



Máster Universitario en Aprendizaje y Enseñanza de Segundas
Lenguas

Second Language Learning and Teaching



Facultad de Educación

UNIVERSIDAD DE CANTABRIA

FACULTAD DE EDUCACIÓN

MÁSTER UNIVERSITARIO EN APRENDIZAJE Y ENSEÑANZA DE SEGUNDAS LENGUAS /
SECOND LANGUAGE LEARNING AND TEACHING

Introducing word formation through the use of native corpora to
promote word depth knowledge in English as a second language
students in Secondary Education

Introducción de la formación de palabras a través de corpus
nativos para favorecer la profundidad de vocabulario en
alumnado de inglés como segunda lengua en Educación
Secundaria

AUTORA: Ana María González Martínez

DIRECTORA: Prof. Evelyn Gandón Chapela

Firma del Director/a:

FECHA: Septiembre 2020

DECLARACIÓN PERSONAL DE NO PLAGIO

D. / Dña Ana María González Martínez

NIF 72152042F

estudiante del Máster Oficial en Aprendizaje y Enseñanza de Segundas Lenguas / Second Language Learning and Teaching de la Universidad de Cantabria, curso 20 - 20 como autor/a de este documento académico, titulado:

Introducing word formation through the use of native corpora to promote word depth knowledge in English as a second language students in Secondary Education [Introducción de la formación de palabras a través de corpus nativos para favorecer la profundidad de vocabulario en alumnado de inglés como segunda lengua en Educación Secundaria]

y presentado como Trabajo Fin de Máster, para la obtención del título correspondiente,

DECLARO QUE

este TFM es fruto de mi trabajo personal, que no copio, que no utilizo ideas, formulaciones, citas integrales e ilustraciones diversas, sacadas de cualquier obra, artículo, memoria, etc., (en versión impresa o electrónica), sin mencionar de forma clara y estricta su origen, tanto en el cuerpo del texto como en la bibliografía.

Así mismo, soy plenamente consciente de que el hecho de no respetar estos extremos es objeto de sanciones universitarias y/o de otro orden.

En Santander, a 8 de septiembre de 2020

Fdo:



Summary

This paper analyses the types and different possibilities offered by linguistic corpora, aiming to evaluate the benefits and constraints of their direct use in the classroom. Once this has been described, I inspect the concepts of word, vocabulary and lexical unit. Furthermore, I turn to morphology to examine the processes of affixation and derivation in word formation, and more specifically, how these two processes are approached in the language education curricula in Spain. Finally, I explore the pedagogical bases and most common approaches to learning vocabulary in English, among which we can find the direct use of corpora in the classroom through Data-Driven Learning. To put these elements into practice, the last part of this paper is based on the creation of a learning unit aimed for upper secondary English learners with an intermediate level within the Spanish curricular framework.

Keywords: corpus linguistics, Data-Driven Learning, English language learning, inductive learning, lexis, lexical competence, vocabulary, word depth, word formation.

Resumen

En el presente trabajo se analizan los tipos y las diferentes posibilidades ofrecidas por los corpus lingüísticos, con el objetivo de evaluar los beneficios y desventajas de su uso directo en el aula. Una vez esto ha sido descrito, se inspeccionan los conceptos de palabra, vocabulario y unidad léxica. Además, se examinan los procesos de afijación y derivación en la formación de palabras y cómo se tratan estos dos procesos en el currículum de educación de lenguas en España. Finalmente, se exploran las bases pedagógicas y los enfoques más comunes del aprendizaje de vocabulario en inglés, entre los cuales se encuentra el uso directo de los corpus en el aula a través del *Data-Driven Learning*. Para poner estos elementos en práctica, la última parte de este trabajo está basada en la creación de una unidad didáctica dirigida a estudiantes de inglés de la etapa de Bachillerato con un nivel intermedio de inglés, en el contexto del marco curricular español.

