cash
93. DTI
94. officers
95. boss
96. tax
97. giant
98. MPs
99. Leader
100. added

The analyses of the lists can be even more useful when we look at concordance lines to see how words behave. For example, when comparing the frequencies of prepositions *in, on* and *at,* which are frequent prepositions found in written registers, we see that *in* is by far much more common than the other two. In fact, *in* comes to be at least four times greater, as shown in the Figure 4.1.

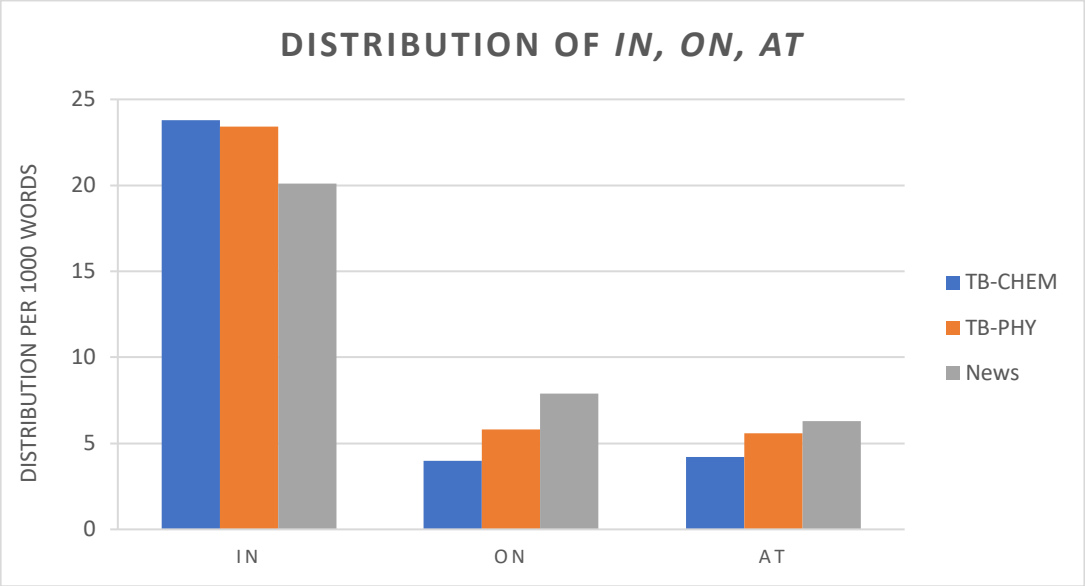


Figure 4.1: Normalized frequencies of prepositions per 1000 words in TB-Chem, TB-Phy, and News

Many EFL textbooks tend to group those three prepositions and teach them together due to semantic similarities in their use to indicate place and time. However, more attention could be given to the preposition *in*, as students are at least four times more likely to see it in a text than they are to find *on* or *at*.

In addition, we have observed that these prepositions are very productive in association with other words (verbs, nouns, and adjectives), thus being part of phrases. For example, in chemistry and physics, common collocates to *in* on the left form the combinations *shown in, results in, as in (figure/the case), an increase in,* and collocates on the right include *(soluble/insoluble/dissolves) in water, in figure, in terms of, in fact, in general.* As for *on* and *at*, different patterns have been found, including: *depend/depends/depending on, based on, acting on, on the surface, on the basis of, on*

Earth; the temperature at, moving at, located at, at rest, (not) at all, at least, at room temperature, temperature at which.

We, therefore, see these phrases as a good starting point when teaching these prepositions, instead of presenting the ones normally found in English for Reading textbooks, which tend to consider their uses in General English, such as *in the morning, on weekends, at night*, most commonly found in news texts, for instance. As stated by O'Keeffe, McCarthy & Carter (2007), the way that language clusters into combinations of words, the so-called chunks, has an impact on what we teach in our vocabulary lessons and, therefore, on how learners acquire vocabulary and fluency.

Another group of words that is present in our list of the most frequent words and are also often taught together as a class is the determiners/pronouns *some, any, no.* In this case, the three of them have a fairly similar distribution in the corpora, which seems reasonable to have them combined for teaching purposes. However, one striking difference is the fact that, in general English, *any* is generally found in questions and negatives (sentences with *not*, or negative words such as *without*), as attested by The Oxford English Dictionary and the Longman Dictionary of Contemporary English. And this could explain why textbooks tend to focus on *any* following that trend. However, that is not the case when it comes to academic English. In TB-Chem and TB-Phy, *any* is found almost exclusively in affirmative sentences, as seen in Fig 4.2. Examples 17 and 18 below show *any* in affirmative and negative sentences, respectively.

[17] *In performing any calculation, especially one involving a number of steps, there will always be slight discrepancies.* [PHY 11]

[18] *Another class of forces doesn't involve any direct physical contact. [PHY 11]*

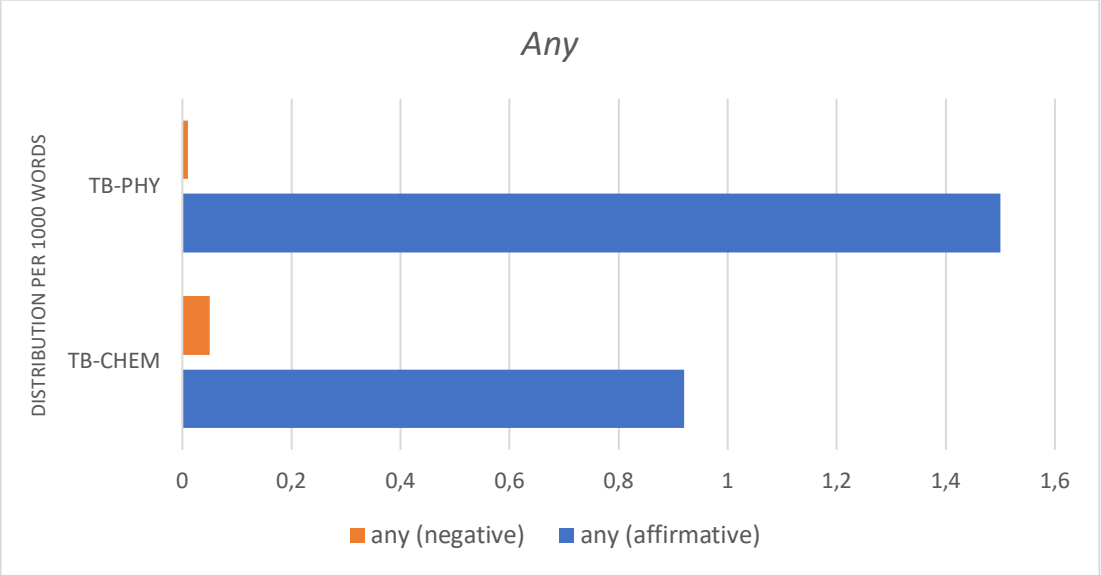


Figure 4.2: Distribution of *any* in affirmative and negative sentences in TB-Chem and TB-Phy.

The fact that *any* is found primarily in affirmative sentences has a great impact for Brazilian learners, as in Portuguese it will have different translations depending on the sentence structure. When in affirmative sentences, it translates into *qualquer/quaisquer* (singular and plural forms respectively), whereas in questions it will mean *algum/alguns (singular/plural),* and in negatives *nenhum/nehuma (masculine/feminine).* Thus, for a Brazilian student who is learning a large amount of English in a short period of time for academic reading purposes, it would be much more profitable to focus on *any* being used in affirmative sentences.

Another point of interest is the presence of modal verbs among the most frequent words in the corpora. In TB-Chem there is the occurrence of *can, will,* and *may,* and in TB-Phy we found *can, will, would* and *must*. However, the distribution of these in comparison to the other modal verbs varies considerably, as we can see in Figure 4.3.

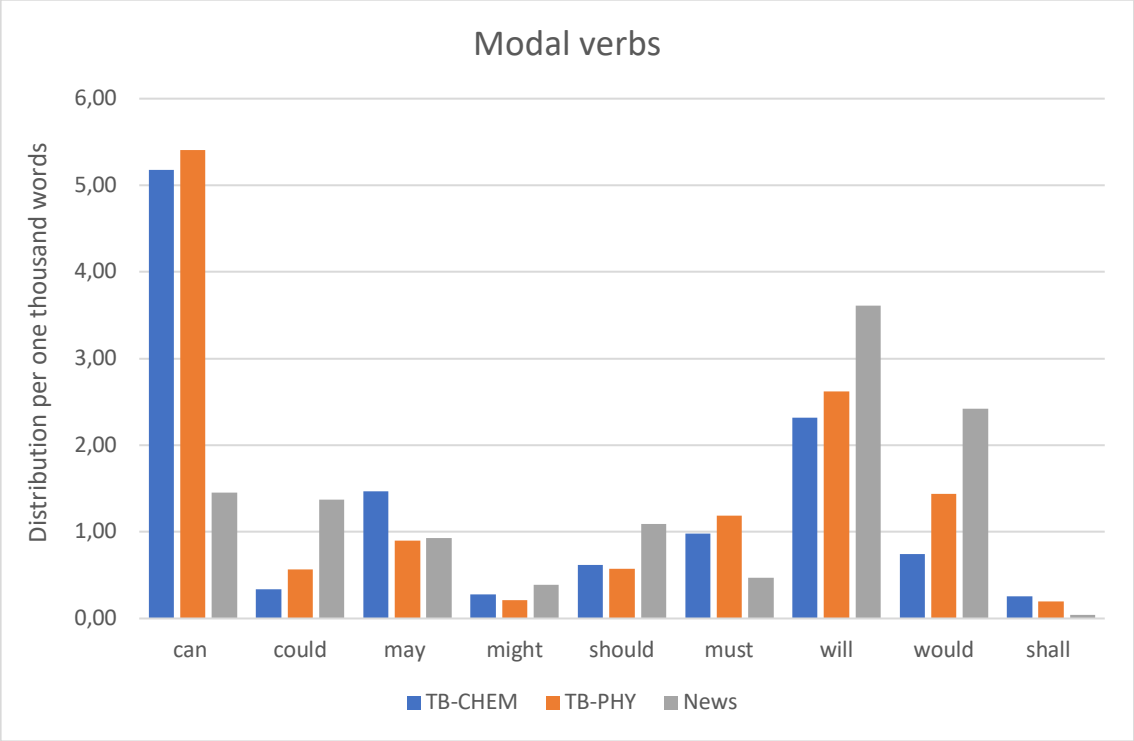


Figure 4.3: Distribution of modal verbs per 1000 words in TB-Chem and TB-Phy, and News

Among modal verbs, Chemistry and Physics textbooks show a strong preference for the modal *can*, which contrasts to news, where *can* is almost five times less frequent . When looking at its common collocates, we find that in both chemistry and physics *can* collocates with the verb *be* nearly 50% of the times. In order to demonstrate how significant this is, the next most common verb that collocates with *can* is *see*, occurring in less than 2% of the cases in both corpora. This preference for the verb *be* happens because most instances of this modal come in passive constructions, where the overt identification of the human agent of the main verb is avoided so that the focus of the message conveyed falls back on the action described by the main verb, as shown in examples 19 and 20:

[19] *Scanning probe microscopes **can** also **be** **used** to move individual atoms into desired positions*. [CHEM 03]

[20] *One of the forces **can** **be** **considered** to be the negative of the other, and they add algebraically to zero*. [PHY 10]

The next most frequent modal in TB-Chem and TB-PHY is *will* and, again, *be* is its most common collocate. This is true because *will* also tends to appear in passive constructions. However, although this pattern is rather common, the verb phrase *will be* is also commonly followed by an adjective phrase, especially in TB-CHEM, as we can see in example 21:

[21] *The volume of the solution **will** then **be greater or smaller** than the sum of its parts.* [CHEM-01]

It is interesting to observe that the verb phrase *will be* is so productive that, in Chemistry, its frequency is equivalent to the sum of the 19 next most common lexical verbs; in Physics, the frequencies of the 16 verbs most frequent lexical verbs added altogether correspond to the frequency of *will be*, as illustrated in Figures 4.4 and 4.5.

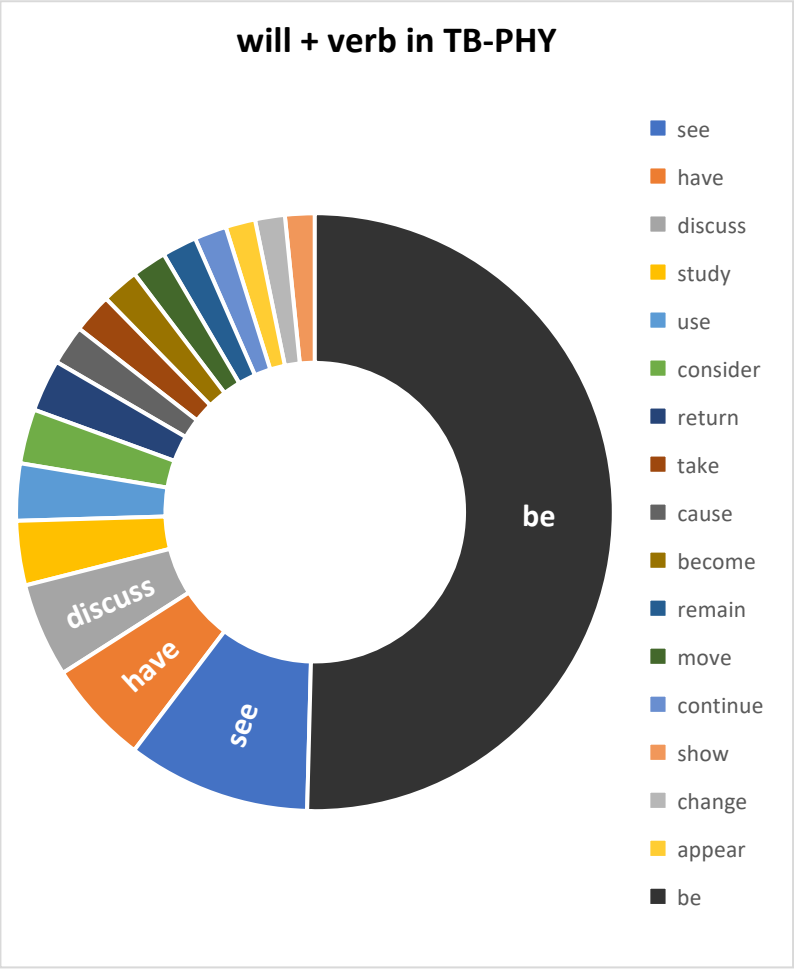


Figure 4.4: Distribution of *will + verb* in percentage in TB-PHY

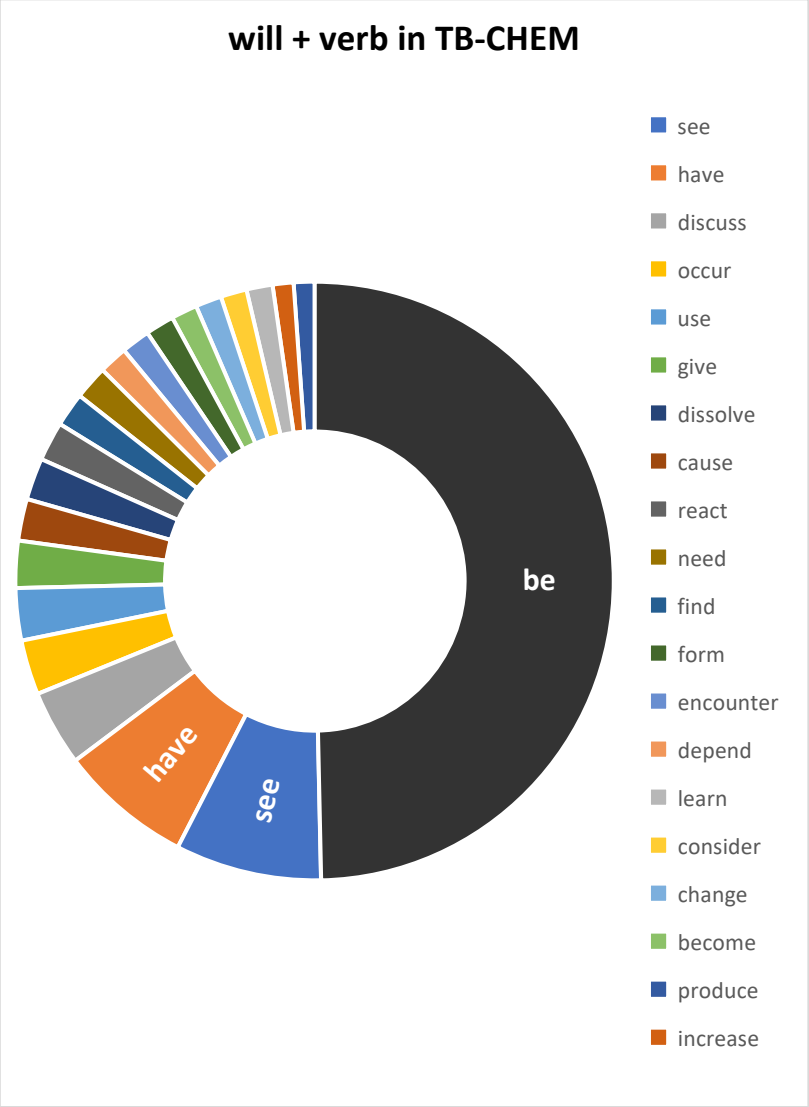


Figure 4.5: Distribution of *will* + *verb* in percentage in TB-CHEM

As for the modal verb *would*, we see it displays a similar discourse function in chemistry and physics, but presents a different use in news. In TB-Chem and TB-Phy, *would* commonly comes in adverbial clauses containing the subordinator *if* with the purpose of presenting some hypothetical situation (example 22). It is also used to describe hypothetical situations even when the subordinator *if* is not present (example 23). In news, on the other hand, *would* is also commonly used as a linguistic resource that journalists make use of in order to avoid being too straightforward when making an assertion, as we can see in example 24.

[22] ***if*** *we have pH data for the titration of an acid as a function of the volume of base added (pH vs. VB), a plot of the first derivative dpH/dVB against VB* ***would show*** *a maximum at the equivalence point.* [CHEM 12]

[23] *Any estimates under ten thousand are small compared with Earth's total population, but a million or more* ***would be*** *alarming.* [PHY 11]

[24] *It* ***would be*** *foolish and unseemly as well as counter-productive for the Prime Minister to enter into an undignified slanging match with him every time he does so.* [News - BNC]

4.2.1. Implications for teachers and materials developers

Applying these findings into teaching and materials development could be done in different ways. The most direct form would be by giving students the opportunity to have access lists of words. Folse (2011), for instance, recommends that when using lists of words for the purpose of teaching vocabulary, they should be used according to frequency, rather than lists containing words that are similar phonologically, orthographically, or semantically. In addition, the wordlists presented in this study (or even larger lists) can be a good starting point for selecting what vocabulary should be included in pedagogical materials.

As indicated in Mishan & Timmis (2015, p. 141), some ways in which wordlists can be incorporated into materials would be through: (1) lists of topic-related vocabulary for each unit; (2) constructing texts to incorporate the target vocabulary; and (3) selecting vocabulary from texts which have been included in the materials for other purposes. The authors, however, alert to the fact that frequency cannot be the only criterion for including or excluding words from materials. In our case, for instance, we see the word *oxygen* figures as a very frequent keyword in Chemistry textbooks, but *nitrogen* does not. Since these two gases make up 99% of dry air in the atmosphere of the Earth[11], it

[11] Source: The National Geographic Society, available at: https://www.nationalgeographic.org/encyclopedia/atmosphere/

would be unreasonable to exclude the word nitrogen from an exercise simply because it is not so frequent in the corpus.

It is also important that students are given the chance to not only read the words, but also write them. This will ensure they not only receive input but also produce output as a means of assimilating those words more effectively. Example activities adapted from Mishan & Timmis (2015) could be as follows:

Example Activity 1:

The words in the box are all taken from the text you have just read. Divide the words in the box into two groups: those that are good electrical conductors, and those that are good electrical insulators:

copper	aluminum	rubber	glass	plastic	mercury
steel	concrete	ceramic	quartz		

electrical conductors	*electrical insulators*

The activity could then be followed by a fill-in-the blanks exercise for more practice. The sentences in the exercise would be taken from selected concordance lines from our corpora, such as:

1. _____ is not only an important material for making food containers, but also an extremely important construction material.
2. One of the most commonly used units of pressure is the millimeter of _____, abbreviated mmHg.

Moreover, leaners should be exposed to new vocabulary from time to time, so as to be able to acquire new words more effectively. In this sense, revision exercises are a must in any material intended to teach vocabulary in order to help students to actively recall words (McCarten, 2007). Words could be revised frequently, but with longer intervals between each time they are recycled: for example, a day after presentation, then a week after, then a month after (Mishan & Timmis, 2015).

4.3. Analysis of N-Grams

In this subsection, we discuss the frequency, functional classification, and syntactic patterns of n-grams (also referred to as lexical bundles) found in our corpora. We begin our analysis with the corpus of chemistry textbooks (TB-Chem). A total of 68 different 4-word bundles were found; yet, for the purpose of this study, the 50 most common bundles will be analyzed, as shown below.

1. as shown in figure
2. can be used to
3. in the case of
4. of the periodic table
5. in the presence of
6. on the other hand
7. in the form of
8. the energy of the
9. as a result of
10. in terms of the

11. is a measure of
12. as a result the
13. the plane of the
14. to the number of
15. the surface of the
16. as a function of
17. in this chapter we
18. one of the most
19. is equal to the
20. is one of the
21. it is possible to
22. at the end of
23. in this case the
24. is the same as
25. the structure of the
26. a large number of
27. it is important to
28. the temperature at which
29. is added to the
30. the formation of a
31. the presence of a
32. the sum of the
33. in the absence of
34. is an example of
35. the same as the
36. in the formation of
37. the difference between the
38. the end of the
39. are said to be
40. in this section we
41. the bottom of the
42. the center of the
43. with respect to the
44. a small amount of
45. an example of a
46. the number of electrons
47. the size of the
48. at the same time
49. in the next section
50. in the same way

In this list contemplating the 50 most common 4-word n-grams, little overlapping was found. In fact, only one overlapping bundle[12] was observed: *A measure of the vs. is a measure of*, which is a case of a longer 5-word bundle that was retrieved as a 4-word sequence. The complete bundle would in fact be *is a measure of the.*

It is interesting to note that while we see n-grams that are typically found in academic textbooks (e.g. *in this chapter we, as shown in figure*), we also find word sequences that seem to be typical of the field of chemistry specifically (e.g. *of the periodic table, the energy of the, the surface of the, the number of electrons*). This could have important implications when we think of the applications of this type of finding for pedagogical materials developers in the context of English for Academic Purposes, especially from a lexical point of view.

Besides being described from a lexical perspective, studies on n-grams investigate these word sequences in relation to their function in context. As discussed in section 2.3, different taxonomies have been proposed, such as the classification into the categories of referential bundles, discourse organizers/text organizers, and stance bundles (Biber, Conrad & Cortes, 2004; Cortes, 2004). Nevertheless, in spite of being widely used by many other authors, this classification presents some variations. For instance, while Biber, Conrad & Cortes (2004) classify *as a result of* and *on the basis of* as referential bundles, Cortes (2004) considers them as text organizers. We therefore chose to adopt the taxonomy proposed in Simpson-Vlach & Ellis (2010), who came up with a relatively more stable and longer list of functional categories.

The 50 most productive 4-word bundles in TB-CHEM were classified and the three functional categories they fall into are: referential bundles, discourse organizers, and stance bundles. Results are shown in Fig. 4.6 below.

[12] Overlapping bundles have been an issue described in previous studies, especially those dealing with the analyses of 3 and 4-word bundles such as Chen and Baker (2010). The authors distinguish between 'complete overlap' and 'complete subsumption'. Efforts have been made to try to automatize the process of eliminating overlappping bundles, such as in the study by Grondona (2015).

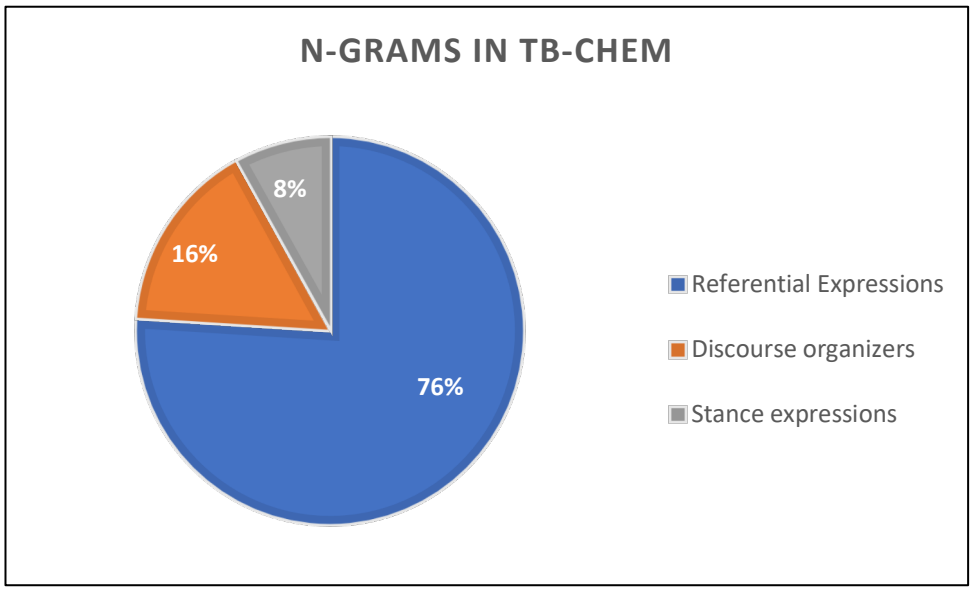


Figure 4.6: Distribution of n-grams in TB-CHEM according to functional categories.

We see from Fig.4.6 that n-grams in Chemistry textbooks rely heavily on Referential expressions, accounting for 76% of the total. Discourse organizers come second in this list, with 16% of the occurrences, followed by stance expressions with 8%. These figures are in line with findings from previous studies that also reported referential expressions as a dominant category in academic textbooks. For instance, Biber & Barbieri (2007) found that referential expressions are considerably more productive than stance expressions and discourse organizers in not only textbooks, but also in academic prose (e.g. research articles) and institutional writing (e.g. university catalogs, brochures).

However, another picture arises when we look at the n-grams from a syntactic point of view. Our analyses is based on the syntactic categories previously defined in Biber et al. (1999, p. 997), as shown below:

Table 4.1: Distribution of 4-word n-grams in TB-CHEM according to syntactic patterns.

Structure	Distribution in %	Examples
Noun phrase + of	32%	*a large number of, a small amount of, an example of a, one of the most, the bottom of the, the center of the, the end of the, the energy of the, the formation of a, the number of electrons, the plane of the, the presence of a, the size of the, the structure of the, the sum of the, the surface of the*
Other noun phrases	6%	*the difference between the, the same as the, the temperature at which*
Prepositional phrase + of	22%	*as a function of, as a result of, at the end of, in terms of the, in the absence of, in the case of, in the form of, in the formation of, in the presence of, in this case the, to the number of*
Other prepositional phrases	18%	*as a result the, at the same time, in the next section, in the same way, in this chapter we, in this section we, of the periodic table, on the other hand, with respect to the*
Passive + prep phrase fragmente	4%	*is added to the, can be used to, as shown in figure*
Anticipatory it + verb/adj	4%	*it is important to, it is possible to*
Be + noun/adjectival phrase	10%	*is a measure of, is an example of, is equal to the, is one of the, is the same as*
Others	4%	*are said to be*

The distribution of n-grams in TB-CHEM, according to syntactic patterns, reveals a strong preference for noun phrase and prepositional phrase in this type of register. Together they are responsible for 78% of n-grams in the corpus, showing that, as far as n-grams are concerned, textbooks employ a much larger number of nouns than verbs. This preference for phrasal rather than clausal-formed bundles structures has also been

found in previous studies, such as Biber, Conrad & Cortes (2004), Chen & Baker (2010), and Dutra & Berber-Sardinha, (2013), indicating that referential expressions are highly productive in academic prose as a whole, not only in the register of textbooks.

As for physics, n-grams present similar characteristics to those in chemistry, both in terms of functional categories and syntactic patterns. A comparison of the 50 most frequent n-grams found in chemistry and physics reveals that nearly 50% of these n-grams are the same in both disciplines, which shows how closely related these disciplines are. In the list below we see frequent n-grams in TB-Phy, and the ones in boldface are those that are also found among the top 50 in chemistry.

1. **in the case of**
2. **on the other hand**
3. **as shown in figure**
4. **in terms of the**
5. the magnitude of the
6. the fact that the
7. **is equal to the**
8. is proportional to the
9. **the surface of the**
10. **at the same time**
11. the speed of light
12. **as a function of**
13. is given by (formula)
14. is shown in figure
15. newton's second law
16. **the size of the**
17. **can be used to**
18. in the direction of
19. of the order of
20. the mass of the
21. **the energy of the**
22. **with respect to the**
23. **in this case the**
24. **it is possible to**
25. **the center of the**
26. the length of the
27. the kinetic energy of
28. **in the form of**
29. **in this chapter we**
30. **is the same as**
31. **in the absence of**
32. **the case of a**
33. in the same direction
34. at the end of
35. **the end of the**
36. to the direction of
37. the motion of the
38. **as a result of**
39. **in this section we**
40. the square of the

41. the velocity of the
42. the radius of the
43. the sum of the
44. in the opposite direction
45. is due to the
46. as long as the
47. we have seen that
48. is known as the
49. of conservation of energy
50. the case of the

As regards their functional categories, n-grams in TB-Phy show a similar tendency to those in Chemistry. Referential expressions are the most prominent, accounting for 82% of occurrences, followed by discourse organizers (14%), and stance expressions (4%), as seen in Fig. 4.7. Indeed, this result does not come as a surprise since, as stated before in this section, referential expressions are the most prominent category in textbooks.

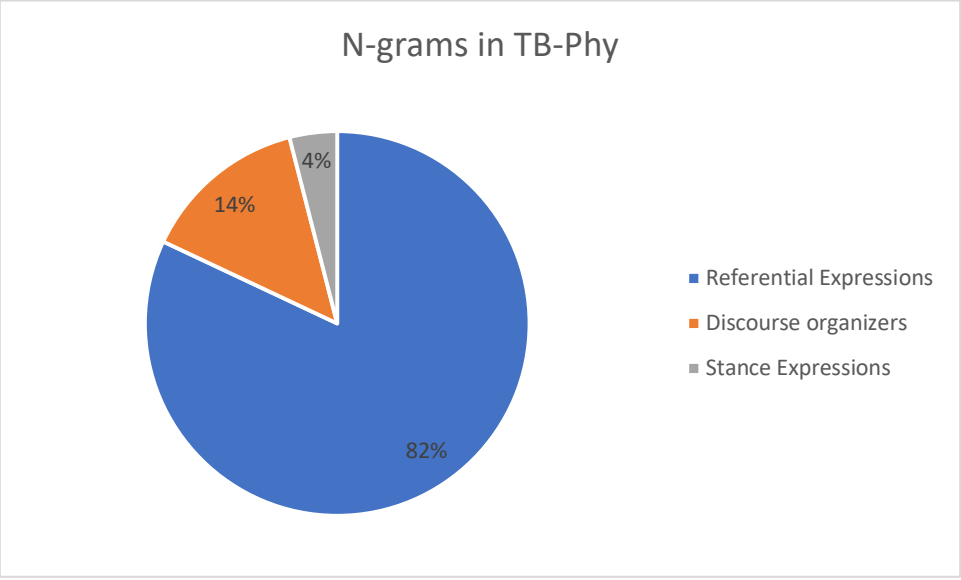


Figure 4.7: Distribution of n-grams in TB-Phy according to functional categories.

It is interesting to note that n-grams in physics include many sequences of words that relate to quantifying expressions such as *the magnitude of the*, *is equal to the*, *the mass of the, the sum of the.* In addition, we see n-grams that refer to the idea of objects

moving in space, which is a common area of investigation in physics, including: *in the direction* of, *the kinetic energy of*, *the velocity of the*, as well as n-grams that are used in the description of formulas, such as *the square of the, is given by (formula).*

As regards their syntactic patterns, n-grams found in physics follow a similar trend to those in chemistry, i.e. a larger concentration of n-grams of the type noun phrase + of, and prepositional phrases (followed or not by 'of'), as seen in the table 4.2.

Table 4.2: Distribution of 4-word n-grams in TB-Phy according to syntactic patterns.

Structure	Distribution in %	Examples
Noun phrase + of	36%	*the magnitude of the, the surface of the, the speed of light, newton's second law, the size of the, the mass of the, the energy of the, the center of the, the length of the, the kinetic energy of, the case of a, the end of the, the motion of the, the square of the, the velocity of the, the radius of the, the sum of the, the case of a, the case of the*
Other noun phrases	2%	*the fact that the*
Prepositional phrase + of	22%	*in the case of, in terms of the, in the absence of, in the form of, as a function of, in the direction of, of the order of, in this case the, to the direction of, as a result of, of conservation of energy*
Other prepositional phrases	18%	*at the same time, in this chapter we, in the same direction, on the other hand, with respect to the, at the end of, in this section we, in the opposite direction, as long as the*

Passive + prep phrase fragment	8%	*can be used to, is given by (formula), as shown in figure, is known as the*
Anticipatory it + verb/adj	2%	*it is possible to*
Be + noun/adjectival phrase	10%	*is equal to the,, is the same as, is proportional to the, is shown in figure, is due to the*
Others	2%	*we have seen that*

Whereas chemistry and physics display a relatively similar pattern as regards functions and syntactic patterns of n-grams, news presents a rather different picture. First, when generating a list of n-grams with the cut-off point previously defined (40 occurrences per million words), we were not able to obtain a comparison list of the 50 most common n-grams; in fact, the output list contains 24 n-grams only, as shown in the list below. This relative small number of n-grams probably has to do with our cut-off point, and our corpus size (approx. 500k words). Other studies on lexical bundles in the news register used different methodologies and were able to come up with longer lists of n-grams, such as Wang (2017), who used a corpus of approximately 11.7 million words and a cut of point of 20 words per million

1. the end of the
2. per cent of the
3. at the end of
4. for the first time
5. from m to m
6. jumped p to p
7. added p to p
8. at the same time
9. p to p after
10. as a result of
11. by the end of
12. is likely to be

13. is one of the
14. the rest of the
15. a member of the
16. the ft se index
17. from p to p
18. the six months to
19. in an attempt to
20. in the six months
21. on the other hand
22. one of the most
23. in the first half
24. will have to be

In our list, we found a 29% similarity between news and chemistry, and 25% similarity between news and physics. This similarity is seen especially because of n-grams that function as discourse organizers, such as *at the same time*, *as a result*, *on the other hand.* In contrast, other n-grams found in news are referential expressions that are content-specific, mainly relating to the topic of finances, such as *in the first half, the ft se index, from m to m*, *jumped p to p, added p to p, p to p after,* and *from p to p ,* where *p* stands for points in the stock market, and *m* stands for *million*, as we can see from examples 25, 26, and 27 .

[25] *Shareholders get a dividend rise of 0.8p to 6.6p. The shares **jumped 8p to 507p**, 32p below their all-time peak of 539p.* [News - BNC]

[26] *Beer profits slumped **from £54.2m to £51m** but the leisure side, Center Parcs, kept up its momentum with maintained occupancy of 90% (...).* [News - BNC]

[27] *Top up at Wessex WESSEX Water saw profits rise 11.3% to £44.3m **in the first half** helped by a tight control of costs and improved results from its non core businesses (...).* [News - BNC]

4.3.1 Implications for teachers and materials developers

Teaching vocabulary from a formulaic perspective could a good means of helping students improve their proficiency in a language. As Cortes (2006) puts it, proficient language use can be measured by the use of frequent word combinations. In this sense, exposing students to frequently occurring n-grams, could help them familiarize with academic discourse, thus facilitating language acquisition, as shown by Neely and Cortes (2009) and Eriksson (2012). Although most studies approach formulaic sequences

in the context of the teaching of writing skills - especially academic writing - we believe that n-grams could also serve the purpose of giving students relevant input for the purposes of vocabulary acquisition and practice in the context of English for Reading.

To illustrate that, we now bring a sample activity that could be used in the context of teaching reading skills aimed at enhancing vocabulary. After generating the lists of frequent n-grams in the corpora, we can use selected concordance lines so as to expose students to the wider context in which those n-grams are inserted. In example activity 2 we chose to work with the lexical bundles *in the presence of* and *in the absence of*. As a first step, the teacher should pre-teach the words *present* and *absent* by writing on the board the name of a student who is present in class that day. Then the teacher elicits from students the name of a student who is not present in class, and also writes it on the board as follows:

"Student A is present in class today "

"Student B is absent in class today" → absent = not present

The teacher will then show how the words *present* and *absent*, used as adjectives, can be transformed into nouns with the addition of the suffix -*ence*: present → presence; absent → absence. The teacher then writes the following sentence on the board and asks students to translate it: "*metal rusts <u>in the presence of</u> oxygen*". After that, the teacher elicits from students what the opposite of "in the presence of" would be.

Students will then do an exercise in which they are requested to fill in the blanks with the word sequences *in the presence of* and *in the absence of.* Since they will be exposed to authentic language taken from TB-Chem and TB-Phy, it is advisable that students do this activity in pairs in order to help one another with unknown vocabulary. The disciplines to which the sentence excerpts belong are shown in square brackets.

Example activity 2

A) Read the sentences below and fill in the blanks with in the presence of or in the absence of

1. _____, a moving object with mass m will travel in a straight line with velocity vector v. [Physical Chemistry]

2. A number of plants (e.g. cassava, sugar cane, some varieties of white clover) and fruits are natural sources of HCN. The origins of the HCN are cyanoglucosides such as amygdalin (e.g. in almonds, peach and apricot stones, apple seeds) and linamarin (in cassava). The release of HCN from certain plants (cyanogenesis) occurs _____ specific enzymes. For example, the enzyme linamarase is present in the cell walls of cassava plants. [Inorganic Chemistry]

3. Activated charcoal is a finely divided form of carbon and is manufactured from organic materials (e.g. peat, wood) by heating _____ of reagents that promote both oxidation and dehydration. [Inorganic Chemistry]

4. Alkynes can be completely reduced to alkanes by reaction with two molar equivalents of hydrogen gas _____ catalysts such as platinum or palladium. [Organic Chemistry]

5. All photodetectors produce a small response _____ light. This dark current could arise from spontaneous emission of electrons from the cathode of a photomultiplier tube or spontaneous generation of electrons and holes in a semiconductor device. [Analytical chemistry]

6. Bodies that are rotating tend to remain rotating, while nonrotating bodies tend to remain nonrotating. _____ outside influences, a rotating top keeps rotating, while a top at rest remains at rest. [General Physics]

7. It is well known that, _____ air resistance, all objects dropped near the Earth's surface fall toward the Earth with the same constant acceleration under the influence of the Earth's gravity. [General Physics]

8. One of the most important consequences of hydrogen bonding is that it causes water to be a liquid rather than a gas at 25 °ΔC. Calculations indicate that _____ hydrogen bonding, water would have a boiling point near –80 °ΔC and thus would be a gas at room temperature. [Organic Chemistry]

9. Small charged particles can accelerate _____ an electric field. [Electrodynamics]

10. The idea that momentum is conserved when no external force acts is elevated to a central law of mechanics, called the law of conservation of momentum, which states:

_____ an external force, the momentum of a system remains unchanged. [General Physics]

11. The important thing to keep in mind is the idea of net force. In a vacuum or in cases in which air resistance can be ignored, the net force is due only to gravity. _____ _____air resistance, however, the net force is less than the force due to gravity because of the opposite force arising from air resistance. [General physics]

When correcting the activity, the teacher could ask students to explain the reason for choosing their answers by explaining the context to which those sentences belong. This is also a good opportunity to explain any unknown vocabulary or draw their attention other relevant structures that appear in the sentences.

4.4. Verb Tense and Aspect

This subsection addresses how tense and aspect behave in chemistry and physics textbooks. The news corpus is also investigated for comparison purposes, but our analysis is not focused on detailed discussions of this register. We investigate tense and aspect both quantitatively and qualitatively by looking at total counts of each tense/aspect marker in our three corpora (TB-Chem, TB-Phy, and News) as well as common patterns and functions they present. The verb tenses/aspects studied were: present simple, past simple, present progressive, past progressive, present perfect, and past perfect.

4.4.1. Distribution of verb tense and aspect across registers

As regards their distribution, tense/aspect markers are especially distinct concerning the present simple and past simple. As we can see in Fig 4.8, both chemistry and physics textbooks have a strong preference for the use of the present simple, while news will tend to use the past simple more frequently. For all other verb tenses/aspects, news outnumbers chemistry and physics with respect to frequency counts. Below we break down our discussion into sections, where we treat each verb tense/aspect separately.

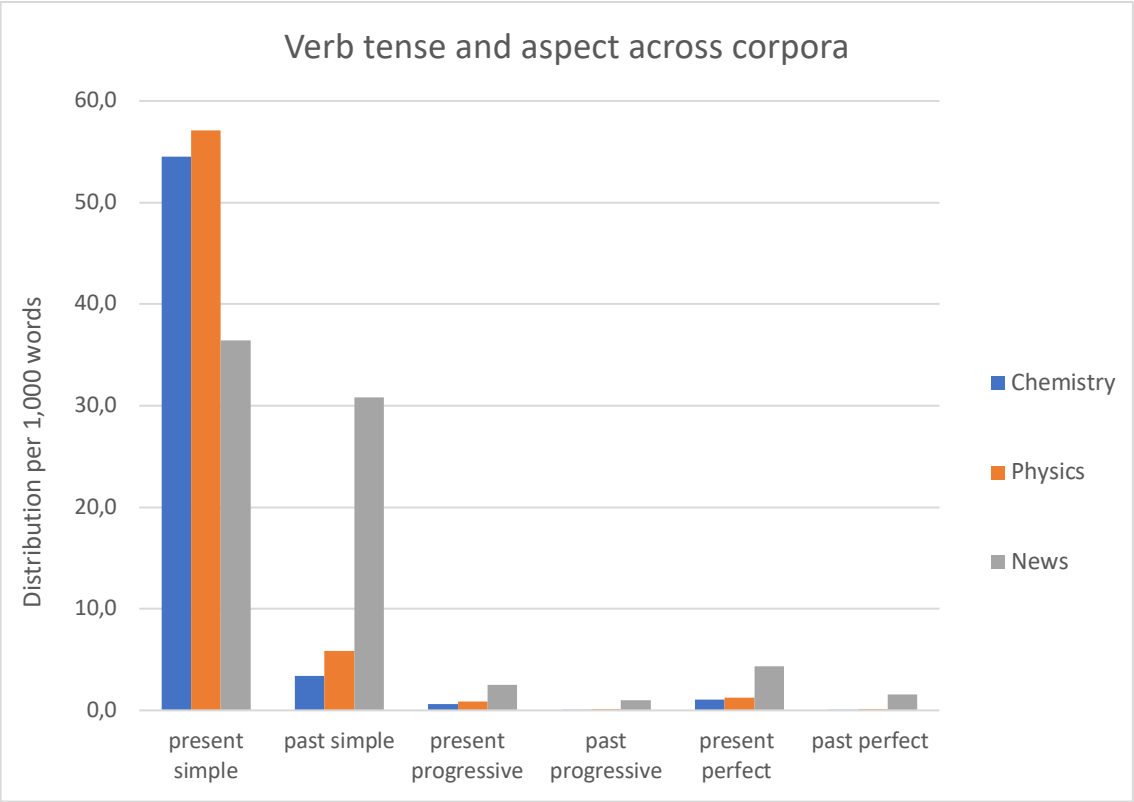


Figure 4.8: Distribution of verb tense/aspect in TB-CHEM, TB-PHY, and News per 1,000 words.