Palabras clave: lingüística de corpus, Data-Driven Learning, aprendizaje de inglés, aprendizaje inductivo, léxico, competencia léxica, vocabulario, profundidad de vocabulario, formación de palabras.

Table of contents

Summary	3
Tables and figures.....	6
1. Introduction	7
2. Theoretical framework.....	9
2.1 Corpus linguistics	9
2.1.1 Types of corpora.....	13
2.1.2 Uses and applications of corpora	17
2.1.3 Pedagogical applications of corpora to language teaching and learning	19
2.1.4 Advantages and disadvantages of using corpora in the language classroom .	21
2.2 Lexis, words, and their role in English language learning.....	24
2.2.1 What is the lexis of a language?.....	25
2.2.2 Words	28
2.2.3 Word knowledge	34
2.2.4 Lexis and the development of language proficiency.....	37
2.2.5 The role of lexis in the Spanish Secondary Education curricula.....	39
2.3 Teaching and learning vocabulary	41
2.3.1 Which words should we teach?.....	41
2.3.2 How do we learn vocabulary?.....	43
2.3.3 Approaches for teaching vocabulary	47
2.3.4 Learning vocabulary through electronic corpora and Data-Driven Learning	50
3. Didactic proposal	53
3.1 Context.....	53
3.2 Methodology	55
3.3 Competences and aims.....	57
3.3.1 Competences	57
3.3.2 Aims	59
3.4 Resources	60
3.4.1 Material resources	60
3.4.2 Corpora.....	60
3.5 Activity sequence.....	65
3.6 Assessment.....	70

4. Conclusion	71
Bibliography.....	74
Appendices	82

Tables and figures

Figure 1. Example of a List search in COCA.....	62
Figure 2. Example of a Chart search in COCA	62
Figure 3. Example of a Collocates search in COCA	63
Figure 4. Example of a Compare search in COCA	63
Figure 5. Example of a KWIC search in COCA	64
Figure 6. Example of a Browse search in COCA.....	64
Figure 7. Example of a Word search in COCA.....	65
Table 1. Word knowledge dimensions: form. Adapted from Nation (2000)	35
Table 2. Word knowledge dimensions: meaning. Adapted from Nation (2000)	35
Table 3. Word knowledge dimensions: use. Adapted from Nation (2000).....	36
Table 5. Activity sequence structure.	66

1. Introduction

Corpus linguistics has completely changed the landscape of language study. Thanks to corpora, linguistic data are accessible for researchers like never before, and their typology and applications have extended to areas like semantics, translation, or language education. Language teaching and learning have been greatly influenced by corpora, although most research has focused on their indirect applications through syllabi preparation and the creation of language teaching materials, like reference works. Due to factors like teacher unawareness, lack of teacher training or time resources, corpora are still a long way from finding their place inside the language classroom, especially outside the tertiary education or the teaching of language for specific purposes, such as scientific or academic writing (Gabrielatos, 2005). Furthermore, researchers like Römer (2006) have pointed out the constraints of learners using corpora directly, such as their incompatibility with all learning styles. Nevertheless, corpora are tools that may help to improve different aspects of language learning, like language awareness. Moreover, corpora promote learners' competence in Information and communication technology (ICTs) and encourage students through a learner-centred approach, with vocabulary learning being one of the areas that can benefit the most from it.

Generally, it has been accepted that acquiring a solid vocabulary is essential for every step in the language learning process. Canale & Swain (1980) consider vocabulary essential for the acquisition of communicative competence, and Barcroft (2004) defends that vocabulary carries more importance in the meaning of a text than grammar, as vocabulary errors may turn the message incomprehensible. Word learning goes beyond the amount (Nation, 2000). Nation (2000) establishes that there is a concept referred to as *vocabulary depth*, which implies knowing diverse aspects of words, including their morphological features. Although it is accepted that paying attention to word formation is an important vocabulary learning strategy that helps learners with meaning retention (Nation, 2000), numerous teachers assume that these processes do not require explicit teaching, because they are assumed to be inferred mechanically as the learner progresses (Tahaineh, 2012). Nevertheless, this is not the case for all learners, and many of them will acquire incomplete vocabulary knowledge, which may hinder their competence in the target language.