4.4.2. The present simple

The present simple is the most used tense-aspect marker in the three corpora. In the distribution per one thousand words, chemistry and physics show similar counts (54.7 and 57.2, respectively), while news presents a total count of 36.5, as shown in Fig 4.9. Notably, it is the only tense-aspect marker in which chemistry and physics display a larger number of occurrences than news.

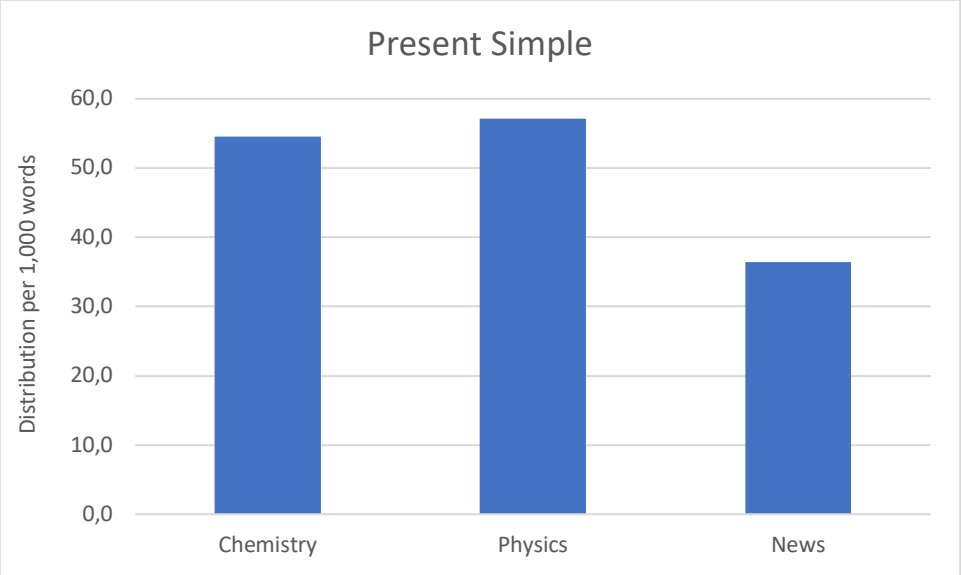


Figure 4.9: Distribution of the present simple in TB-CHEM, TB-PHY, and News per one thousand words

In chemistry and physics textbooks, the simple present is mostly employed in its prototypical use, i.e. to refer to actions and events in the present. Typically, it is used in TB-Chem and TB-Phy for three purposes: i) to refer to scientific facts or general truths (examples 27 and 28), ii) to describe or exemplify natural phenomena (examples 29 and 30), or iii) to explain/describe scientific experiments/calculations (example 31 and 32).

[27] *Polar solvents **dissolve** polar and ionic solutes; nonpolar solvents **dissolve** nonpolar solutes*. [CHEM 01]

[28] (...) *a massive brick wall with a thickness of 24 cm, for instance, **has** a critical frequency of about 100 Hz*. [PHY 03]

[29] *As an example of a process that **takes** place spontaneously yet **absorbs** heat, **think** about what **happens** when you **take** an ice cube out of the freezer. The ice spontaneously **melts** to give liquid water above 273 K* (...). [CHEM 01]

[30] *Because we **are** almost always in contact with Earth, gravity **is** more familiar to us as something that **presses** us against Earth rather than as something that **accelerates** us*. (PHY 10)

[31] *We then **balance** the half-reactions individually and **add** them together.* [CHEM 02]

[32] *When it **hits** the seat, the air flow **is** interrupted and the under pressure caused by it **vanishes.*** [PHY 03]

Although the present simple is used to refer to actions and events in the present, it can also be used to refer to the past or even the future (Greenbaum & Quirk, 1990), especially in fiction and conversation registers (Biber et al., 1999). In chemistry and physics, the present simple with a past meaning (historic present) is rarely used; however, it is more common when found in sentences that project the message to the future. This is especially the case when it appears in dependent (subordinate) clauses, particularly in adverbial clauses, as shown in examples 33 and 34, where subordinators are underlined and the verb phrase is in boldface.

[33] *If the flask **contains** 100 mL at the end point, what is the error in the titration in milliliters* (...)? [CHEM 12].

[34] (...) *the problem is completely solved as soon as we **know** the global electromagnetic energy density*. [PHY 06]

In fact, the present simple appears rather commonly in dependent clauses in our corpora. In chemistry, this happens in approximately 26% of clauses carrying the present simple, and 23% in physics. Among these, adverbial clauses are the most frequent, followed by noun and adjective clauses. In TB-Chem the most common subordinators are *that*, *which*, and *when*, while in physics the most common subordinators are *that*, *which*, *when*, and *if*.

Finally, a noticeable difference between chemistry and physics on one side, and news on the other, is the lexical verbs used in the present simple in these corpora. In chemistry the most productive verb forms are *see*, *shows*, *contains/contain*, *occurs*, *gives*, *increases*, *depends*, *need*, and *requires.* In physics we have *see*, *consider*, *shows*, *moves*, *depends*, *use*, *become*, *gives*, *means*, and occurs. News on the other hand, present a

rather different list, which includes: *say/says*, *believe*, *seems*, *want*, *make*, *think*, *know*, *comes*, and *get*.

These lexical choices demonstrate that chemistry and physics on one side, and news on the other, portray clearly different discourse functions, as we can see when we classify them according to Biber et al's. (1999) semantic domains. Chemistry and physics rely more on verbs that activity verbs expressing non-volitional actions (*show*, *move*, *give*), occurrence verbs that typically report physical events (*become*, *increase*, *occur*), existence verbs, which report a state that exists between entities (contain), and causative verbs, which indicate that some person or inanimate entity brings about a new state of affairs (require). News, on the other hand, presents more communication verbs used to report people's statements and declarations (say/says), and mental verbs, which denote activities not involving physical action and experienced by humans (*think*, *believe*, *know*, *want*)

4.4.3. The past simple

The past simple is the second most used tense-aspect marker in the three corpora, being less frequent than the present simple only. Chemistry presents the least number of occurrences of the past simple per one thousand words (2.51.), followed by physics (5.16), and news (30.88), as shown in Fig. 4.10.

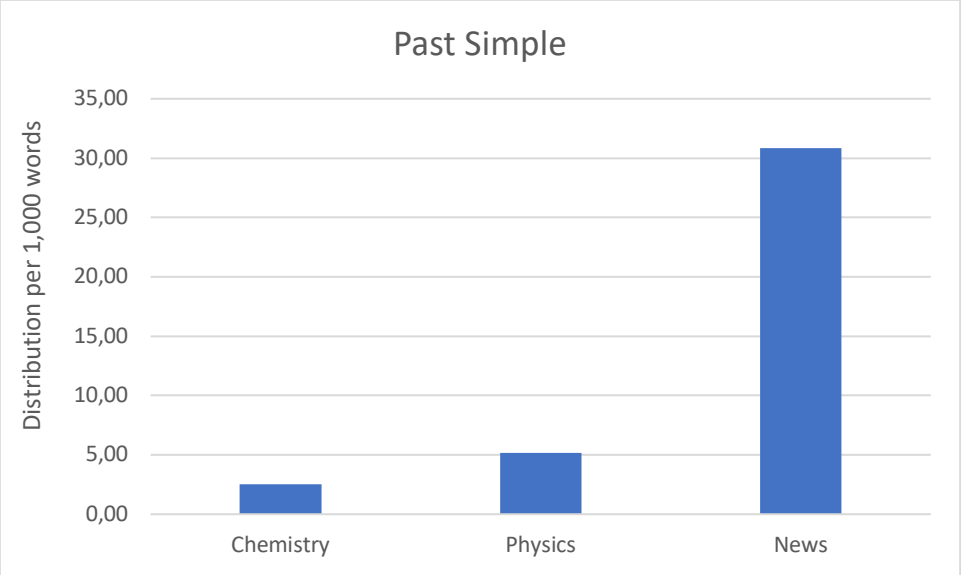


Figure 4.10: Distribution of the past simple per one thousand words in TB-Chem, TB-Phy, and News.

Although the frequency of the past simple might look similar for chemistry and physics, this is not exactly the case. Total counts are masked by the discrepant number of occurrences of the past simple in the news register, which makes the bars in the chart look misleadingly similar. In fact, physics presents almost 106% more occurrences of this tense/aspect marker than chemistry. Yet, In both chemistry and physics textbooks, the past simple presents two main rhetorical functions: i) to introduce a historical background to what is being discussed (examples 35 and 36); and ii) to refer to a discussion previously done in the chapter/book (examples 37 and 38).

[35] *In 1926 Erwin Schrodinger **published** an equation that gives correct solutions for the energy levels of the hydrogen atom. Shortly afterward, Heitler and London **showed** that the Schrodinger equation, as it has come to be called, predicts the existence of (...)* [CHEM 05]

[36] ***Were*** *there more-fundamental physical laws beneath the statistical rules? Einstein **thought** there **were**, but Bohr **said** no, there **was** no deeper level to be understood*. [PHY 17]

[37] *In Chapter 8 we **saw** that primary alcohols are oxidized to aldehydes, which are subsequently easily oxidized to acids.* [CHEM 14]

[38] *As we **learned** in the preceding section, the work done by the gravitational force on an object moving between any two points near the Earth's surface is (...)* [PHY 12].

Interestingly, physics makes much more use of historical backgrounds (as illustrated in examples 35 and 36) than chemistry does, which explains why physics textbooks employ the past simple more frequently. In addition, physics - but not commonly chemistry - uses the simple past to describe hypothetical situations so as to illustrate a given physical phenomenon, commonly using a conditional clause (relative clause) as illustrated in example 40.

[39] *If you **exerted** the same force on every part of the ball, the ball would remain perfectly round as it **accelerated**. But if you **pulled** harder on one side than the other, the different pulls would stretch the ball*. [PHY 10]

Although physics presents a higher number of occurrences for the past simple, chemistry displays a more even distribution of this tense-aspect marker among its disciplines. Except for one book of inorganic chemistry, the past simple is distributed somewhat evenly among analytical, inorganic, organic, general, and physical chemistry. This, however, is not true when it comes to physics, where the past simple is preferred in introductory books of general physics, and also in modern physics, thermodynamics and electrodynamics. In contrast, it is very little employed in classical mechanics and relativity.

As for verbs used in the past simple, we see they are quite varied and reflect the discourse functions this tense-aspect marker possesses. In chemistry, the most common lexical verbs used in the past simple include: *bonded*, *used*, *found*, *saw*, and *emitted*. In physics, among the most common lexical verbs we have: *saw*, *found*, *discovered*, *made*, *showed*, and *learned*. Interestingly, verbs such as *see and learn* rank among the most frequent because authors commonly use them to refer to a discussion previously done in the chapter/book, as stated earlier in this section.

In contrast, for news, among the lexical verbs used in the past simple in we have *said*, *told* and *added*, which are used in news reports to inform what people said about/in a given event. Other very productive lexical verbs used in news include: *took*, *came*, *made*, and *went*. Indeed, news presents a much higher frequency of verbs in the past simple (over 6 times more than chemistry and physics), as seen in Fig. 4.10; yet, this does not come as a surprise, as one of the basic discourse functions of news texts is to narrate past events.

4.4.4. The present perfect

The present perfect is the third most common tense-aspect marker is our corpora, being less frequent than the present simple and past simple, and more frequent than the past perfect, present progressive and past progressive. It is worth noting that for this analysis, we have considered occurrences of the present perfect in the active voice.

Chemistry and physics display a similar distribution for the present perfect, 1.04 and 1.28 per one thousand words respectively. News, on the other hand, presents a higher frequency of this time reference, 4.33, as shown in Fig. 4.11.

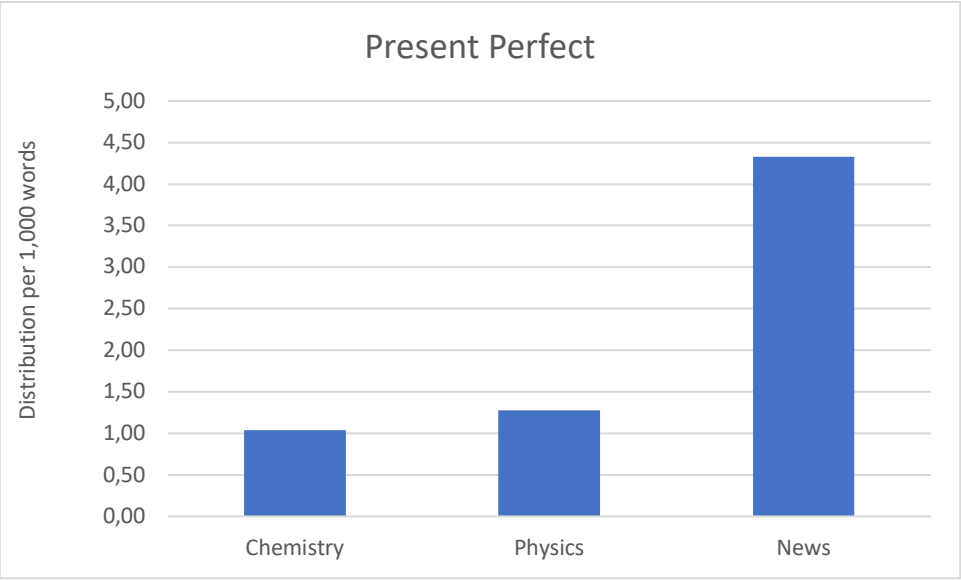


Figure 4.11: Distribution of the present perfect in TB-Chem, TB-Phy, and News.

Typically, the present perfect refers to events taking place in the past but connected with the present; yet, the function it serves in chemistry and physics diverges from that found in the news register. In chemistry and physics, the present perfect is mostly controlled by the pronoun *we*; 50.2% and 58.7%, respectively. This tendency shows that it is used for two purposes: i) in a dialogical manner, where the writer interacts with the reader by reminding them of some information previously discussed in the book, as shown in examples 40 and 41, and ii) to refer to a an action, operation or definition that has just been mentioned or taught in the chapter, illustrated in examples 42 and 43.

[40] *When investigating the fabric of atoms, **we have seen** that atomic absorption and emission spectra produced identical line spectra*. [CHEM 07]

[41] *What **we have discussed** so far is that the Copenhagen interpretation of quantum mechanics defines matter in terms of the measurements of* (…). [PHY 09]

[42] *In the example above, **we have replaced** one-fourth of the Si4þ by Al3þ*. [CHEM 15]

[43] *As in this treatise **we have adopted** the latter method of investigation, we naturally adopt the second expression as giving the most significant form to the kinetic energy*. [PHY 02]

Interestingly, in TB-Chem and TB-Phy, when the present perfect does not follow the pronoun *we*, it is mostly controlled by non-human subjects. This can be explained by the nature of the content found in chemistry and physics textbooks, i.e. the description and explanation of natural phenomena. In this respect, it loses its dialogical function and assumes a pragmatic purpose of connecting past events to the present moment, as shown in examples 44 and 45. In the examples, subjects are underlined and verb phrases are in boldface.

[44] (…) *when all the liquid water **has become** vapor at 100 ℃, the temperature begins to rise once again*. [CHEM 03]

[45] (…) *the magnetic moment is again aligned along the positive z-direction, but <u>the sign of the spin state vector</u>* ***has changed***. [PHY 19]

Unlike chemistry and physics, the present perfect used in news is not commonly followed by the pronoun *we*; hence, it does not serve a dialogical purpose, nor does it refer to actions that have just been mentioned in the text. Instead, it relates more to what Quirk et al. (1985, p.192) refer to as "[a] state leading up to the present moment", or "indefinite event(s) in a period leading up to the present". This means that when events reported in news texts employ the present perfect, they intend to inform about an event that took place in the past while emphasizing its high significance in the present. This, according to Chovanec (2014), has pragmatic consequences in the news texts because "by using the simple past tense, the writer makes the commitment to refer to the event as completed. The present perfect, by contrast, entails the writer's evaluation of the event as relating to, and being relevant for, the reader" (p.214). This assertion is illustrated in examples 46 and 47.

[46] *Proposals for the development of a Taurus system* ***<u>have been</u>*** *under consideration for much of this decade*. [News – BNC]

[47] *British Department of Transport officials* ***<u>have expressed</u>*** *concern about the probable restrictions and urged 'an open and free investment climate'*. [News – BNC]

In each of the examples presented in this section, different main verbs have been shown; however, some main verbs are more commonly used with the present perfect than others, depending on the corpus we are looking at. As this analysis does not include the present perfect in the passive voice, our results differ from those found in Biber et al. (1999, p.463), where they state that "The verb *been* is the most common present perfect form in all registers except conversation" [emphasis added]. Unlike the present simple, where the verb *be* is the most productive, its participle form *been* used for the present perfect does not follow the same trend, as we can see in Fig. 4.12.

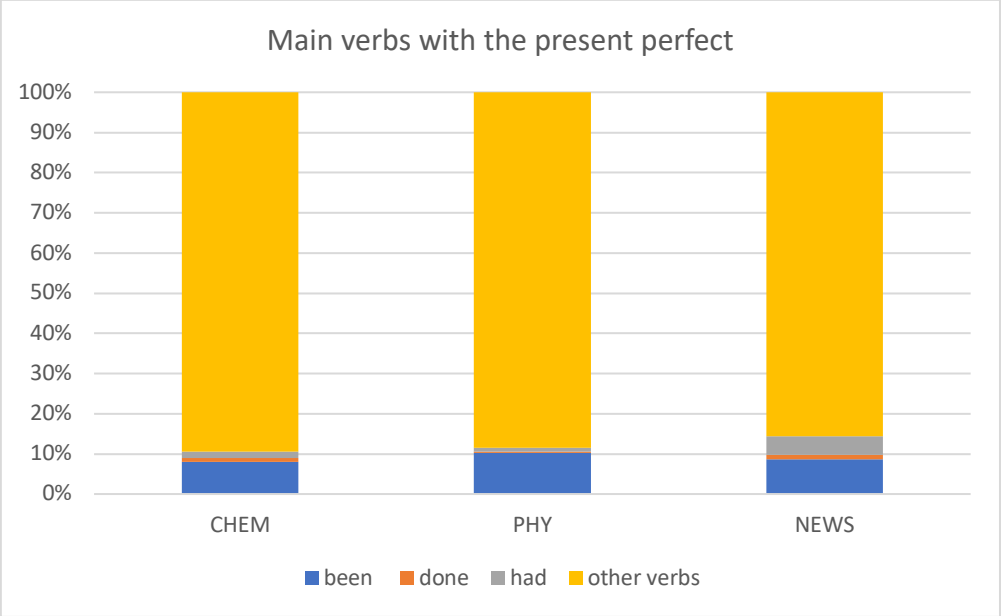


Figure 4.12: Distribution of main verbs used with the present perfect in percentage in TB-Chem, TB-Phy, and News.

The fact that authors of chemistry and physics textbooks establish a dialogue with the reader is reflected in the verbs they use. This explains why phrases such as *as we have already seen* appear in the texts. Among the verbs used as a main verb with the present perfect, the most productive include:

- For chemistry: seen, become, developed, mentioned, described;
- For physics: seen, discussed, shown, learned, reached, moved;
- For news: got, taken, come, made, given, found, gone.

Another difference that is noticed when we look at the distribution of the present perfect in TB-Chem and TB-Phy has to do with disciplinary variation. In chemistry, although organic chemistry employs the present perfect more frequently than the other disciplines, there is not such a large difference in the distribution of this tense-aspect marker among physical, inorganic, general, and analytical chemistry, as shown in Fig. 4.13.

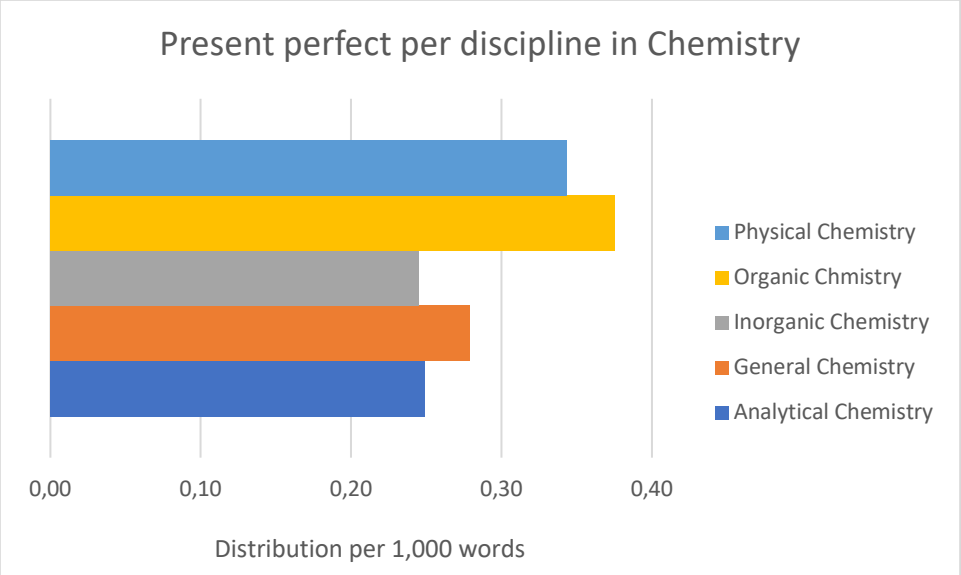


Figure 4.13: Distribution of the present perfect among disciplines in TB-CHEM per 1,000 words

In contrast, physics presents great variation in the distribution of the present perfect when we look at its disciplines. Notably, three disciplines use it more frequently, namely thermodynamics, solid state physics and general physics. Five other disciplines employ the present perfect more moderately: optics, classical mechanics, quantum mechanics, electromagnetism, and wave motion. In contrast, the disciplines that least use the present perfect are electrodynamics, modern physics, acoustics, condensed matter physics, and relativity, as shown in Fig 4.14.

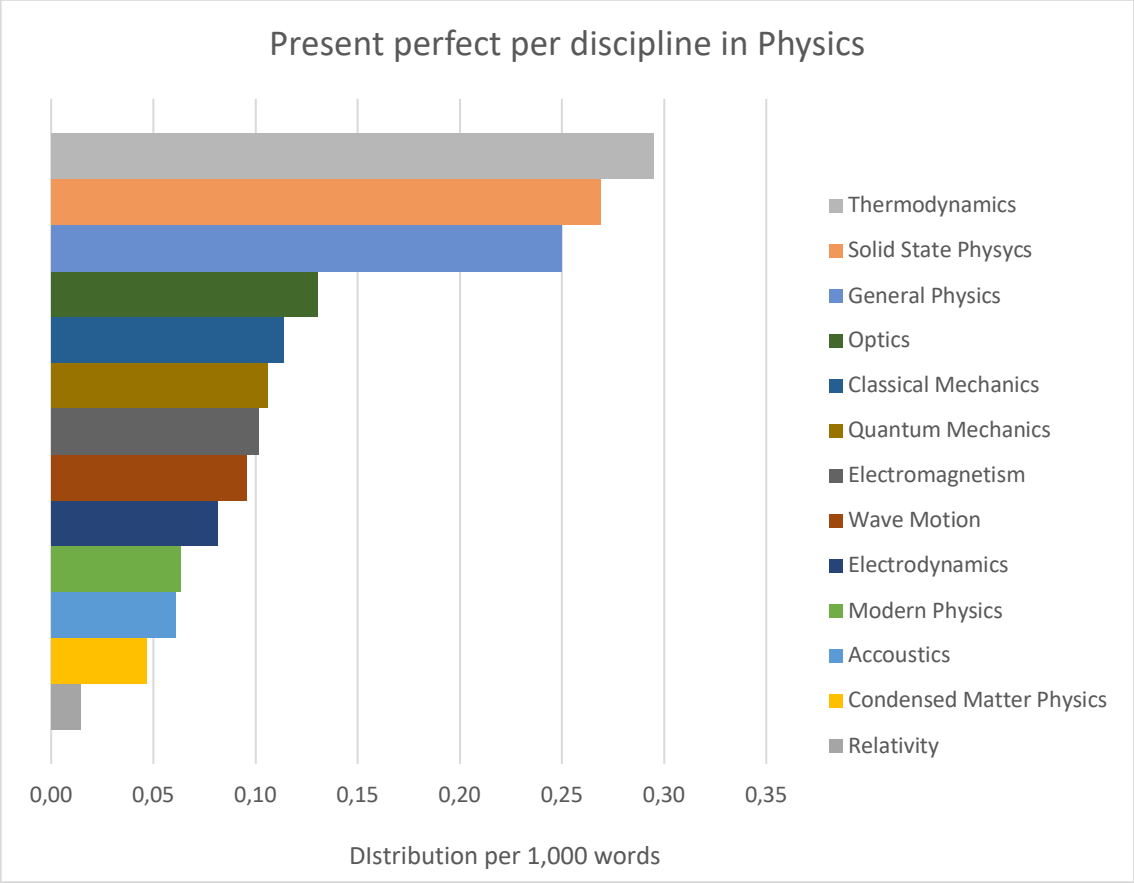


Figure 4.14: Distribution of the present perfect among disciplines in TB-Phy per 1,000 words

Finally, when teaching the present perfect, English for Reading textbooks frequently state that the word *since* is commonly used with this tense-aspect marker. It is, therefore, implied that *since* is used as a preposition (example 48). Nevertheless, in chemistry and physics textbooks, *since* is predominantly used as a subordinate conjunction, meaning *for the reason that*, or *because*, as illustrated in example 49. In addition, it is also very rarely used as an adverb (example 50), especially in chemistry and physics. The distribution of *since* according to its part of speech classification can be seen in Fig. 4.15.

[48] *It has been recognized **since** 1923 that boron is an essential plant micronutrient*. [CHEM 08]

[49] *(...) if we want to compare the phases of the vibrations, we must be more careful, **since** in Optics, several micrometers represent several wavelengths* [PHY 06]

[50] *Miss Wright simply walked out of the hospital at 2pm on Sunday, May 23, and has not been seen* ***since***. [NEWS - BNC]

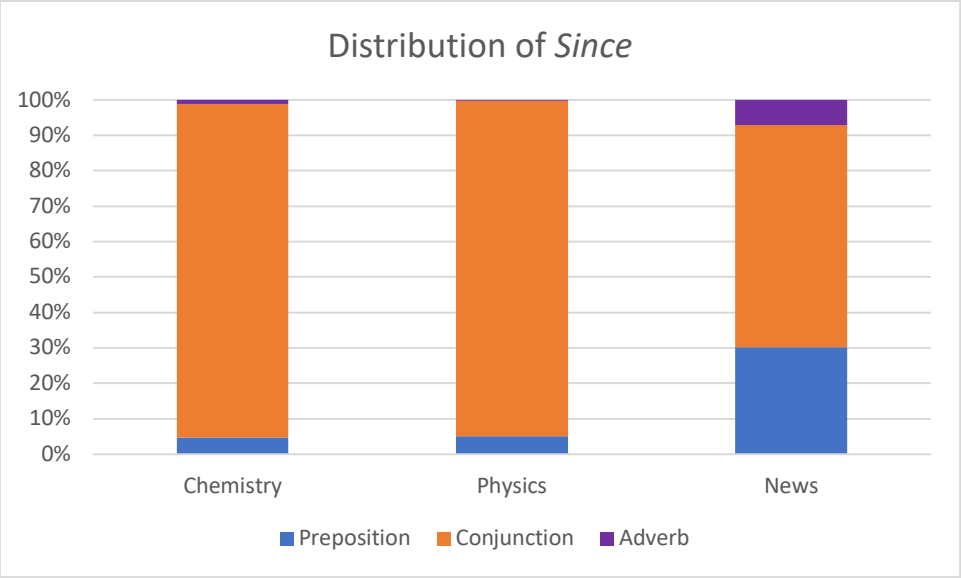


Figure 4.15: Distribution of *since* in TB-CHEM, TB-PHY, and News, in percentage

4.4.5. The past perfect

The past perfect is the second least used tense-aspect marker in our corpora, being more frequent than the past progressive only. This is especially true for chemistry and physics, whose counts total 0.06 and 0.13 occurrences per one thousand words, respectively. News, on the other hand, presents 1.58 occurrences of the past perfect per one thousand words, the most productive among the three corpora, as shown in Fig. 4.16.

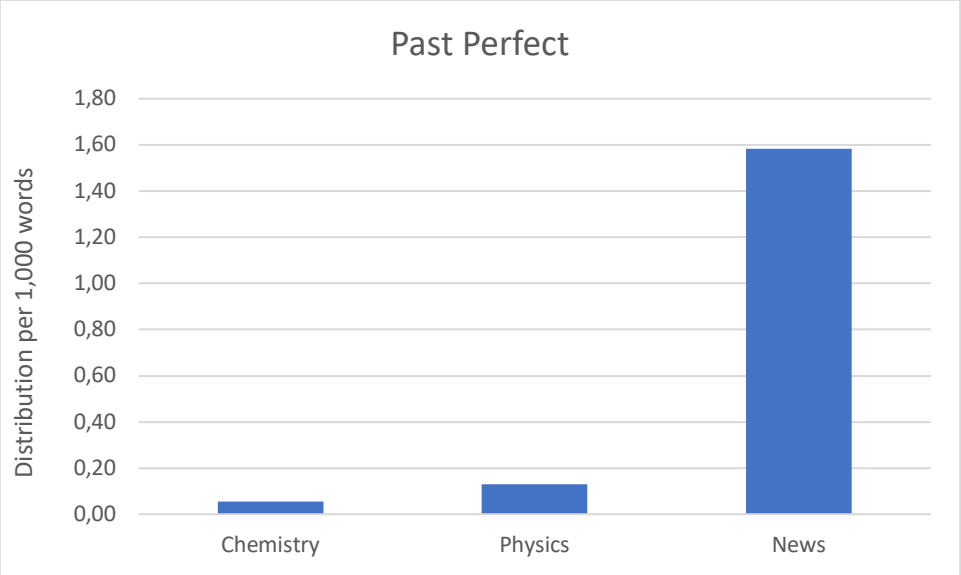


Figure 4.16: Distribution per one thousand words of the past perfect in TB-Chem, TB-Phy, and News.

Although the past perfect is more frequent in Physics than in Chemistry, it portrays a similar function in both corpora. In most instances, the past perfect is used in textbooks to narrate past historical facts or events in order to contextualize what is being discussed in the chapter. In chemistry, this tendency is particularly more evident in the field of inorganic chemistry, whereas in physics it is more prominent in thermodynamics, as illustrated in examples 51 and 52.

[51] *In ancient times, sodium carbonate was obtained from the places where brine solutions* ***had evaporated*** *and dry lake beds*. [CHEM 15]

[52] *Newton* ***had*** *already* ***noticed*** *that, in contradiction to dynamics, an irreversible loss of motion is involved in each hard collision*. [PHY 17]

In contrast, news texts very commonly employ the past perfect to narrate events described by third parties (i.e. the writer informs what other people said), by using indirect speech (reported speech). It is know that when indirect speech is used, the verb phrase typically carries the past tense and what varies is the aspect marker. Therefore, something that was said in the present simple can be reported in past simple; likewise, something that was said in the past simple or present perfect can be reported in past

perfect. Examples 53 and 54 show sentences carrying the past perfect in indirect speech taken from the news corpus.

[53] *The defendant's mother became suspicious because she **had overheard** a conversation about drugs*. [NEWS - BNC]

[54] (...) *a police spokesman said it seemed nobody **had broken** in, though children were seen running away*. [NEWS - BNC]

Whilst in chemistry and physics the main verbs that form the past perfect are rather varied, in news some verbs are more frequently used with this tense-aspect marker. Among the most productive verbs used for the formation of past perfect in news, we have: *taken*, *come*, *given*, *suffered*, *gone*, *made*.

In addition, another notable difference among the three corpora lies in the use of adverbs in the middle of the present perfect construction, that is, an adverb between the auxiliary *had* and the main verb. Constructions such as *x had already pointed out*, or *x had just started*, appear in 17% of the occurrences of the past perfect in physics, and in 19% in news. Curiously, chemistry does not present any instances of this type of construction. It does, however, present a few occurrences of an adverb between the auxiliary *had* and the main verb when we take passive voice constructions into consideration. In physics, the most used adverb is *already*, while news presents a variety of adverbs to fill in the same slot, such as *already*, *also*, *just*, *never, and recently*. Examples 55, 56, and 57 illustrate these occurrences.

[55] *Compton and Pauli **had previously suggested** that the electron could possess an intrinsic magnetic moment* (...). [PHY 19]

[56] *The MiGs **had also entered** the zone 20 minutes earlier, but fled to safety when asked to identify themselves*. [News-BNC]

[57] *Due to the difficulty of controlling conditions in the real world, ocean tests on the scale he suggested **had never been done***. [CHEM 04]

4.4.6. The present progressive

The present progressive is the third least used tense-aspect marker in our corpora, being more frequent than the past perfect and the past progressive, and less frequent than the present simple, the past simple, and the present perfect. Its distribution per one thousand words totals 0.6 in chemistry, 0.9 in physics, and 2.5 in news, as shown in Fig. 4.17.

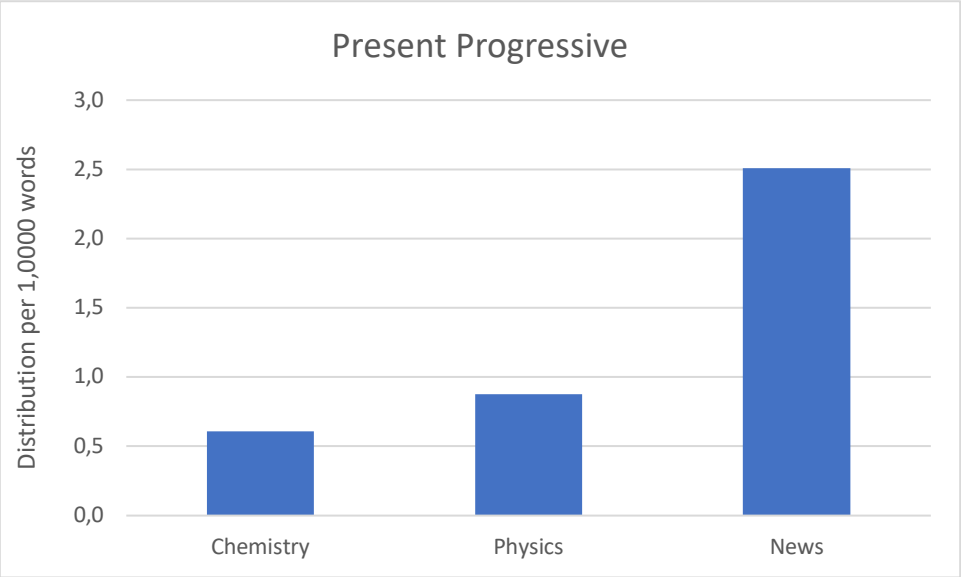


Figure 4.17: Distribution of the present progressive per 1,000 words in TB-Chem, TB-Phy, and News

When explaining certain processes, textbook authors choose the present progressive to make their explanation more realistic, as if it were actually happening at the moment. This creates a visual effect for the reader who is interpreting the message. In examples (58) and (59), the choice for the present simple instead of the present progressive would probably sound as though the events were static rather than dynamic.

[58] *For example, imagine a source **is radiating** over a broad range between 300 and 700 nm*. [CHEM 12]

[59] *Because the molecules **are moving** with high speeds in random directions, many of them will cross the imaginary barrier*. [PHY 11]

In addition, the present progressive is used in chemistry and physics to talk about ongoing events that started in the past and are likely to continue. Semantically, it performs a similar function as that of the present perfect progressive (e.g. *x has been happening*), as shown in examples 60 and 61. This function is also very commonly found in the news corpus (example 62)

[60] *The concentration of water vapour in the atmosphere has remained steady over time, but concentrations of some other greenhouse gases **are rising***. [CHEM 06]

[61] *Photovoltaic cells **are gaining** popularity in building materials, roofing, tiles, and even transparent windows*. [PHY 18]

[62] *In today's markets, where increasingly companies **are looking** to raise funds beyond their own national borders, agreement on an objective standard for measuring the financial health of a business is vital*. [News - BNC]

The present progressive is also employed in chemistry and physics textbooks when authors are establishing a closer dialogue with the reader. In this case, the present progressive is controlled by subject pronouns *you* and *we*, as illustrated in examples 63 and 64.

[63] *Thus, when you **are calculating** the net energy released or absorbed during a reaction, you need to be careful about plus and minus signs*. [CHEM 03]

[64] *To put it another way we **are taking** averages over an interval Δt, within which there are unphysical processes as noted*. [PHY 17]

Although the present progressive is not very used in chemistry and physics textbooks as a whole, in some disciplines its use is virtually non-existent. In chemistry, these disciplines are physical chemistry and inorganic chemistry, and in physics, the disciplines that avoid the present progressive are classical mechanics, relativity and condensed matter physics.

Finally, the present progressive is also used to indicate and action/event that will take place in a near future, thus performing a similar function as that given by the semi modal *be going to*. However, the application of this tense-aspect marker with a sense of futureness is only found in the news corpus, illustrated in example 65.

[65] *In the RNCM on Friday they **<u>are playing</u>** Sir Peter Maxwell Davies's Trumpet Concerto, with Hakan Hardenberger as soloist*. [News- BNC]

4.4.7. The past progressive

The past progressive is the least used tense-aspect marker in the three corpora. In chemistry and physics, it totals 0.02 occurrences per one thousand words, while in news the distribution reaches 0.16, as seen in Fig 4.18.

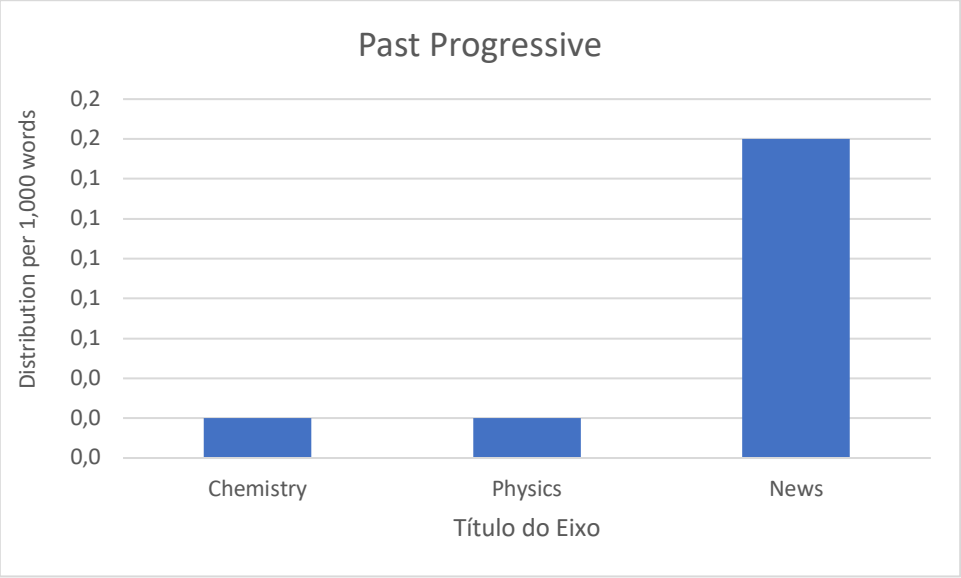


Figure 4.18: Distribution of the past progressive per 1,000 words in TB-Chem, TB-Phy, and News

In chemistry and physics, the past progressive - though little used - commonly appears in adverbial conditional clauses, which means the message refers to a hypothetical situation, as seen in examples 66 and 67.

[66] *If the atom **were radiating** energy constantly, then conservation of energy would require the orbit to decay* (...) [CHEM 20]

[67] *If the particle **were describing** a circle then r is a constant* (...) [PHY 01]

News, on the other hand, typically employs the past progressive to narrate ongoing events that took place in the past (example 68), also frequently using indirect speech (example 69).

[68] *The woman who **was holding** the baby is white, in her 30s, with brown, tidy collar length hair*. [NEWS - BNC]

[69] *A spokesman for President Clinton said assessment teams **were studying** the storm damage*. [NEWS - BNC]

4.4.8. Implications for teachers and materials developers

When we come to think of the teaching of verb tense and aspect in the context of ESP, it is best that we take into account the specific functions each verb tense/aspect display, rather than informing students of all the possible uses they may present. For instance, it is known that the present simple can be used to express past and future events, as illustrated in examples 1 and 2 respectively, from Huddleston and Pullum (2002, p.131).

1) *I **hear** we're getting some new neighbours*. [past meaning]
2) *The next high tide **is** around 4 this afternoon*. [future meaning]

Although the present simple can be used to indicate past and future events, this use is limited to specific registers (e.g. conversation, news). Thus, in our case, it would make more sense to expose learners to the actual uses of the present simple in the context of chemistry and physics academic prose, especially textbooks. For example, we could explore the present simple being used to explain/describe scientific

experiments/calculations. After explaining the transformations undergone by the present simple when it is used in interrogative sentences, i.e. the insertion of the auxiliary verb (do/does), students could work in pairs to practice transforming statements into questions, as in the example activity below. Sentences were taken from a physics textbook by Serway & Jewett (2004), which is included in our corpus.

Example activity 3:

1) Transform the statements below into questions:

a. The acceleration of the ball increases.
_____.

b. The position of the stone depends on the location chosen.
_____.

c. The stone's velocity depends on the choice of origin.
_____

2) Work in pairs. Complete the sentences with *do* or *does* and discuss the answers to the questions.

a. A student at the top of a building of height *h* throws one ball upward with a speed of v_1 and then throws a second ball downward with the same initial speed, v_1 . How _____ the final velocities of the balls compare when they reach the ground?

b. A ball is thrown upward. While the ball is in free fall, _____ its acceleration (a) increase (b) decrease (c) increase and then decrease (d) decrease and then increase (e) remain constant?

c. A stone is thrown vertically upward from the roof of a building. _____ the position of the stone depend on the location chosen for the origin of the coordinate system? _____ the stone's velocity depend on the choice of origin? Explain your answers.

and should not pose much difficulty to the teacher, since the answers to the questions are given in the book where the sentences were taken. Moreover, students play an

active role in the context of the ESP class and therefore they usually bring to class the background knowledge needed for the comprehension of scientific concepts.

It could also be interesting that during the presentation of the present simple tense in this activity, the subject of *free falling objects* be used to introduce the discussion of this verb tense. Thus, students would be able to discuss and revise this physical concept, besides dealing with verbs such as *throw*, *increase*, *decrease*, etc. besides their morphological variation in third person singular *throws*, *increases*, *decreases, etc.* in order to build up active vocabulary before the actual practice.

The morphological changes in the verb are even more evident in the past simple. In order to approach these changes, students can be exposed to the differences in morphology between regular and irregular verbs. In this sense, in addition to recycling vocabulary, students would actively identify what changes the verb phrase undergoes when used in the past simple. After being taught the rules that govern the past simple, students could do an exercise in which they are requested to transform verbs that are in imperative sentences (which are morphologically identical to the present simple) into the past simple, as shown below in example activity 4. The text used was taken from Bishop (2006), which is included in our corpus.

Example activity 4

1) Read an explanation on how to name binary covalent compounds:

1) If the subscript for the first element is greater than one, **indicate** the identity of the subscript using prefixes from the table below. We do not write mono- at the beginning of a compound's name.
Example: *We start the name for N2O3 with di-*.