Since no ultimate teaching method for vocabulary has proven to be efficient for each student in fulfilling all their word knowledge needs, it will be necessary to explore the advantages that corpus-based activities offer for English language students through a process that combines both explicit and inductive vocabulary learning. This way, both teachers and students teach and acquire word formation processes in a real language context. For this reason, this paper aims to demonstrate that this can be achieved through the elaboration of a corpora-based learning unit. The programme will focus on the study of morphology within English as a First Language curriculum in Spain. In this unit, corpora are the base for the study of words and their formation processes and, in greater terms, vocabulary and language learning. This is the main objective of the paper, which, at the same time, is subdivided into four, more specific objectives. The first three objectives are aligned with the second chapter, which belongs to the theoretical framework of this work. The last objective corresponds with the creation of a learning unit.

The first subsection of this framework corresponds with the first specific objective, and it aims to explore the realm of Corpus Linguistics, paying attention to its origins, the types of corpora, and their applications, focusing on their use in language education. An analysis of the advantages and disadvantages that their direct implementation brings into the classroom is offered afterwards. The second subsection deals with the different concepts associated with lexis and types of lexical units, focusing on words. Therefore, words and their formation processes are examined, paying special attention to affixation and derivation, as well as analysing the two dimensions of word knowledge: vocabulary size or breadth and vocabulary depth.

The second specific objective of the paper consists in analysing the role of morphology within word knowledge and the relevance that morphology has for language learning, as represented in the language curricula in Spain.

To conclude the literature review, the third subsection examines the principles of vocabulary acquisition, deductive and inductive learning approaches, and vocabulary learning through corpora, focusing on the methodology of Data-Driven Learning. This corresponds with the third specific objective, and it aims to investigate and derive the implications of teaching vocabulary hands-on corpora.

The last part of this project (and fourth specific objective) is consolidated with the design of a learning unit in which all conclusions reached from these implications are put together in the third chapter. In the selection of activities, a combination of both explicit and inductive learning of word formation through affixes is present. Each activity will be evaluated according to specific criteria, and learners' progress will be recorded in a final portfolio that showcases their learning process.

2. Theoretical framework

In this section I provide a theoretical review that will set the foundation for the didactic proposal presented in section 3. First, in subsection 2.1, I examine the concept of corpora and corpus linguistics, the different types of corpora and the areas they have contributed to, paying special attention to language teaching and learning. Further, I inspect the notion of lexis and lexical unit to focus on the concept of word in subsection 2.2. The purpose of this is to review the different processes of word formation and the two dimensions of vocabulary knowledge: size and depth. After this, I examine how lexis contributes to the development of the language proficiency and set vocabulary teaching and learning within the Spanish curricular framework. Finally, in subsection 2.3, I explore the processes by which vocabulary is acquired, with a focus on the deductive-inductive debate and how learning can take place through electronic corpora and Data-Driven Learning.

2.1 Corpus linguistics

Characterizing corpus linguistics (CL henceforth) begins with its placement within the field of linguistics. Applied linguistics has traditionally been associated with language teaching (McCarthy & O'Keeffe, 2010). This assumption is not far from being right in terms of the socially accountable character that applied linguistics has, but this field has actually undergone significant changes and though it once was a synonym of language teaching, nowadays it covers a wide range of matters involving the application of language for solving real-life problems (Hunston, 2002), including speech therapy, translation and interpreting, and the central point in this paper, corpus linguistics.

But before a definition of CL is provided, the term *corpus* itself must be defined first. The noun *corpus* (plural *corpora*) is a Latin term that means "body",

(Etymological Dictionary, n.d.) and has been incorporated into the English language in fields such as medicine. In linguistics, corpora have been defined in the following way by Cheng (2012):

A corpus is a collection of texts that have been compiled for a particular reason. In other words, a corpus is [...] a collection of texts based on a set of design criteria, one of which is that the corpus aims to be representative (p.3).