Number of atoms	Prefix	Number of atoms	Prefix
1	mon(o)	6	hex(a)
2	di	7	hept(a)
3	tri	8	oct(a)
4	tetr(a)	9	non(a)
5	pent(a)	10	dec(a)

2) <u>Attach</u> the selected prefix to the name of the first element in the formula. If no prefix is to be used, **<u>begin</u>** with the name of the first element.
Example: *We indicate the N2 portion of N2O3 with dinitrogen.*

3) <u>Select</u> a prefix to identify the subscript for the second element (even if its subscript is understood to be one). *Leave* the *a* off the end of the prefixes that end in *a* and the *o* off of mono- if they are placed in front of an element whose name begins with a vowel (oxygen or iodine).
Example: *The name of N2O3 grows to dinitrogen tri-.*

4) <u>Write</u> the root of the name of the second element in the formula as shown in the table below.
Example: *The name of N2O3 becomes dinitrogen triox-.*

Element	Root	Element	Root	Element	Root	Element	Root
C	carb	N	nitr	O	ox	F	fluor
		P	phosph	S	sulf	Cl	chlor
				Se	selen	Br	brom
						I	iod

*<u>5) **Add**</u>* -ide to the end of the name.
Example: *The name of N2O3 is dinitrogen trioxide.*

2) Now, give the names of *NO2* and *N2O5* and write the steps you followed to do it. Use the underlined verbs in the text to help you. The first sentence has been done for you.

<u>N2O5: *We indicated the identity of the subscript (di).*</u>

<u>NO2</u>

<u>____:</u>

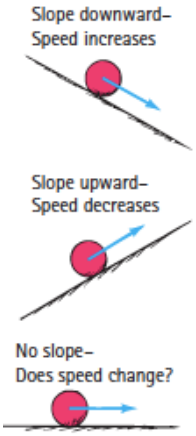
As stated in section 5.4.2, the past simple is also commonly used in chemistry and physics textbooks to introduce a historical background to the content being discussed. In this sense, it could be interesting to expose learners to this function of the past simple through texts that portray this characteristic. For instance, students could be invited to discuss about concepts brought by early scholars when they attempted to prove a certain phenomenon, as in the example activity 5 below. The texts were taken from Hewitt (2002), which is included in our corpus.

Example activity 5

1) The pictures below illustrate two of Galileo's experiments in the early 17th century. Which experiments do the illustrations refer to?

A)

B)



2) Decide if the phrases/sentences below used to describe Galileo's experiments refer to picture A or picture B above

	Picture
balls rolling on upward-sloping planes lost speed.	( )
did not fall twice as fast	( )
dropped objects of various weights	( )
fell together and hit the ground at the same time	( )
motions of various objects on plane surfaces	( )
objects hit the ground together	( )
the less the friction, the more the motion approached constant speed	( )

3) Read the text below about two of Galileo`s experiments. Choose from sentences A-G the one that fits each gap.

Galileo's Experiments

Leaning Tower

It was Galileo, the foremost scientist of the early 17th century, who gave credence to the Copernican view of a moving Earth. He accomplished this by discrediting the Aristotelian ideas about motion. Although [1] in Aristotle's views, Galileo was the first to provide conclusive refutation through observation and experiment.

Galileo easily demolished Aristotle's falling-body hypothesis. Galileo is said to have dropped objects [2] to compare their falls. Contrary to Aristotle's assertion, Galileo found that a stone twice as heavy as another did not fall twice as fast. Except for the small effect of air resistance, he found that objects of various weights, when released at the same time, [. 3] On one occasion, Galileo allegedly attracted a large crowd to witness the dropping of two objects of different weight from the top of the tower. Legend has it that many observers of this demonstration who saw the objects hit the ground together scoffed at the young Galileo and continued to hold fast to their Aristotelian teachings.

Inclined Planes

Galileo was concerned with how things move rather than why they move. He showed that experiment rather than logic is the best test of knowledge. Aristotle was an astute observer of nature, and he dealt with problems around him rather than [4]. Motion always involved a resistive medium such as air or water. He believed a vacuum to be impossible and therefore did not give serious consideration to motion in the absence of an interacting medium. That's why it was basic to Aristotle that an object requires a push or pull to keep it moving. And it was this basic principle that Galileo rejected when he stated that, if there is no interference with a moving object, [5]; no push, pull, or force of any kind is necessary.
Galileo tested this hypothesis by experimenting with the motions of various objects on plane surfaces tilted at various angles. He noted that balls rolling on downward-sloping planes picked up speed, while balls rolling on upward-sloping planes lost speed. From this he reasoned that balls rolling along a horizontal plane would neither speed up nor slow down. The ball would finally come to rest not because of its "nature," [6.]. This idea was supported by Galileo's observation of motion along smoother surfaces: When there was less friction, the motion of objects persisted for a longer time; the less the friction, the more the motion approached constant speed. He reasoned that, [7], a horizontally moving object would continue moving indefinitely.

A. but because of friction
B. fell together and hit the ground at the same time.
C. he was not the first to point out difficulties
D. in the absence of friction or other opposing forces
E. it will keep moving in a straight line forever
F. of various weights from the top of the Leaning Tower of Pisa
G. with abstract cases that did not occur in his environment

4.5. Phrasal grammatical complexity

In this section we analyze two of the most prominent phrasal grammatical complexity indices in academic prose (Biber & Gray, 2010) i.e. attributive adjectives and premodifying nouns in TB-Chem, TB-Phy and News. We address issues concerning their distribution as well as function patterns found in our corpora. At the end of the section, some sample activities to be worked in the context of English for Reading lessons are presented.

4.5.1. Attributive adjectives as nominal pre-modifiers

In this section we look at the distribution of attributive adjectives in NPs in our three corpora. Hence, we are looking at how *adjective + noun* combinations are spread in TB-Chem, TB-Phy, and news, as well as what types of adjectives mainly modify nouns in those texts. Examples 70 and 71 below illustrate this type of phrasal grammatical complexity feature, showing adjectives in bold and nouns underlined.

[70] *The geometry of these **hybrid** orbitals helps us to account for the **actual** structures and bond angles observed in **organic** compounds*. [CHEM 17]

[71] *A group representing families of the **British** victims marked the anniversary by handing in a petition to 10 Downing Street calling for a **public** inquiry*. [News]

A first glance at the distribution of this feature shows that chemistry and physics textbooks display fairly similar counts, (62.25) and (61.21) times per 1,000 words respectively, while news presents a lower total of 50.67 times per 1,000 words, as seen in Fig. 4.19. This means that that attributive adjectives modifying nouns is around 22% more frequent in chemistry and physics than it is in news.

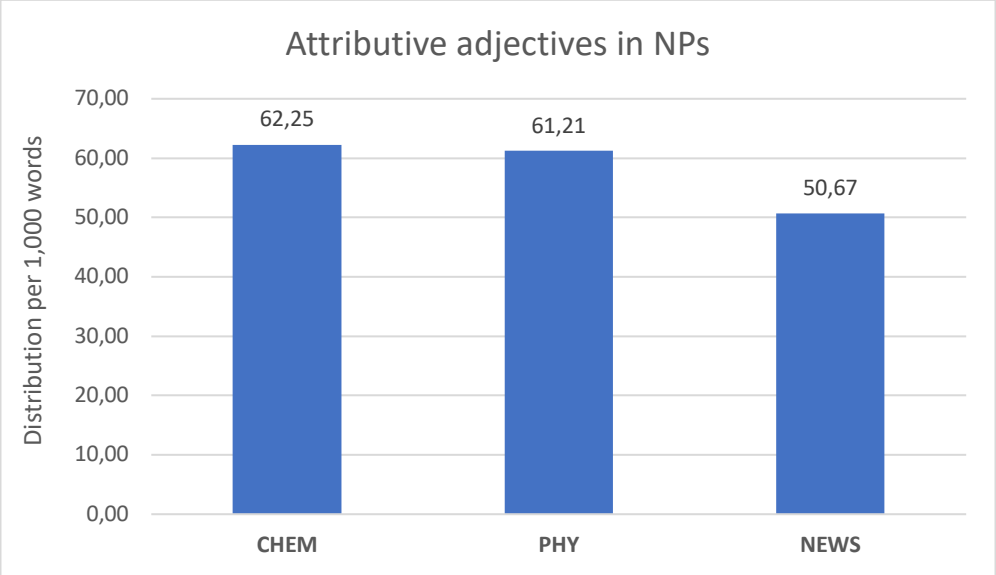


Figure 4.19: Distribution of attributive adjectives modifying nouns in TB-Chem, TB-Phy, and News.

These figures, however, present some interesting differences when looked at from a semantic point of view. Semantically, adjectives can be grouped into two different categories: descriptors and classifiers. As discussed in section 2.7.2, Biber et al. (1999, p.508) state that descriptors are "prototypical adjectives denoting such features as color, size and weight, chronology and age, emotion (...) and other characteristics", while classifiers are used "to delimit or restrict a noun's referent, by placing it in a category in relation to other referents". These two broad categories can be further divided into finer classes, as done in Biber & Gray (2016), namely:

- Descriptive (Physical): color or physical appearance, size or weight, age and frequency.
- Descriptive (Evaluative): evaluations or emotional states .
- Classifying (Relational/ Classificational): delimit the referent of a noun in relation to other referentes.
- Classifying (Topical/Affiliative): identify the type or subject area of a noun, or designate the group that a referent belongs to.

Examples of adjectives belonging to each of these subcategories and found in our corpora are shown in Table 4.3 below.

Table 4.3: Examples of adjectives taken from TB-CHEM, TB-PHY, and NEWS according to semantic categories.

	Descriptive (Physical)	Descriptive (Evaluative)	Classifying (Relational/ Classificational)	Classifying (Topical/Affiliative)
TB-CHEM	*purple colors*	special cases	residual blood	hydrochloric acid
TB-PHY	solid body	useful application	valid answer	optical axis
NEWS	old airfield	hard decisions	public company	Australian beetle

The distribution of attributive adjectives as nominal pre-modifiers according to semantic categories reveals a different picture from the evenly-distributed figures shown in Fig. 4.19. From a random sample of 200 adjectives in each corpus, what we see is that chemistry and physics show a preference for classifiers, whereas in news the distribution of adjectives is more uniform, as we can see in Fig 4.20.

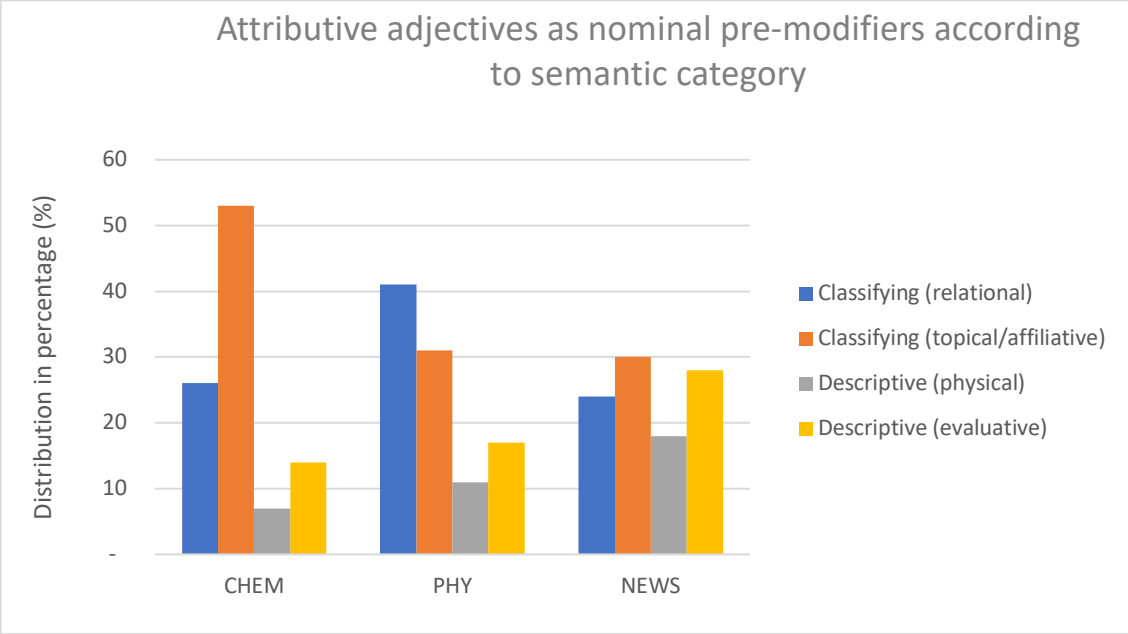


Figure 4.20: Distribution of attributive adjectives modifying nouns in TB-Chem, TB-Phy, and News according to semantic category

Chemistry relies heavily on the use of classifiers, especially those of the topical type, which accounts for more than 50% of the occurrences, which is similar to what Dutra et.

al (in press) found in their corpus of chemistry research articles, even though they looked at longer sequences of adj/noun modification. This preference for classifiers in textbooks happens mainly because chemistry uses this type of adjective to differentiate among the types of nouns repeatedly used in this discipline. For instance, *acid* and *bond* are frequent nouns used throughout chemistry, but it is through the addition of a classifying adjective to those nouns that different disciplines in chemistry distinguish among the many types of acids and bonds that exist, such as *sulfuric acid*, *hydrochloric acid*, *carbonic acid*, *nitric acid*, etc. and *metallic bonds*, *disulfide bonds*, *ionic bonds*, *covalent bonds*, etc. Examples 72 and 73 below bring examples taken from TB-Chem to illustrate the modification of other common nouns (reactions, compounds, chemistry) with the addition of an attributive adjective.

[72] (...) *are also characteristic of some **biochemical** reactions, and they can behave in ways that are beneficial or detrimental to the organism*. [CHEM 05]

[73] *Alkyl groups may also act as bridges in **inorganic** compounds, a function rarely encountered in **organic** chemistry*. [CHEM 16]

Classifiers are also the preferred adjective category in physics, accounting for more than 70% of the occurrences. However, physics shows a preference for classifying topical/affiliative (41%) followed by classifying relational adjectives (31%). This latter category performs a similar function in physics as it does in chemistry, i.e. it is used to differentiate frequently used nouns, such as *field*, *motion*, and *acceleration,* as shown in examples 74 and 75.

[74] *Electrodynamics is the interaction of matter via the **electromagnetic** field*. [PHY 02]

[75] (...) *and after some cooling has taken place, the **gravitational** field of the matter of the universe delivers about 1 MeV units of energy to* (...). [Phy 09]

The classifying relational category in physics, on the other hand, will be used in physics to modify nouns used in the description and/or explanations of physical concepts, experiments, and certain natural phenomena, as shown in examples 76, 77, and 78.

[76] *Air vibrations can be understood in terms of either **vertical** air displacement or changes in air pressure.* [PHY 16]

[77] *Quantities which take **different** values in **different** frames of reference, but which combine to form a law which is applicable in all frames* (...). [PHY 01]

[78] *One technique for measuring the **specific** heat of a solid or liquid is to raise the temperature of the substance to some value* (...) [PHY 11]

Interestingly, some disciplines tend to favor the use of attributive adjectives more than others. In chemistry, the use of attributive adjectives as nominal pre-modifiers in especially common in analytical chemistry, while in physics this tendency is seen is solid state physics, but little noticed in relativity.

In contrast, the distribution of attributive adjectives as nominal pre-modifiers in News shows a different pattern. News texts employ the four semantic categories of adjectives more homogeneously. Although the classifying topical/affiliative is the most frequent - showing a similarity with chemistry - there is a striking difference in its use between these two registers. While chemistry uses this type of adjective mostly with a topical function i.e. to identify the subject area of a noun, news uses it mostly with an affiliative function, i.e. to express the national or religious group to which a referent belongs, as we can see from examples 79 and 80:

[79] *The NAACP denounces The Silence of the Lambs because Hannibal the Cannibal shares his name with an **African** general* (...) [News - BNC]

[80] *This is Walton's part in the Women's World Day Of Prayer. Women from **Christian** congregations throughout the town will be attending.* [News - BNC]

Concerning the use of descriptors, news employs it to a much greater extent than do chemistry and physics. As regards descriptive (physical) adjectives, we found that news

employed it 100% more times than did chemistry and physics textbooks; similarly, descriptive (evaluative) adjectives are also 75% more frequent in news. Another distinctive factor concerning news is the fact that, according to Biber et al. (1999, p. 514) it "employs a wider range of common adjectives than any other register". The authors also add that, in news "many of the most common attributive adjectives [...] are words derived from, or closely related to, nouns, especially those ending in -al, such as political and national".

Below we present a table with the most frequent attributive adjectives modifying nouns as well as the most frequent nouns modified by attributive adjectives. Remarkably, some similarity is noticed between chemistry and physics, but that cannot at all be seen when we compared them to news.

Table 4.4: Common attributive adjectives modifying nouns and common nouns modified by attributive adjectives in TB-Chem, TB-Phy, and News.

	Common attributive adjectives modifying nouns	Common nouns modified by attributive adjectives
Chemistry	*atomic, large, negative, organic potential, single*	*charge, electron, energy, field, force, line, motion, value*
Physics	*basic, electric, high, magnetic, mechanical, single, various*	*acid, atom, bond, charge, compound, energy, force, molecule, reaction*
News	*bad, big, further, full, general, great, high, local, new, old, other, original, potential, public, real, social, special, top, wrong*	court, election, idea, investment, life, market, measures, Party, problems, security, site

4.5.2. Nouns as nominal pre-modifiers

In this section we look at the distribution of nouns functioning as modifiers of other nouns, thus forming noun-noun sequences. Given that these sequences do not contain a function word to make their meaning relationship explicit, we also investigate the

different meaning relations expressed by these noun-noun sequences by categorizing them into subclasses according to the logical relationship that exists between the modifying noun and head noun. Examples 81, 82, and 83 below present some noun-noun sequences for illustration.

[81] *A proton from the **halogen acid** must protonate the **oxygen atom**, providing an alcohol as* (...). [CHEM 14]

[82] *High-temperature plasmas created in **fusion reactors** may someday allow humankind to harness the **energy source** of the sun*. [PHY 11]

[83] *He said that **market forces** were being 'inhibited by **state ownership** and **government support** for producers whose viability was doubtful.'* (NEWS- BNC)

As regards their distribution, chemistry presents the greatest number of noun-noun sequences (31.4), followed by news (28.2) and physics (25.4), as shown in Fig 4.21.

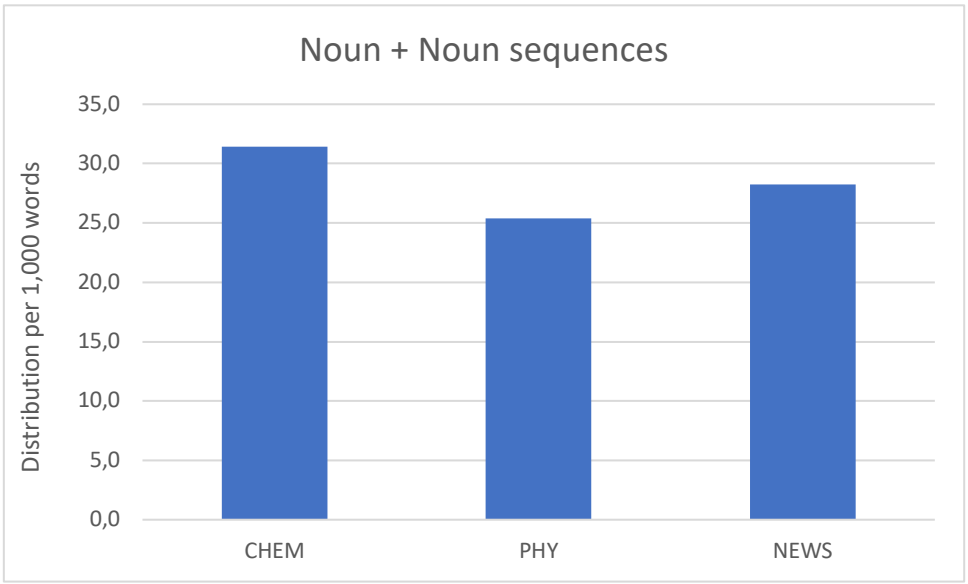


Figure 4.21: Distribution of nouns premodified by nouns in TB-Chem, TB-Phy, and News

In comparison to physics, not only does chemistry present more instances of noun-noun combinations, but it also possesses a larger repertory of possible candidate words to fill in the slots for N1 (premodifying noun) and N2 (head noun). As shown in Table 4.5 below, in a hypothetical noun 1 + noun 2 (N1 + N2) sequence, physics will have, in raw counts, around 1500 candidate words to fill in N1 and 2400 words to fill in N2, whilst these numbers for chemistry are around 1900 and 2900[13], respectively. This shows that although nouns can both modify other nouns (when in N1 position) or be modified by other nouns (when in N2 position), the number of possible modifiers is smaller than that of the modified (head) nouns. News, on the other hand, presents a much greater number of possible candidates for both N1 and N2 than chemistry and physics.

Table 4.5: Approximate number of candidate words to fill in N1 and N2 slots in a noun + noun sequence. Data from TB-CHEM and TB-PHY

	Approximate number of candidate words for N1	Approximate number of candidate words for N2
Chemistry	1900	2900
Physics	1500	2400
News	3300	3900

Although the range of possibilities concerning potential candidates to fill in the N1 and N2 slots is strikingly large, the probability that these slots will be occupied by any random noun is rather small. In chemistry, the N1 position is occupied, one third of the times, by only 1.5% of all possible nouns. This means that out of the 1900 possible nouns, only 29 are appear more frequently in N1 position, namely: *hydrogen, carbon, electron, energy, water, bond, gas, metal, sodium, quantum, wave, oxygen, pH, phase,*

[13] Counts for possible nouns filling in the N2 slots distinguish between singular and plural forms; here, electron and electrons, for example, are treated as two different words.

amino, surface, ion, p, valence, reaction, carbonyl, Alkyl, transition, symmetry, resonance, dipole, methyl, end, and *equivalence*, as shown in Fig. 4.22.

Similarly, in physics, despite the roughly 1500 possibilities, the N1 position is occupied, one third of the times, by 20 nouns (representing 1.3% of all possible nouns), namely: *wave, quantum, energy, sound, time, electron, field, vector, air, phase, lattice, unit, reference, point, particle, incident, surface, crystal, heat,* and *space,* as shown in Fig 4.23. Notably, in both chemistry and physics, most frequent nouns in N1 position tend to be of the non-countable type (e.g., energy, sound), whereas those in N2 position tend to be countable nouns (atom, theory).

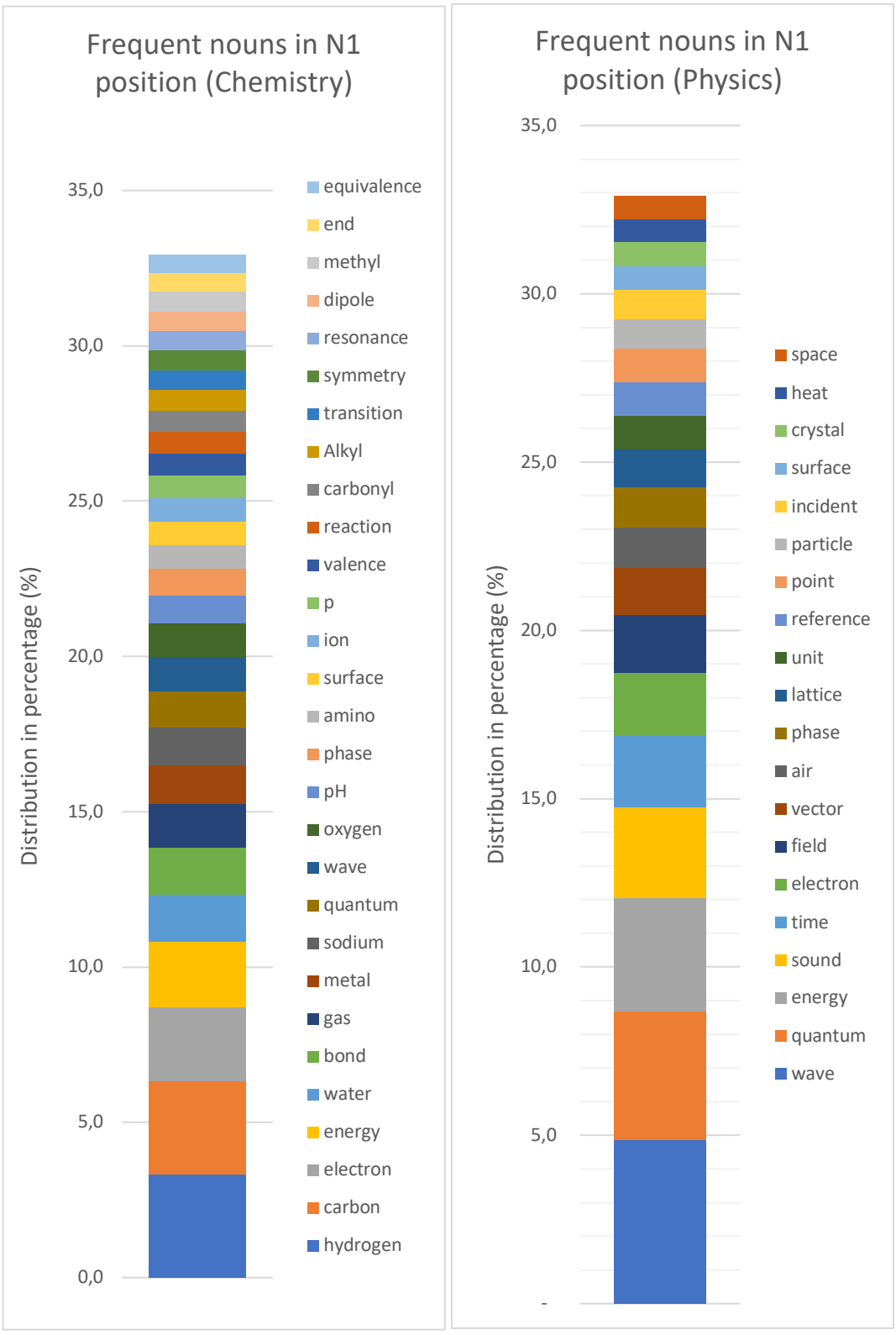


Figure 4.22 and Figure 4.23: Frequent nouns in N1 position in TB-CHEM and TB-PHY, respectively.

The tendency of presenting few frequent nouns occupying a certain slot is also noticed when we look at nouns in N2 position. Although the range of possible N2 nouns is even

wider than that of N1 nouns, as seen in Table 4.5, 25 nouns in N2 position in chemistry account for nearly one third of the occurrences, namely: *atom(s), ion(s), group(s), molecule(s), point(s), state(s), bond(s), energy(ies), orbital(s), level(s), reaction(s), change(s), function(s), temperature(s), dioxide, solution(s), curve(s), chloride, pressure(s), pair(s), density, electrode(s), moment(s), gas(es), particle(s), area(s).* In physics, 22 nouns in N2 position account for approximately 20% of occurrences, namely: *wave(s), function(s), theory(ies), vector(s), mechanics, state(s), system(s), point(s), level(s), force(s), energ(ies), velocity(ies), field(s), time, difference(s), pressure, structure(s), interval(s), frame(s), density, lattice, temperature.*

News also presents a considerable number of noun-noun sequences. However, as presented by Biber et al. (1999), although news has by far the greatest number of premodifying nouns that frequently combine with many head nouns, in our corpus only six premodifying nouns are, at the same time, very productive and extremely frequent: *world*, *police*, *home*, *health*, *city*, and *government*. The authors also argue that this could be due to space-saving in texts, as "the extremely productive use of nouns as premodifiers in news results in a very dense, integrated packaging of information" (p.594).

Another point of analysis concerning these noun sequences has to do with the meaning relations established between them. As Biber & Gray (2010, p.12) point out, this type of construction results in a "loss of explicitness", as there is "no grammatical clue to help the reader know what the meaning relation is between the two nouns". Let us take, for example, the same noun (electron) as a pre-modifier in three sentences:

[84] *The energy of the **electron beam** can be decreased from the typical 70 electron volts (eV) to 20–25 eV, where much less fragmentation occurs.* [CHEM 17]

[85] *The first moment < x > of the **electron distribution** along some coordinate x is (...)* [PHY 14]

[86] *Going to the limit of space-time points leads to the well known infinities of Quantum Field Theory (and classical* ***electron theory****) which require* (...). [PHY 17]

In the examples above, although we have the same pre-modifier (*electron*), its meaning in relation to the head nouns is rather different. In 84 (*electron beam),* N2 is made from/consists of N1; in 85 (*electron distribution*), N1 is the object of the process described in N2; and in 86 (electron theory), N2 is about N1. Because in each of these examples the meaning relation is implicit, the reader - especially the L2 reader - might have some difficulty to interpret it. Here we have seen three possible meaning relations to explain how one noun modifies the other, but Biber et al. (1999, 2002) propose 12 semantic categories to classify the different possible semantic relations. Still, this classification is not exhaustive and, according to these authors, there is a number of noun + noun sequences that do not fit neatly into any of these categories, as we can see in Table 4.6 below.

Type	Meaning	Chemistry	Physics	News
1 composition	N2 is made from N1; N2 consists of N1: e.g. glass windows = windows made from glass	*amino acid, ethyl acetate, isopropyl alcohol, metal alloy, multielement analysis, hydrogen atom, hydrogen bond, blood cell, carbon chain, sodium chloride, nitrogen compound, carbon dioxide, heat energy, radiation field, neon gas, carboxyl group, alkyl halides, sodium ion, water molecule, calcium nitrate, aluminum oxide, sugar solution, salt water*	*hydrogen atom, laser beam, steel cable, air column, wave components, quartz crystals, laser light, water molecules, crystal momentum, sound wave*	*gold medal*
2 purpose	N2 is for the purpose of N1; N2 is used for N1: e.g. pencil case = case used for pencils; war fund; nursery programs	*X*	*billiard ball, probability calculus, test charge, proportionality coefficient, vector diagram, force field, reference frame,*	*savings account, unemployment benefit, emergency calls, computer application, school fees, life insurance, health services, winter home*
3 identity	N2 has the same referent as N1 but classifies it in terms of different attributes e.g. women algebraists = women who are algebraists	*X*	*X*	*police officers*

4 content	N2 is about N1; N2 deals with N1: e.g. algebra text = a text about algebra	*transduction approaches, force constant, gas chromatography, phase diagram, heat capacity curve, electron configuration*	*quantum number, quantum system, quantum theory, displacement vector*	*education bill, investment board, history books, music business, election campaign, heath care, law commission, press conference, computer games, abortion law, investment levels, job losses, interest rates, radio reports, living standards, peace talks, asset values, health warnings,*
5 objective	N1 is the object of the process described in N2, or of the action performed by the agent described in N2; N2 is the object of the process described in N1 e.g. egg production = X produces eggs; discharge water = water that has been discharged	*energy absorption, enzyme activity, temperature adjustment, ion analyzer, ionization chamber, ion detector, energy distribution, ionization energy, dispersion forces, bond formation, detection limit, air oxidation, transfer reactions,*	*shock absorber, energy conservation, plastic deformation, x-ray diffraction, electron oscillators, energy transformation*	*project management*
6 subjective	N1 is the subject of the process described in N2; N2 is usually a nominalization of an intransitive verb; N2 is the subject of the process described in N1. e.g. child development = children develop; labor force = a force that is engaged in labor	*rotation axis, entropy change, color transition,*	*transmission axis, velocity change, air flow, sound signals,*	*investment group*

7 time	N2 is found or takes place at the time given by N1: e.g. summer conditions = conditions that occur during the summertime	*X*	*X*	*Christmas eve, evening flight, weekend games*
8 location	N2 is found or takes place at the location given by N1. N1 is found at the location given by N2, e.g. corner cupboard = a cupboard that is located in the corner; notice board = a board where notices are found	*surface area*	*zone boundaries, spring constant, space coordinates,*	*race area, school building, heart attack, home office, dressing room, boarding school, police station, trade zone*
9 institution	N2 identifies an institution for N1: e.g. insurance companies = companies for (selling) insurance	*X*	*X*	*automobile association, district council, finance department, city hall, software house, stock market, welfare organizations,*
10 partitive	N2 identifies parts of N1: e.g. cat legs = legs of a cat	*bond angle, electron charge, , electron density, valence orbitals,*	*arc length,*	*X*
11 specialization	N1 identifies an area of specialization for the person or occupation given in N2; N2 is animate:	*X*	*X*	*travel agent, health authority, insurance broker, executive committee, defense minister, food retailers, defense secretary, police spokesman,*

	e.g. finance director = director who specializes in finance			
12 other	Sequences that do not fit neatly into any of the above categories	*heat capacity, valence electron, energy level, molar masses, quantum mechanics, dipole moments, symmetry operation, electron pairs, absorption peaks, end point, gas phase, diffraction pattern, equivalence point, gas properties, reaction rates, redox reactions, valence shell, sample solution, transition state, membrane structure, gas systems, room temperature,*	*free-fall acceleration, phase accumulation, oscillation amplitude, deflection angle, electron approximation, unit area, coordinate axes, exponential behavior, material body, heat capacity, unity cell, electron concentration, boundary conditions, resonance curves, electron density, frequency dependence, time derivatives, phase difference, transmission directions, pressure distribution, entropy effects, valence electrons, reaction force, wave front, energy gap, energy interactions, time interval, energy levels, photon mass, wave packets, wave properties, temperature scale, energy source, crystal structure, room temperature, frequency variables, sound velocity*	*diesel car, world championships,*

Table 4.6: Frequent Noun + Noun sequences found in TB-Chem, TB-Phy, and News, classified according to semantic categories defined by Biber et al. (1999, 2002).

As we can see from Table 4.6, for chemistry, out of the eleven categories proposed in Biber et al. (1999, 2002), two were especially productive, namely: category 1 - composition, and category 5 - objective. Interestingly, category 1 deals precisely with things made from/ consisting of another (e.g. *hydrogen atom*, *sodium ion*), and relate to noun sequences whose head noun functions as a sort of 'hypernym'. This notion of hypernymy lies in the fact that there are a number of frequently used nouns that if treated in isolation would function semantically as a broad category and thus require the combination with other nouns to specify its meaning. For instance, the noun *ion* alone would represent a broad category (a hypernym), but in combination with other words, more specific meanings are conveyed, thus forming hyponyms such as *chloride ion*, *hydroxide ion*, and *hydrogen ion*. As for category 6, we see it relates to processes that are realized by nominalizations in discourse, a common feature of academic prose, as discussed in Biber & Gray (2013, 2016). In addition, many noun sequences would not fit in any of the categories and were grouped under category 12 - other. Five categories were not filled in, namely, 2 - purpose, 3 - identity, 7 - time, 9 - institution, and 11 - specialization.

Semantic categories found in physics are similar to those in chemistry. The main difference between the two is found in category 2 - purpose, which is more common in physics than in chemistry; in addition, more noun-noun combinations in physics fall in category 12 - other, which shows that such a categorization is not sufficient to incorporate most possible semantic realizations of these noun-noun combinations. Similarly, expect for category 2, all other categories not common in chemistry are also not common in physics, namely: 3 - identity, 7 - time, 9 - institution, and 11 - specialization.

Interestingly, noun-noun sequences found in news fit in the above semantic categories much more neatly than they do in chemistry and physics. This may be explained by the fact that nouns employed in news are less technical and thus more easily categorized. As we can see from Table 4.6, noun sequences in news fit in all of the 11 semantic categories proposed, which reinforces the claim that such a semantic categorization is more appropriate for non-technical registers.

Another difference regarding premodifying nouns can be seen when we look at three-word noun sequences (noun + noun + noun). As shown in Fig 4.24, these sequences are more frequently found in news (3.0 occurrences), followed by chemistry (2.2) and physics (1.3) per one thousand words.

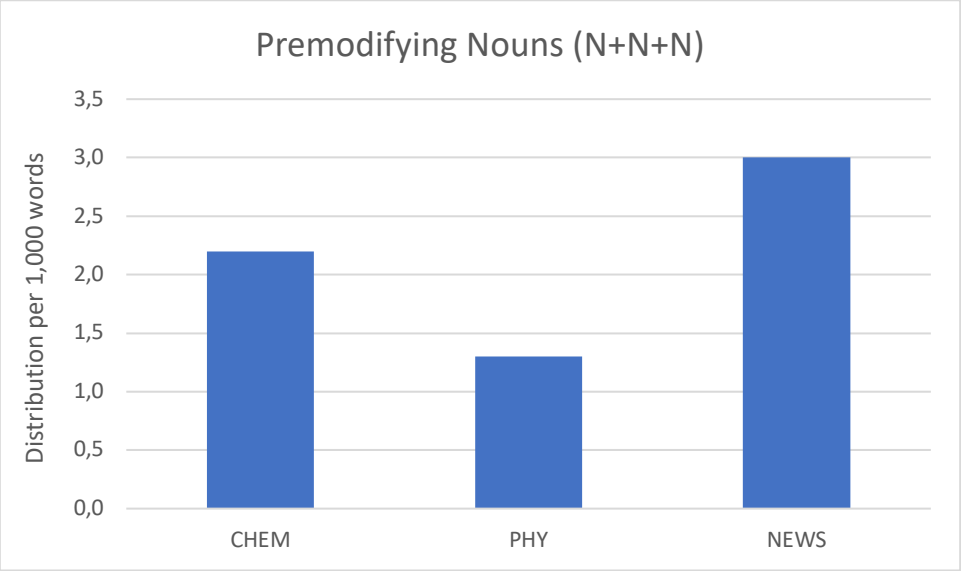


Figure 4.24: Distribution of three-word noun sequences in TB-Chem, TB-Phy, and News per 1,000 words.

The fact that news presents more instances of this type of sequence corroborates the assertion found in Biber et al. (1999, p.597) when they state that ”in news, longer premodifier sequences are slightly more common [than other registers]” (emphasis added), which highlights the efficiency of this type of pre-modification insofar as it allows for the “packing (of) dense informational content into as few words as possible”. On the other hand, when we compare chemistry and physics, we notice a greater difference in the distribution of these three-word noun sequences. This happens mainly because chemistry presents a large number of compound nouns (formed by noun + noun), such as *amino acid, silver nitrate,* and *hydrogen peroxide*. Noun compounding in the form just described is very productive in chemistry discourse because it is used to designate chemical elements/compounds, which are words that appear rather frequently in chemistry textbooks in the form of N + N. Examples 87 and 88 below illustrate this assertion.

[87] *Myoglobin is an example of a small globular protein, consisting of a single* ***amino acid*** *chain*. [CHEM 01]

[88] *Any ionic compound that has the same arrangement of cations and anions as NaCl is said to have the* ***sodium chloride*** *crystal structure.* [CHEM 04]

When three nouns occur in sequence, they could display two different pre-modification patterns: 1) the two first nouns modify the head noun ([n+n] + n), and 2) the first noun modifies a compound head noun (n + [n+n]). Example 89 below exemplifies the first case and example 90 exemplifies the second. The modifier is in boldface.

[89] *Interestingly, the calcium carbonate reacts with the acidic soil to form* ***carbon dioxide*** *gas, which in the atmosphere helps to keep rainwater slightly acidic*. [CHEM 03]

[90] *If a hard wall is shifted with respect to the reference plane x = 0 by a distance d (see Fig. 6.7a), a* ***plane*** *air cushion with the thickness d is formed* (...). [PHY 03]

In a random sample of 200 concordance lines, physics and news show similar counts when we look at the distribution of three-word noun sequences with regard to the pre-modification patterns just described; both patterns are evenly distributed in these registers (≃ 50% for each pattern). Chemistry, on the other hand, shows a preference for pattern 2, which accounts for around 63% of the occurrences. This preference could be explained, as stated before, by the fact that in chemistry vocabulary we find a large number of compound nouns (n+n) that could not be used otherwise, as these compound nouns serve the purpose of naming elements that are present in chemistry discourse (e.g. carbon dioxide, sodium nitrate).

4.5.3. Implications for teachers and materials developers

The combination of adjective + noun forms a rather simple structure in English, and there is normally no variation in this structure. As stated by Eastwood (2002, p.197), “there is usually a fixed order of adjectives before a noun" , and according to Biber et al. (1999, p. 505), when adjectives have an attributive role they “typically precede the head

noun". In Portuguese, on the other hand, attributive adjectives display the opposite word order, i.e. they usually follow the head noun, as we see in the translated example below:

E.g.: *Even the wind (...) is a form of **solar power***. [PHY 10]
adj noun

*Mesmo o vento (...) é um tipo de **energia solar***.
noun adj

Although in Portuguese the most common word order is noun + adjective, there are times in which the word order is adjective + noun. This, happens "when an adjective is used not to differentiate or specify, but rather to mention an inherent quality of the noun" as stated by Whitlam (2011, p.38). Below we show two examples Whitlam brings to illustrate this assertion:

E.g. *na **longínqua China**.*
in **far-away China** (i.e. there is no 'nearby China')

O restaurante serve uma **deliciosa comida italiana**.
The restaurant serves **delicious Italian food**. (i.e. all the food is delicious)

Thus, while in Portuguese the word order concerning attributive adjectives modifying nouns is variable, in English this word order is more fixed, which could make it easier for students who are reading texts written in English as an L2. However, low-proficiency students very often find it difficult to interpret/translate adjective + noun sequences, especially when there is more than one adjective modifying a noun (example 91), or when there is one or more adjectives modifying two nouns (a noun + noun sequence) (example 92).

[91] *The [**homonuclear [diatomic molecules]]** discussed Section 5.2 are nonpolar molecules*. [CHEM 16].

[92] *Because mercury and alcohol have [**different [thermal [expansion properties]]]**, when one indicates a temperature of 50° C, say, the other may indicate a slightly different temperature*. [PHY 11]

As we have seen in the discussion of this section, among the classes of adjectives, classifiers are more frequent than descriptors in chemistry and physics. These classifiers, however, especially those of the topical type, tend to be easily translated by Brazilian learners - even low-proficiency ones - since classifying/topical adjectives used in chemistry and physics tend to be scientific words that have similar correspondents in Portuguese, such as sulfuric (*sulfúrico*) or electromagnectic (*electromagnético*). Hence, when preparing teaching materials that deal with attributive adjectives, we do not need to focus more on classifiers (due to frequency) than on descriptors; instead, we try to balance both classes of adjectives.

Below we bring a sample activity that deals with adjectives + noun sequences. The text used in the activity was adapted from Housecroft & Sharpe (2012), a textbook that is part of our corpus. In this activity students will first identify, among ten phrases, the part of speech of each word: noun or adjective (exercise A). By the end of this step, the teacher will make sure students have realized that adjectives (modifiers) precede the noun (the modified item).

Following that, students will read a text (exercise B) and fill in the gaps with words from exercise A. This type of activity will help them not only memorize the vocabulary just learned but also, as stated by Grellet (1995), help them find out the meaning of words based on context. This activity could be followed by, for example, text comprehension questions that would lead them better understand details of the text. Alternatively, students could create/complete a schema or diagram (either in English or Portuguese depending on the objective) about the industrial manufacturing process for the production of iron, such as the sample in exercise C.

Example activity 6

A) In the phrases below, write A for adjective and N for noun.

crude steel
enormous scale
exact nature
high Si content
highest temperature zone
industrial manufacturing processes
major raw materials
secondary elements
small amounts
wrought iron

B) Read the text below and fill in the blanks with words from exercise A.