This definition expands that of Biber et al. (1998), who describe a corpus as a “large and principled collection of natural texts” (p.4). Hunston (2002) also contributes to this “natural” component with his definition:

Linguists have always used the word corpus to describe a collection of naturally-occurring examples of language consisting of anything from a few sentences to a set of written texts or tape recordings, which have been collected for linguistic study. More recently the word has been reserved for collections of texts or parts of them that are stored and accessed electronically (p.2).

Based on these descriptions, one may conclude that corpora are large collections of naturally-occurring language compiled from texts of varied sizes and genres based on criteria set by researchers to study a particular language. However, to fulfill a language analysis, more than the text collection is needed. Hunston (2002) considers that a corpus by itself does not provide anything other than storing language. This author highlights that it is the software use and the electronic storage which allows researchers to approach these collections in ways that would not be possible through other means. Being able to store language data through computers thanks to the technological advances that took place in the last decades of the 20th century fully defined what we nowadays understand by corpus. No exaggeration is made if one states that linguistic corpora have completely changed the study of language.

With computer-based corpora researchers can access data like never before in terms of quantity and quality (Sinclair, 1999). This is what corpus linguists do: they compile and investigate corpora (Cheng, 2012; McEnery & Hardie, 2012). Their work is based on making generalizations about different aspects of language, like lexis or grammar, based on patterns of language use (Stubbs, 2004). Even though there is no

ultimate guide of how CL analysis should be carried out, Biber et al. (1998, p.4) determine its basic characteristics by stating that it must be:

- Empirical, based on data observation.
- Based on a corpus, which must be large and composed of natural texts.
- Based on the use of computers, combining automatic and interactive processes.
- Reliant on quantitative and also qualitative analytical techniques.

Thanks to language analysis, new insights have been provided to words, phrases, grammar, or semantics, even those that were assumed to be fairly understood by scholars. As an example, Cherifi (2019) describes in his study the importance of corpora for clarification in these cases, proving that the word *believe* is not the most suitable for academic writing as it denotes judgement or value. This demonstrates that although CL is a methodology that leads researchers to approach linguistic information objectively, their intuition is vital to interpret the findings (Sinclair, 1999). For example, investigators may suggest not using the verb *analyse* in an American paper, as data suggests that it occurs more frequently in British than in American English.

This method of searching through large amounts of text looking for patterns in words and phrases has its origins in the Middle Ages, with biblical scholars using concordances to study the Bible, like Anthony of Padua (1195–1231) or Cardinal Hugo of St Caro (1200-1263) (McCarthy & O’Keeffe, 2010). Other works and authors were also the subjects of concordancing studies, like Shakespeare, who would be studied later in the 18th century. Even though these concordances were performed by hand, the essence of the technique is still present in the software programs that we currently use.

It was the structuralist linguists who set the foundation for corpus linguistics in the 1950s when the idea of collecting real data came into its own (Llamazares, 2008). This first type of electronic corpora began to produce the first concordances by the end of the 1950s, a time in which processing the number of 60,000 words was a complex task that took more than twenty-four hours (McCarthy & O’Keeffe, 2010).

The appearance of Chomsky in the linguistic landscape at the end of this decade represented a shift in the field of linguistics in which the focus was on linguistic competence rather than performance (Tognini Bonelli, 2010). Chomsky (1965) made a

distinction between competence, which he defined as “the speaker-hearer’s knowledge of his language” (p.4) and performance, that is, “the actual use of language in concrete situations” (p.4). According to Chomsky (1965), performance could not be a direct reflection of a user’s competence, as there are diverging factors present in their natural productions. For instance, it is likely to encounter false starts or spontaneous changes in mid-discourse. This theory originated criticism towards corpus studies, as these were not considered valid tools to investigate speakers’ linguistic competence.

Despite the harsh criticism and the new shift in the trend in the field of linguistics, researchers continued working on what would be the second generation of corpora during the decades of the 60s and 70s, now influenced by the emergence of computers. These corpora were very small compared to the ones that are used nowadays (Stubbs, 2004). From this period, the creation of the first electronic corpus of written language took form as the Brown corpus. It was compiled at Brown University by Nelson Francis and Henry Kucera, and is still in use, containing one million words of written American English from different text types and topics published in 1961 (McCarthy & O’Keeffe, 2010).

The decade of the 1970s was a period of consolidation in which corpora spread to diverse languages and typologies (McCarthy & O’Keeffe, 2010). Even though development was still slow because of the limitation of the available technology, other corpora were created, like the Lancaster-Oslo/Bergen Corpus (LOB), which compiled samples of British English written in 1961, analogously to the Brown corpus, or the Survey of Spoken English (SSE) carried out by J. Svartvik at the University of Lund, which would give way to the London-Lund Corpus of Spoken English (LLC) (Llamazares, 2008).

Once the criticism towards corpora was overcome, and with new possibilities offered by the new technological advances on the horizon, electronic corpora became essential tools for language study since the decade of the eighties (Tognini Bonelli, 2010). This novel access to computers, together with the invention of hardware like scanners, recorders, or encoding systems, like Unicode¹, enabled the creation of

¹ Unicode is a universal character encoding standard for written characters that enables the user to share text data in multiple languages. It was preceded by other systems like ASCII or ISO, but these systems

different types of corpora that included more complete information (Tognini Bonelli, 2010). Since the late 1980s, new materials for language learning, such as dictionaries and grammars based on authentic language usage, like the COBUILD English Dictionary, were created (Stubbs, 2004). By the end of the 1990s, some corpora consisting of hundreds of millions of words were already created, including the Bank of English (BoE) and the British National Corpus (BNC) in the United Kingdom, which remain points of reference (Stubbs, 2004).

It would be extremely difficult to perform a search manually in the corpora available today because of their size. The development of fast software has been vital in the development and evolution of corpora, so technology can be highlighted as the principal factor in the growth of corpus linguistics. Nowadays, there is a wide selection of corpora that represent many languages and serve different purposes, as will be shown in the following subsections.

2.1.1 Types of corpora

There are diverse types of corpora depending on the purpose that was sought when they were collected. As there are a large number of corpora available and they are increasingly growing, an approximate classification can only be made. A general classification on the major English language corpora extant is summarized hereinafter, based on the work of Tognini Bonelli (2010), Lee (2010) and Römer (2010).

2.1.1.1 General and specialized corpora²

General or reference corpora attempt to be a source for all the features of a language (Römer, 2010). They contain several million words and include a selection of a wide range of text types from different registers and varieties of the language in use (spoken, written, genres...) (Römer, 2010). Typical examples of these types of corpora are the COBUILD Bank of English (BoE); the International Corpus of English (ICE); the British National Corpus (BNC) or the BYU Corpus of Contemporary American English

only worked with English characters. Unicode facilitated the creation of digital text collections in different languages. See Allen et al. (2014).

² See <https://www.sketchengine.eu/corpora-and-languages/corpus-types/> for a description of the types of corpora and <https://www.sketchengine.eu/corpora-and-languages/corpus-list/> for a list of corpora based on size and language.

(COCA). These corpora have inspired the creation of other national corpora spreading to a variety of languages across the world, like the Italian CORIS/CODIS; the German COSMAS; the Spanish Corpus del Español; the Portuguese Corpus do Português; the Russian Reference Corpus (BOKR); the Peking University Corpora, or the Korean National Corpus, among others.

In contrast, specialized corpora are collections of texts from a particular field of expertise or produced by a specific group of people (Römer, 2010). They are usually smaller in size than general corpora, custom-compiled most of the time and they have a specific purpose, like their application to language teaching. As they are not general or national corpora, they do not aim to represent a language as a whole. Instead, they represent specialized and narrowed areas of it (Römer, 2010). Some examples of specialized corpora include the Michigan Corpus of Academic Spoken English (MICASE), the Medical Web Corpus, or the English Language Newspaper Corpus (SiBol³).