Iron production

The _____ for the commercial production of Fe are haematite (Fe_2O_3), magnetite (Fe_3O_4) and siderite ($FeCO_3$). The extraction of iron is carried out on an _____ to meet the consumer demands for both iron and steel. In 2010, China and Japan led the world in the production of _____. The _____ for iron can be summarized as follows. Iron ore is mixed with limestone ($CaCO_3$) and coke in a blast furnace in which temperatures vary from ≈750 to 2250 K. Carbon is converted to CO in the _____, but both C and CO may reduce the iron ore:

$2C + O_2 \rightarrow 2CO$
$Fe_2O_3 + 3C \rightarrow 2Fe + 3CO$
$Fe_2O_3 + 3CO \rightarrow 2Fe + 3CO_2$

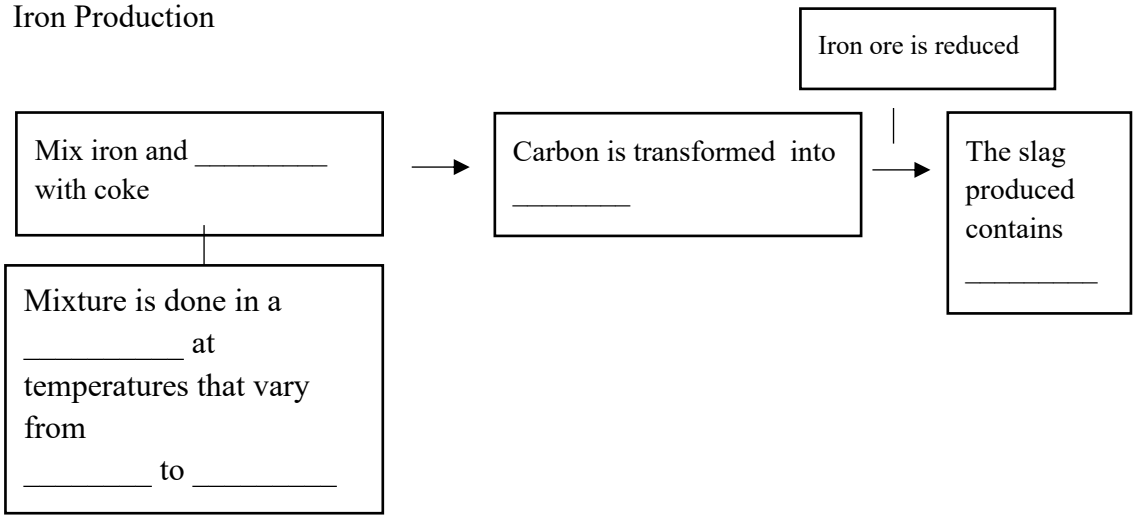
The function of the limestone is to remove impurities and the product of these reactions is slag, which contains, for example, calcium silicate. Molten Fe from the furnace is collected and cooled in salt-moulds as pig iron, which contains 2–4% C plus _____ of P, Si, S and Mn. After remelting and moulding, the product is cast iron; this is brittle and its _____depends upon the relative amounts of _____. A _____ results in the C being in the form of graphite, and the cast iron so formed is called grey cast iron. On the other hand, white cast iron forms

when the Si content is low and carbon is present within the iron–carbon phase cementite, Fe3C.

The puddling process is used to convert cast iron to _____. During this process, C, S and other impurities are oxidized, leaving wrought iron with <0.2% C content. Unlike cast iron, wrought iron is tough and malleable and is readily worked; its applications, in wrought iron railings and window and door grills, are widespread.

Adapdted from: Housecroft, C. E; Sharpe, A. G. (2012). *Inorganic chemistry*.

C) Complete the diagram with information from the text.



We now bring a sample activity aimed at teaching noun + noun combinations. When dealing with noun + noun sequences, it is important to be aware that such a construction is not as common in Portuguese as it is in English. In Portuguese when two nouns occur in sequence, they are most often separated by a hyphen, as shown in the following examples: *cirurgião-dentista* (dental surgeon), *tenente-coronel* (lieutenant colonel), *tia-avó* (great aunt) or *editor-chefe* (editor-in-chief). Hardly ever do we see two nouns in the same noun phrase not being hyphenated, as in the case of *efeito estufa* (greenhouse effect). What generally happens, however, is that a noun modifying another noun requires the addition of a preposition between them, as in: *medalha de ouro* (gold medal), *átomo de carbono* (carbon atom). Nevertheless, though the most frequent, this

is not the only possible way a noun + noun sequence can be translated. When translating a noun + noun sequence from English into Portuguese, we come across three main possibilities:

a) the inversion of the nouns with the addition of a preposition between them: e.g. carbon dioxide (*dióxido de carbono*); energy levels (*níveis de energia*); uncertainty principle (*princípio da incerteza*).

b) the inversion of the nouns without the addition of a preposition: e.g. carbonyl group (*grupo carbolina*); room temperature (temperatura ambiente); aluminum foil (*papel-alumínio*)

c) the inversion of the elements, being the first noun translated as an adjective: e.g quantum mechanics (*mecânica quântica*); salt water (*água salgada*)*; gas state* (*estado gasoso*).

In addition, when discussing noun + noun sequences in a language class with low-proficiency students, it is important to make sure they understand that not only is this type of construction rather common in English academic prose, but also that word order in English is an issue they have to aware of; that is, the modifying noun comes before the head noun in English, as opposed to what happens in Portuguese.

Below we bring a sample activity that aims to teach and practice noun + noun combinations. The text and the table shown in the activity were adapted from Bishop (2002), a book that is part of our corpus of chemistry textbooks.

As a pre-reding activity, the teacher starts a discussion on recycling, pre-teaching vocabulary such as landfill, recycle bin, and trash. During the discussion, the teacher includes questions regarding materials that can be recycled, attitudes and policies concerning recycling, how recycling is dealt with in the town where students live, etc. Following that, the teacher tells students they will read a text about 'recycling synthetic

polymers' and students are encouraged to say what they know about synthetic polymers.

The teacher, then, writes two questions on the board. Students are supposed to read the first paragraph of the text to find the answer to these questions:

- What happens to a used bottle if you toss it into the trash?
- What happens to a used bottle if you put it in the recycle bin?

After students give the answer to the questions asked, the teacher solves questions they might have concerning the vocabulary of the first paragraph.

Now, students will be asked to read the rest of the text. In order to encourage reading, a while-reading task will be proposed. Students will read the text and fill in the blanks with the correct word, paying attention to their part of speech. The words are: recycling, recycled, recycle, and polymers. After correcting the activity, the teacher solves any remaining vocabulary questions, and students will answer some text comprehension questions.

Example activity 7

A) Read the text and fill in the blanks with the correct word from the box.

recycling	*recycled (2x)*	*recycle*	*polymers*

Recycling Synthetic Polymers

You finish off the last of the milk. What are you going to do with the empty bottle? If you toss it into the trash, it will almost certainly go into a landfill, taking up space and serving no useful purpose. If you put it in the recycle bin, it's likely to be melted down to produce something new.

Between 50 and 60 billion pounds of synthetic polymers are manufactured each year in the United States—over 200 pounds per person. A large percentage of these _____ are tossed into landfills after use. This represents a serious waste of precious raw materials (the petroleum products from which synthetic polymers are made), and exacerbates concerns that the landfills are quickly filling up.

Some synthetic polymers can be _____ and some cannot. So-called thermoplastic polymers, usually composed of linear or only slightly branched molecules, can be heated and formed and then reheated and reformed. Therefore, they can be _____. On the other hand, thermosetting polymers, which consist of molecules with extensive three-dimensional cross-linking, decompose when heated, so they cannot be reheated and reformed. This makes them more difficult to _____.

In 1988, the Plastic Bottling Institute suggested a system in which numbers embossed on objects made of polymers tell the _____companies what type of polymer was used in the object's construction.

Adapted from: Bishop, M. (2002). *Introduction to chemistry*.

B) Answer the questions about the text 'Recycling synthetic polymers'.

1) What do the following numbers refer to?

- 50 and 60: _____
- 200: _____
- 1998: _____

After correcting the activity by asking different students to answer each question, the teacher will write on the board the following phrase extracted from the text*: raw materials = the <u>petroleum products</u> from which synthetic polymers are made.* The teacher should elicit from the students the meaning of what is written on the board and will draw their attention to the phrase *'petroleum products'.* The teacher then explains

that this is a noun + noun combination, and tells students it is a type of construction that is common in English, but that is realized differently in Portuguese (generally, in Portuguese, there will be a preposition between the two nouns, and the head noun comes first). The teacher gives some more examples of this type of construction, and asks students to translate them*: hydrogen atom, electron pair, carbon dioxide, energy levels*.

After that, students will be given a table also adapted from Bishop (2002) containing examples of uses for virgin polymer, as well as examples of uses for recycled polymer. Among these examples are words formed by noun + noun combinations. In pairs, students will complete the missing words on the table by deducing which products could be transformed into recycled polymer. As this is not a straightforward deduction, this activity will be a good way to practice the noun + noun combinations, since during the correction of the activity students will be pronouncing those noun + noun combinations more than once in order to check their answers with the teacher.

C) Complete the table below with the information given:

Examples of uses for virgin polymer: *beverage containers, milk bottles, detergent bottles, food, wrap, sandwich bags, yogurt containers.*

Examples of uses for recycled polymer: *compost bins, coat hangers, office accessories, carpet fibers*.

Symbol and Abbreviation	Name of Polymer	Examples of Uses for Virgin Polymer	Examples of Uses for Recycled Polymer
	poly(ethylene terephthalate	_____, boil in food pouches	detergent bottles, _____, fleece jackets

	high density polyethylene	_____, _____, mixing bowls, toys, plastic bags	compost bins, detergent bottles, curbside recycling bins
	poly(vinyl chloride)	_____, vegetable oil bottles, blister packaging, plastic pipes	detergent bottles, tiles, plumbing pipe fittings
	low density polyethylene	shrink-wrap, plastic _____, squeeze bottles	films for industry and general packaging
	polypropylene	_____, grocery bags, carpet fiber, food wrap, luggage	_____, curbside recycling bins
	Polystyrene	plastic utensils, clothes hangers, foam cups and plates	_____, _____, CD boxes

Students will then use the input from the table to form sentences (exercise D). This activity will be another chance to practice new vocabulary concerning noun + noun combinations. For exercise E, students will fill in the blanks with noun + noun sequences. These are frequent sequences found in concordance lines in our corpus.

D) Fill in the blanks with information from the table above to form true sentences.

_____ can become _____

_____ can be made into _____

_____ can be transformed into _____

_____ can be turned into _____

_____ can be recycled into _____

_____ can be used to make _____

E) Fill in the blanks with the words from the box.

carbon dioxide	*light waves (2x)*	*air resistance*
hydrogen atom	*air bags*	*water molecules*

1. Since a _____ contains one proton and one electron, a hydrogen cation is simply a proton because it has lost its single electron

2. _____, for example, consist of two hydrogen atoms joined by covalent bonds to a single oxygen atom.

3. Most gaseous systems contain a mixture of gases. Air is a mixture of nitrogen gas, oxygen gas, xenon gas, _____ gas, and many others.

4. Seat belts and _____ in cars also provide a mechanism for transferring energy to prevent or reduce injury.

5. In the real world heavier weights are seen to fall more rapidly than lighter weights. But this is due to the different _____ to these falling weights.

6. _____ are microscopic; light has a wavelength in the range of 10^{-7} m. The only way we can observe the diffraction of _____ is through experimentation

5. CONCLUSION

This study stemmed from an identified deficiency in the field of English for Academic Purposes (ESP) in Brazil, more specifically, English for Reading courses aimed at science and technology undergraduate majors. We realized that the vast majority of published textbooks for English for Reading take a more generalist approach and are more focused on teaching general English rather than Academic English. In this respect, these books frequently rely on newspaper/magazine texts as the source of reading input to students. Hence, our purpose was to show that the hard sciences (here represented by chemistry

and physics) possess different linguistics features in discourse that sharply contrast to those found in news texts.

In terms of their vocabulary, the most frequent keywords in chemistry and physics diverge dramatically from those found in news. Words such as *atoms*, *electrons*, *energy*, and *force,* commonly used in the study and description of natural and scientific phenomena, are frequently used in chemistry and physics textbooks, while in news, frequent keywords include proper names (Peter, Thatcher, Kinnock), reference to cities, countries and nationalities (London, France, German), as well as words related to finances (shares, profits, dividend), and politics (MP, party, election).

Function words and modal verbs also present different patterns when we compare chemistry, physics, and news. For instance, we have found that students are almost four times more likely to see the preposition *in* in chemistry and physics than the prepositions *on* and *at*. In addition, the determiner *any* was found to be extremely productive in affirmative sentences in chemistry and physics, which contrasts to the usual information found in ESL/EFL books that emphasize the use of *any* in negative sentences. From a semantic point of view, the use of *any* in affirmative sentences brings important implications to students, as the words that correspond to *any* in Portuguese are different depending on whether it is used in affirmative, negative, or interrogative sentences. As for modal verbs, we see that physics and chemistry show a strong preference for *can*, especially in passive constructions, being almost 4 times more frequent than in news. News, on the other hand, presents *will* and *would* as the most common modal verbs.

As regards the distribution of the 50 most frequent 4-word n-grams, we found that chemistry and physics present a 50% similarity between them, while news presents a 29% similarity with chemistry, and 25% similarity with physics. In addition, the n-grams in chemistry and physics that correspond to those in news mostly belong to the category of discourse organizers (*at the same time*, *as a result*, *on the other hand*). Discourse organizers are, in fact, not very productive in chemistry and physics; in these disciplines, the most productive functional category is that of referential expressions (76% in

chemistry, and 82% in physics), followed by discourse organizers, and lastly by stance expressions. In terms of their structural patterns, both chemistry and physics show a preference (more than 50%) for the constructions *noun* phrase + *of* (*the surface of the, a large number of*), and *prepositional phrase + of (in terms of the, as a function of*).

The investigation of verb tenses/aspects also revealed considerable differences in both distribution and discourse functions for chemistry, physics and news. To begin, the simple present is the most used verb tense/aspect marker in chemistry and physics, and it is used especially to refer to scientific facts or general truths. The past simple, on the other hand, is more frequently used in news, being nearly 6 times more frequent than in chemistry and physics. While in news the simple past is used to report/narrate past events, in chemistry and physics it is used especially to introduce a historical background to what is being discussed, and to refer to a discussion previously done in the chapter/book.

The present perfect is also more frequently used in new than in chemistry and physics. Interestingly, our corpora of chemistry and physics textbooks have shown that the present perfect is mostly controlled by the pronoun *we*; 50.2% and 58.7%, respectively, especially in a dialogical manner, where the writer interacts with the reader by reminding them of some information previously discussed in the book. In contrast, in news, it is used so as to entail the writer's evaluation of the event as relating to, and being relevant for, the reader, as discussed in Chovanec (2014). As for the past perfect, it commonly used in news narrate events described by third parties, while in chemistry and physics, it is employed to narrate past historical facts or events in order to contextualize what is being discussed in the chapter.

The present progressive in chemistry and physics was found to be used by textbook authors to make their explanation more realistic, as if it were actually happening at the moment, thus creating a visual effect for the reader who is interpreting the message. Also, it is commonly used to talk about ongoing events that started in the past and are likely to continue. Both the present progressive and the past progressive are considerably more frequent in news than in chemistry and physics. In news, the past

progressive is typically employed to narrate ongoing events that took place in the past, also frequently using indirect speech, whilst in chemistry and physics commonly appears in adverbial conditional clauses (*if*-clauses), referring to a hypothetical situation.

Two of the phrasal grammatical complexity indices were also investigated, namely attributive adjectives as nominal pre-modifiers and nouns as nominal pre-modifiers. As regards *adjective + noun* combinations, this construction is around 22% more frequent in chemistry and physics than it is in News. In addition, when adjectives are analyzed from a semantic perspective, we observe that the category of classifiers is the most productive in chemistry and physics, while news shows a more even distribution of both classifiers and descriptors. As for noun + noun combinations, chemistry presents more instances of this type of construction than physics, and also possesses a larger repertory of possible candidate words to fill in the slots for N1 (premodifying noun) and N2 (head noun). IN a noun + noun combination (N1 + N2), both chemistry and physics present a larger number of potential nouns to fill in the N1 than news, in which only six premodifying nouns are, at the same time, very productive and extremely frequent: *world*, *police*, *home*, *health*, *city*, and *government.* Furthermore, when examined from a semantic perspective according to Biber et. al (1999), noun+ noun combinations in chemistry are more common in category 1 (composition) and category 5 (objective). The same is true for physics, with the addition of category 2 - purpose. News, on the other hand, presents noun + noun combinations in all of the proposed categories, except for category 11 (partitive).

This study has shown that there are several linguistic differences in the registers of chemistry and physics textbooks on one hand, and news on the other. Hence, we maintain that more specific and content-oriented pedagogical materials be prepared in the context on ESP in Brazil. As a final reflection, we would like to state that most English for Reading courses at Brazilian Universities are taught in the freshmen year, as it is believed that students will, after taking the courses, be instrumentalized to read texts in English, which will help them throughout their academic and professional lives. However, we believe that a 60-hour reading course (which is what is commonly found in Brazil) is not enough to get students to master the language thoroughly, and that it

would be most appropriate that more levels should be offered so that students could benefit from a more comprehensive exposure to the language, including other skills instead of just reading skills, as it is done in the *Inglês para fins Acadêmicos* Program, at UFMG.

REFERENCE

Anthony, L**.** (1998). Defining English for specific purposes and the role of the ESP practitioner. *Center for Language Research* 1997 Annual Review, pp.. 115-120.

Anthony, L. (2018). AntConc (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software

Baker, P. (2004). Querying keywords: Questions of difference, frequency, and sense in keywords analysis. *Journal of English Linguistics*, *32*(4), 346-359.

Bechara, E. (1999). *Moderna gramática da língua portuguesa*. Rio de Janeiro: Lucerna.

Biber, D & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.

Biber, D. (1993). Representativeness in corpus design*. Literary and linguistic computing*, 8(4), 243-257.

Biber, D. & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for specific purposes*, 26(3), 263-286.

Biber, D. & Gray, B. (2013). Nominalizing the verb phrase in academic science writing. In Aarts, B., Close, J., Leech, G., & Wallis, S. (Eds.). *The verb phrase in English: Investigating recent language change with corpora*. Cambridge University Press.

Biber, D. & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in

Biber, D., Conrad, S., & Leech, G. N. (2002*). Longman student grammar of spoken and written English*. Pearson Education.

Biber, D., Conrad, S., & Reppen, R. (1998). C*orpus linguistics: Investigating language structure and use*. Cambridge University Press.

Biber, D., Gray, B. & Poonpon, K (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *Tesol Quarterly.* 45(1) 5-35.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English*. London: Longman.

Biber, D.; Gray, B. Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes,* 9 (1), 2–20, 2010.

Burnard, L. (2008). *Reference guide to BNC Baby (second edition)*. Available online at: http://www.natcorp.ox.ac.uk/corpus/baby/manual.pdf

Byrd, P. & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, *5*(5), 31-64.

Cegalla, D. P. (2008). *Novíssima gramática da língua portuguesa.* 48. ed. rev. São Paulo: Companhia Editora Nacional.

CELANI, M. A. A., Holmes, J.Ramos, R. C. G., & Scott, M. (1988). *The Brazilian ESP project: an evaluation*. *São Paulo: EDUC*.

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language learning & technology*, *14*(2), 30-49.

Chovanec, J. (2014). *Pragmatics of tense and time in news*. Amsterdam: John Benjamins Publishing Company.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397-423.

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and education*, *17*(4), 391-406.

Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, 34(2), 213-238.

Coxhead, A. (2013). *Vocabulary and English for Specific Purposes Research*: qualitative and quantitative perspectives. Routledge.

Csomay, E., & Cortes, V. (2010). Lexical bundle distribution in university classroom talk. In *Corpus-linguistic applications* (pp. 153-168). Brill Rodopi.

Dias, R. (2002). *Reading Critically in English*. 3. ed. Belo Horizonte: Editora UFMG.

Dutra, D. P., & Berber-Sardinha, T. (2013). Referential expressions in English learner argumentative writing. Twenty Years of Learner Corpus Research: Looking back, Moving ahead. *Corpora and Language in Use–Proceedings*, *1*, 117-127.

Eastwood, J. (2002). *Oxford guide to English grammar*.7th Edition. Oxford: Oxford University Press.

Eriksson, A. 2012. 'Pedagogical perspectives on bundles: Teaching bundles to doctoral students of biochemistry'. In James Thomas & Alex Boulton (eds). *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press, 195-211.

Evaristo, S., Nunes, C., Rosa, L., Brandão, S., Sampaio, S., Araújo, D., & Franco, E. (1996). *Inglês instrumental; estratégias de leitura*. Teresina: Halley/UFPI.

Fiori-Souza, A. G.; Absy, C. A ; Costa, G.C. & Mello, L. F. (2005). *Leitura em Língua Inglesa: uma Abordagem Instrumenta*l. São Paulo: Disal.

Foley, A. (2009). Using academic word lists in the communicative classroom. *English Australia Journal*, 25(1), 17.

Folse, K. (2011). Applying L2 lexical research findings in ESL teaching. *Tesol Quarterly*, *45*(2), 362-369.

Gaies, Stephen J. T-unit analysis in second language research: Applications, problems and limitations. TESOL quarterly, p. 53-60, 1980.

Gray, B., Staples, S. &Egbert, J. (2018) *Complexity in writing development: untangling two approaches to measuring grammatical complexity*. Workshop at the 14h American Association for Corpus Linguistics (AACL) Conference. Georgia State University. Atlanta, GA.

Greenbaum, S. (1990). *A student's grammar of the English language*. Pearson Education India.

Grellet, F. (1995). Developing Reading Skills: A practical guide to reading comprehension exercises. Cambridge: Cambridge University Press*.*

Gries, S. T. (2009). *Quantitative corpus linguistics with R: A practical introduction*. Chapter 1. Routledge.

Gries, S. T., & Mukherjee, J. (2010). Lexical gravity across varieties of English: an ICE-based study of n-grams in Asian Englishes. *International Journal of Corpus Linguistics*, 15(4), 520-548.

Grondona, C.B. (2015). *Eliminação de pacotes lexicais relacionados ao tópico e de pacotes lexicais em contexto de sobreposição: uma proposta metodológica para os estudos da linguística de corpus*. Master's thesis (Master's in Applied Linguistics) - Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte.

Halliday, M. A. K. (2004). The language of science. London, England: Continuum.

Hirano, K. (1989). Research on T-unit measures in ESL. *Bull. Joetsu Univ. Educ*. 8(2) 67-77.