2.1.1.2 Spoken and written corpora

Spoken or speech corpora refer to multimedia corpora that include recordings of the language, which may be accompanied by orthographic transcripts, phonemic and prosodic markups to facilitate their analysis (Tognini Bonelli, 2010). Among these corpora, one can find the Spoken English Corpus (SEC), the previously mentioned London-Lund Corpus of Spoken English (LLC), and some others which are more specialized in dialects, like the Newcastle Electronic Corpus of Tyneside English (NECTE), the Limerick Corpus of Irish English (L-CIE) or the Scottish Corpus of Texts and Speech (SCOTS).

Written corpora, on the other hand, are those that have been compiled including exclusively written texts from one or different genres (Tognini Bonelli, 2010). Under this classification, there are available corpora such as the British Academic Written English (BAWE), the TIME Magazine Corpus or the Wikipedia Corpus.

³ The acronym SiBol is a result of the name of the project that created the corpus, and it is a word blend of the name of the Universities of Siena and Bologna in Italy.

2.1.1.3 Monolingual and multilingual corpora

According to Lee (2010), monolingual corpora contain texts in one language only, whereas multilingual or parallel corpora inversely contain texts in several languages, usually translations. These corpora usually work paragraph-to-paragraph or sentence-to-sentence (Lee, 2010). Examples of parallel corpora may be the English-Norwegian Parallel Corpus (ENPC), the English-Swedish Parallel Corpus (ESPC) or the Oslo Multilingual Corpus (OMC), which includes texts and translations from German, French and Finnish.

2.1.1.4 Diachronic and monitor corpora

Diachronic corpora cover language at certain moments of intervals of time, portraying the language in use in such period. They allow describing and tracking changes in the language since then (Tognini Bonelli, 2010). The first diachronic corpus was the Helsinki Corpus of English, covering exemplars of texts ranging from the 8th to the 18th century.

Monitor corpora are created to track language as it occurs to observe language change. Instead of suggesting the replacement of materials with more recent samples, they are retained and tagged in time (Tognini Bonelli, 2010). One example is the AVIATOR project, which attempted to collect an annual amount of over ten million words of texts from the Times newspaper. Other examples include the Representative Corpus of Historical English Registers (ARCHER), which covers the early Modern English period up to the present (17th - 20th centuries), or the Corpus of Historical American English (COHA), covering from the early 1800s to the present time.

2.1.1.5 Parsed corpora

A parsed corpus implies that words have been syntactically analysed at a phrasal level and tagged so that it is possible to search by structural syntactic functions (Lee, 2010). They are usually smaller than normal or unparsed corpora, as they usually involve a process of checking by hand to improve their accuracy (Lee, 2010). Examples of tagged corpora are the York-Helsinki Parsed Corpus of Old English Poetry, the Penn-Helsinki Parsed Corpus of Middle English (PPCME), the Penn Parsed Corpus of Modern British English (PPCMBE), or the Parsed Corpus of Early English Correspondence (PCEEC).

2.1.1.6 Learner corpora

Granger (2019, p.1) defines learner corpora as “electronic collections of language data produced by foreign language learners”. This author points out that one of the main characteristics of these corpora is that language acquisition processes are examined through the scope of corpus linguistics to understand foreign language students’ errors and design pedagogical tools that target them. Granger also distinguishes the following characteristics in a learner corpus:

- Electronic format.
- Compilation criteria based on learner characteristics, such as age, gender, or mother tongue.
- Complete discourse is included instead of isolated words or sentences.
- The data they include might come from natural or semi-natural communication from learning tasks in which learners have to choose their own wording (p.1).