Holmes, J., & Celani, M. A. A. (2006). Sustainability and local knowledge: The case of the Brazilian ESP Project 1980–2005. *English for Specific Purposes*, 25(1), 109-122.

Holmes, J., & Celani, M. A. A. (2006). Sustainability and local knowledge: The case of the Brazilian ESP Project 1980–2005. English for Specific Purposes, 25(1), 109-122.

Huddleston, R. & Pullum, G. (2005). *The Cambridge grammar of the English language.* Cambridge: Cambridge University Press. Chapter 4.

Hunston, S. (2002). *Corpora in applied linguistics*. Ernst Klett Sprachen.

Hunt, K. W. (1965). Grammatical Structures Written at Three Grade Levels. NCTE Research Report No. 3.

Hunt, K. W. (1970). Syntactic maturity in schoolchildren and adults. *Monographs of the society for research in child development*, 35(1), iii-67.

Hunt, K. W. (1977). Early blooming and late blooming syntactic structures. Evaluating writing: Describing, measuring, judging, 91-106. Urbana, IL: *National Council of Teachers of English*.

Hutchinson, T., & Waters, A. (1987). *English for specific purposes*. Cambridge university press.

Hyland, K. (2004). *Genre and second language writing*. Michigan: University of Michigan Press.

Hyland, K. (2009) Writing in the disciplines: Research evidence for specificity. *Taiwan International ESP Journal*, 1(1), 5-22.

Hyland, K. (2002). Specificity revisited: how far should we go now?. *English for specific purposes*, *21*(4), 385-395.

Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. English for specific purposes, 27(1), 4-21.

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62.

Hyland, K. (2012). Bundles in academic discourse. Annual review of applied linguistics, 32, 150-169.

Johns, A. M. (2013). The history of English for specific purposes research. In:

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The modern language journal*, *73*(4), 440-464.

Larsen-Freeman, D. & Strom, V. (1977). The construction of a second language acquisition index of development. *Language Learning*, 27(1), 123-134.

Laufer, B. (2009). Second language vocabulary acquisition from language input and from form-focused activities. *Language Teaching*, 42, 341-354.

Laufer, B., & Schmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it?. *RELC journal*, 28(1), 89-108.

Leech, G., & Svartvik, J. (2013). *A communicative grammar of English*. 3rd Edition. Routledge.

Lima, E. P.. *Upstream. Inglês instrumental: petróleo e gás*. São Paulo: Cengage Learning, 2012.

Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, 31(1), 25-35.

McCarten, J. (2007). *Teaching vocabulary. Lessons from the Corpus, Lessons for the Classroom*. Available at: http://perino. pbworks. com/f/McCarten_booklet. pdf

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. New York: Routledge.

Mishan, F. & Timmis, I. (2015). *Materials development for TESOL*. Chapter 8. pp. 141-152. Edinburgh University Press.

Munhoz, R. (2000). *Inglês Instrumental: estratégias de leitura, módulo I*. São Paulo: Textonovo.

Munhoz, R. (2001). *Inglês Instrumental: estratégias de leitura, módulo II*. São Paulo. Nation, P., & Wang Ming-Tzu, K.. (1999). Graded readers and vocabulary. *Reading in a foreign language*, *12*(2), 355-380.

Neely, E., & Cortes, V. (2009). A little bit about: analyzing and teaching lexical bundles in academic lectures. *Language Value*. 1 (1), 17-38.

O'Donnell, R. (1976). A Critique of some Indices of Syntactic Maturity. *Research in the Teaching of English*. 10, (1)

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.

Olivera, N. A. (2003). Para *Ler em inglês–desenvolvimento da habilidade de leitura*. Belo Horizonte: Number One Soluções em Linguagem.

Paltridge, B., & Starfield, S. (Eds.). (2013). *The handbook of English for specific purposes*. West-Sussex: Wiley-blackwell.

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60-71.

Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. Bloomsbury Publishing.

Pitts, M., White, H. & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a Foreign Language* 5 (2), 271–275.

Ramos, R. D. C. G. (2008). ESP in Brazil: history, new trends and challenges. ESP and EAP in Developing and in Least Developing Countries. *IATEFL,* 68-83.

Saragi, T., Nation, I. S. P. & Meister, F. (1978). Vocabulary learning and reading. *System*, 6, 72–78
Scott, M. (2019). *WordSmith Tools version 7*, Stroud: Lexical Analysis Software.

Seale, C., Ziebland, S., & Charteris-Black, J. (2006). Gender, cancer experience and internet use: a comparative keyword analysis of interviews and online cancer support groups. *Social science & medicine*, 62(10), 2577-2590.

Simpson-Vlach, R. & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied linguistics*, *31*(4), 487-512.

Swales, J.M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge & New York: Cambridge University Press.

Torres, M. L. *Inglês instrumental para profissionais da saúde*. São Paulo: Allprint, 2007
university teaching and textbooks. *Applied linguistics*, 25(3), 371-405.

Wang, Y. (2017). Lexical bundles in news discourse 1784–1983. Chapter 6. *Diachronic Developments in English News Discourse*, 97-116.

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Whitman, J. (2011). *Modern Brazilian Portuguese Grammar*: *A Practical Guide*. Routledge.

Wolfe-Quintero, K.; Inagaki, S. & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. University of Hawaii Press.

Wood, D. C., & Appel, R. (2014). Multiword constructions in first year business and engineering university textbooks and EAP textbooks. *Journal of English for Academic Purposes*, 15, 1-13.

APPENDIX A:

Textbooks used in the compilation of TB-Chem and TB-Phy.

CHEMISTRY:

Atkins, P., De Paula, J., & Keeler, J. (2018). Atkins' physical chemistry. Oxford University Press.

Bishop, M.. (2006). *Introduction to chemistry*. 2nd Edition. Chiral Publishing Company.

Christian, G. D.; Dasgupta, P. H.; Schug, K. A. Analytical Chemistry. Seventh Edition. Wiley.

Cooksy, A. (2014). Physical Chemistry: Quantum Chemistry and Molecular Interactions. Boston: Pearson.

Harris, D; Lucy, C. (2016). Quantitaive Chemical Analysis. Ninth Edition. W. H. Freeman and Company, Macmillan.

Higson, S. (2006). Analytical Chemistry. Oxford: Oxford University Press.

Hofmann, A. (2018). Physical Chemistry Essentials. Springer International Publishing.

House, J. E., & House, K. A. (2015). Descriptive inorganic chemistry. Academic Press.

Housecroft, C. E., & Sharpe, A. G. (2012). Inorganic Chemistry. 4th Edition England: Pearson, 673-700.

Klein, D. R. (2012). Organic Chemistry as a Second Language: First Semester Topics (Vol. 1). John Wiley & Sons.

McMurry, J., Ballantine, D. S., Hoeger, C. A., Peterson, V. E., & Castellion, M. (2010). Fundamentals of general, organic, and biological chemistry. Pearson Education.

McQuarrie, D. A. (2008). Quantum Chemistry. Second Edition. University Science Books.

Miessler, G. L., Fischer, P.J., Tarr, D. A. (2004). Inorganic Chemistry. Fifth Edition. Pearson Education.

Ouellette, R. J., & Rawn, J. D. (2015). Principles of organic chemistry. Elsevier Academic Press.

Rogers, D. W. (2011). Concise physical chemistry. John Wiley & Sons.

Solomons, T. G., & Fryhle, C. B. (2014). Organic Chemistry. 12th Edition. John Wiley&Sons. Inc.

Suchocki, J. (2014). Conceptual chemistry. Pearson Education.

Tro, N. J. (2015). Chemistry: Structure and properties (pp. 585-587). Upper Saddle River, NJ: Pearson.

Wade, L. G., Simek, J. W. (2017). Organic Chemistry. Ninth Edition. Pearson.

PHYSICS:

Berman, P. R. (2018). Introductory Quantum Mechanics: A Traditional Approach Emphasizing Connections with Classical Physics. Springer.

Chartier, G. (2005). Introduction to optics. Springer Science & Business Media.

Choquet-Bruhat, Y. (2014). Introduction to general relativity, black holes, and cosmology. OUP Oxford.

Dewar, R. L. (2001). PHYS3001 - Classical Mechanics. Australian National University.

Espinoiza, F. (2017). Wave Motion as Inquiry: The Physics and Applications of Light and Sound. Springer.

Griffiths, D. J. (2014). Introduction to electrodynamics. Fourth Edition. Pearson.

Hewitt, P. G. (2002). Conceptual physics. Pearson Education.

Hofmann, P. (2015). Solid state physics: an introduction. John Wiley & Sons.

Kuttruff, H. (2006). Acoustics: an introduction. CRC Press.

McCall, M. (2010). Classical Mechanics: From Newton to Einstein: A Modern Introduction. John Wiley & Sons.

Mead, C. A. (2018). Collective Electrodynamics I. In Feynman and computation (pp. 29-43). CRC Press.

Nolting, W. (2017). Theoretical Physics 4: Special Theory of Relativity. Springer.

Nowikow, I., Heimbecker, B., Bosomworth, D., & Van Bemmel, H. M. (2001). Physics: Concepts and Connections. Irwin Publishing Limited.

Sachs, M. (2007). Concepts of modern physics: the Haifa lectures. Imperial College Press.

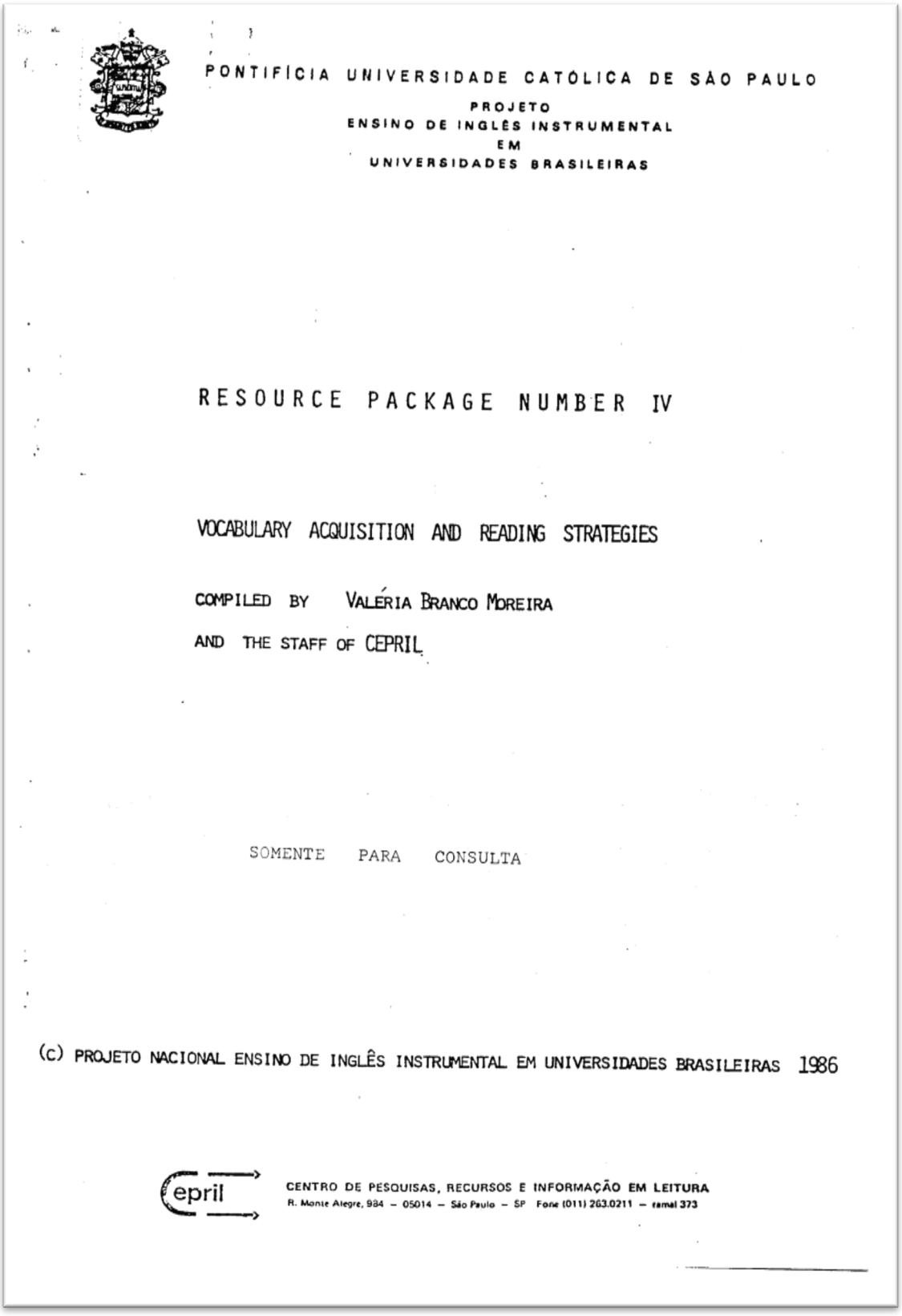
Serway, R. A., & Jewett, J. W. (2018). Physics for scientists and engineers with modern physics. Cengage Learning.

Serway, R., & Vuille, C. (2012). College physics (Vol. 1). Nelson Education.

Sidharth, B.G. (2008) The Thermodynamic Universe: Exploring the Limits of Physics. World Scientific Publishing.

APPENDIX B

Sample materials that make up the 'Resource Packages' used in the Brazilian National ESP Project during the 1980s and 1990s.



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO

PROJETO
ENSINO DE INGLÊS INSTRUMENTAL
EM
UNIVERSIDADES BRASILEIRAS

RESOURCE PACKAGE NUMBER IV

VOCABULARY ACQUISITION AND READING STRATEGIES

COMPILED BY VALÉRIA BRANCO MOREIRA

AND THE STAFF OF CEPRIL

SOMENTE PARA CONSULTA

epril
CENTRO DE PESQUISAS, RECURSOS E INFORMAÇÃO EM LEITURA
R. Monte Alegre, 984 – 05014 – São Paulo – SP Fone (011) 263.0211 – ramal 373

4. Responda as perguntas de acordo com o texto:

a. Qual é a maior evidência da crise econômica em Portugal?

b. Qual é a atitude dos trabalhadores em relação à crise?

c. Qual a atitude dos líderes do CGTP em relação à crise?

d. O que é concluído no último parágrafo?

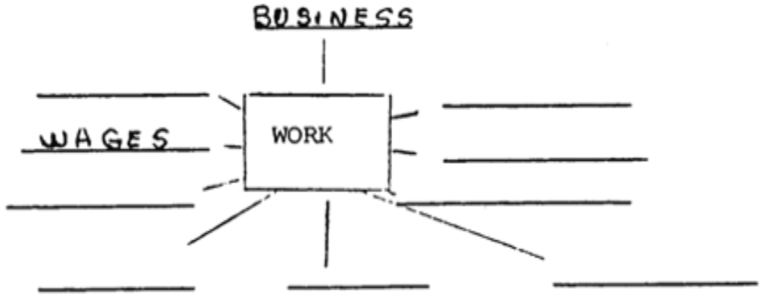
5. Relacione abaixo palavras ou expressões que o ajudaram a caracterizar o problema : a crise econômica

Workers have being going without pay

unpaid workers

ECONOMIC CRISIS

6. Escreva nos espaços abaixo palavras ou expressões que podem ser relacionadas à palavra WORK.



7. Dê o referente de:

linha 6- she _____

17- they_____

59- them_____

linha 60- who_____

81- one_____

_____

8. O que você achou do texto?

9. Você aprendeu alguma coisa nova com o texto?

10. Em que aspectos a crise econômica de Portugal pode ser comparada à sua realidade?

23

PORTUGAL

The Crisis: All Work, No Pay

Maria de Lourdes has worked as a dishwasher in a Lisbon restaurant for nearly two years—without pay. Her employer insists that he cannot afford to pay her until his business improves, and her stake is by now so high that she doesn't want to walk away from it. At Companhia Portuguesa de Pesca, a fishing company in Lisbon, over 300 employees haven't been paid for the past 18 months. Throughout Portugal more than 100,000 workers in about 600 private and state-owned companies have been going without pay for up to two years. But there have been few resignations and hardly any threats of a strike, for the simple reason that the workers know that they may lose their jobs if they challenge their employers.

The unpaid workers are the most dramatic sign of Portugal's economic crisis. After years of financial mismanagement, the country is littered with inefficient, money-losing companies in virtually every sector of the economy. Inflation is running at an annual rate of 32 percent and unemployment has reached 9 percent. The nine-month-old government of Socialist Prime Minister Mário Soares has called on the International Monetary Fund to help Portugal meet its debt obligations. But the IMF-imposed austerity measures have only added to the short-term woes of the companies and their workers. "The situation has no parallel in Europe," says Rui Cacho, an official at Portugal's largest labor union, the procommunist CGTP-Intersindical. "Employers hold back wages so they can keep their doors open and workers carry on working because they feel it is better to wait for their pay than to join the unemployment queues."

Workers and union leaders have been remarkably passive, considering the severity of the economic conditions. Most of the unpaid employees make ends meet by selling possessions or borrowing from relatives. Others take home food from company lunchrooms to help feed their families. Some workers are being paid intermittently, or are getting a portion of their regular wages; many fear that if they quit, they may never see any of their back pay.

Arrests: But the workers' patience is beginning to wear thin. Recent public-opinion polls show the government's popularity declining steadily. Late last month trade-union leaders gathered outside Soares's São Bento Palace offices and demanded an audience. The government responded by arresting 280 of the union officials and charging them with illegal picketing. The powerful communists, who have sat on the sidelines as Soares's Socialist-Social Democratic coalition government has tried to solve the economic mess, have finally decided to act. Five days after the first set of arrests, the CGTP leaders called for a series of strikes, work stoppages and street protests this month—the 10th anniversary of the country's revolution.

It is difficult to determine how effective the communist labor actions will be. The government simply lacks the money needed to pay the workers' back salaries, which now total more than $180 million. And there is no reason to expect any significant improvement in Portugal's economy until 1985. Under the circumstances, most Portuguese workers may decide that a job, even one that pays only sporadically, is better than no job at all.

MARC FRONS with KEN POTTINGER in Lisbon

Drawing by Roy Doty

Tens of thousands of workers face an unpalatable choice between no pay and no jobs at all

NEWSWEEK - April 16, 1984

UNIT 13:

VERB FORMS, TIME AND PROBABILITY

EXERCISE ONE
Skim the text.
Your objective is general comprehension.
What is the dilemma mentioned in the text?

Machine or Organism? –the dilemma of Western Medicine

Is cancer a purely physical phenomenon?
Is 'medicine' a mere synonym for 'medical technology?"

In recent years it appears that Western medicine has adopted the view of man as a machine, rather than the traditional view that man is an organism with mental as well as physical characteristics. Expensive equipment and spectacular operations all demonstrate that this mechanistic view is related to enormous financial pressures.

This is especially true in the case of two 'killer' diseases: cancer and coronary disease. It is well-known that these are influenced strongly by factors such as stress, diet, life-style and environment. The logical solution to combat these problems would be to spend money on health education. A public campaign would encourage people to change their diet, to stop smoking, to take more exercise, and to relax. This campaign would be of minimal cost to the government and patient. Instead however, we wait until the patient becomes ill and spend vast sums of money on diagnosis with scanners and X-ray tomography. Even more money goes on 'cures' such as radiotherapy or chemotherapy for cancer and in the case of heart disease spectacular operations such as by-pass surgery, valve replacement and organ transplants.

The medical industry now spends huge sums of money on cure rather than on prevention. The expensive equipment that this requires then confers increased prestige on the doctors and surgeons who use it. In order to emphasise the *preventive* role of medicine doctors would become teachers or advisers, they would spend less money and offer less spectacular solutions. To change the system would change fundamentally the entire medical profession and the practice of medicine in the Western world.

And yet a change is occurring, with a tremendous increase in the popularity of alternative approaches to medicine. Acupuncture and homeopathy, for example, are two types of medicine which treat the patient as a whole organism rather than a mechanical problem, and consider emotional and mental factors as important in assessing the diagnosis and treatment of a disease. It is possible that this increasing interest in unconventional medicine may, in the future, cause some change in the medical profession and its associated industrial-technical complex. Unfortunately, there are immense financial pressures on doctors to offer spectacular solutions rather than cheap and unimpressive, but effective treatment. This means that any change will take a long time.

Fritjof Capra: Science, Society and the Rising Culture, Harvester Press 1982

EXERCISE THREE
Identify the verbs in the text.
Do they indicate present time, past time or probability?
What is the connection between the verb forms and the two principal ideas which are discussed in the text?

UNIT 13
VERB FORMS: TIME AND PROBABILITY

13.1 Objective

To show how important verb tenses are as textual signals. By being able to identify the form of a verb (i.e. the tense), we can then identify different sections of the text. Thus, it is not necessary to be able to produce a lot of different verb forms, but sufficient to be able to identify what a verb form indicates.

13.2 Suggestions

the exercises themselves are fairly explicit. However it may be necessary to emphasise that the exercises are simply for identification of verbs and have nothing to do with long lists of verbs and irregular verb forms that some students may have been accustomed to deal with in the past.

Note that in the text used, the modals indicate the author's suggestions for the future direction of medicine while the past and present indicate the current view of man as a machine. The contrast is not made in a neatly structured way as in "The Two Cultures" but is much more complex,- identifying the verb forms is a useful tool in sorting this out.

UNIT 14
CONJUNCTIONS IN CONTEXT

14.1 Objective

To enable students to appreciate the importance of conjunctions by introducing some of the more common ones and by showing the way in which a knowledge of conjunctions can help when dealing with unknown sections of a text and in this specific case, unknown vocabulary.

14.2 Suggestions

Again, the exercises are fairly self-explanatory. By introducing the idea of four different types of conjunction we can help the student to relate the ideas in different parts of a text. Some students may wish for a list of conjunctions with their Portuguese equivalents, but it's probably unnecessary. Conjunctions are important but the variety of conjunctions in most academic writing is fairly limited.