The first learner corpora have their origins in the decade of the eighties, and they have expanded since then. The principal learner corpus projects at present are the International Corpus of Learner English (ICLE) and the Louvain International Database of Spoken English Interlanguage (LINDSEI). However, these represent the productions of upper-level students, with a gap in the representation of younger learners of English. The International Corpus of Crosslinguistic Interlanguage (ICCI) is aimed to fill this niche, and it is currently being compiled in European and Asian countries. Finally, examples can be found of specialized corpora of learner English, such as the Learner Business Letters Corpus (Learner BLC) or the Learning Prosody in a Foreign Language Corpus (LeaP Corpus). Learner corpora have also been created for the benefit of young learners and their educators. In 2018, a team of researchers at University of Cantabria (Spain) compiled the Primary Education Learners’ English Corpus (PELEC), which gathers over 60,000 words of written compositions and spoken productions of English as second language students at this educative stage (see Blanco-Suárez, Gallardo-del-Puerto & Gandón-Chapela, 2020). This will allow instructors to examine the most common errors and adapt their teaching practice and materials in order to benefit these learners, who are in a decisive stage in their language learning journey.

Overall, it can be concluded that corpus data are essential for describing language use. As a result, different types of corpora have been applied in diverse fields,

such as translation, cultural analysis, or language teaching, to name but a few. This variety in their application will be examined in more detail in the following subsection.

2.1.2 Uses and applications of corpora

CL is being used extensively. In this subsection, the most common practices in CL, which apply to a wide range of areas in linguistics, are presented. For this purpose, the different uses have been organized into six main blocks based on the work of McEney & Hardie (2012) and Hunston (2002): lexicography and lexical studies; study of language features, such as grammar or semantics; translation; sociolinguistics; forensic linguistics and language teaching and learning. It is this last use of CL which holds most importance in this chapter and for the present dissertation, and, for this reason, it is going to be analysed more thoroughly.

2.1.2.1 Lexicography and lexical studies

Nowadays, it is difficult to find an updated version of a dictionary that does not rely on data retrieved from corpora. These tools allow lexicographers to access and retrieve authentic examples of the usage of different terms based on frequency and collocation data (Hunston, 2002). Some dictionaries, like COBUILD or Longman, include this information in the word entries making it possible to get a description of the word along with information about collocation, register or even group of age. Furthermore, the monitor character of some corpora allows researchers to keep track of language changes over time and update these sources (Hunston, 2002).

2.1.2.2 Study of different language features

Corpus-based studies can be oriented to examine different language features like lexis, grammar, semantics, register and genres, dialects and language varieties, language changes, pragmatics, discourse analysis or stylistics (Hunston, 2002). Corpora provide information about the behaviour of these language features in context and allow observing multiple features. For instance, in the case of lexis, corpora are used to examine patterns of collocation or morphology, by examining the internal structure of words at the sub-lexical level in terms of roots and affixes (Hunston, 2002).

2.1.2.3 Translation

Corpus linguistics allows comparing between source and target texts and examining patterns across languages automatically (Gilquin & Granger, 2010). Corpora are used to analyse how an idea in a particular language can be conveyed in another, comparing their linguistic features and the frequency of term use (Stubbs, 2004). They have also been used in translator training and in the development of computer-based translation systems (Stubbs, 2004). Parallel and multilingual corpora have been especially useful for these purposes.

2.1.2.4 Sociolinguistics

Corpus-based sociolinguistic research has focused on issues such as gender studies, like language and sexism, femininity or sexual identity, and other aspects such as race, age or social class (Hunston, 2002). The increasing availability of corpora that can provide information about the context of the text and metadata has played an important role in this field (Hunston, 2002).

2.1.2.5 Forensic linguistics

Forensic linguistics studies the use of language in court trials and examines linguistic evidence (Hunston, 2002). This is a field in which CL has a protagonist position, as conclusions reached through this study can affect court verdicts. Language is expected to be impartial in judicial contexts, but corpus study of testimonies has shown that it can be evaluative, showing judgement and value (Hunston, 2002). Language choices may affect how the judge, defence or witnesses in a case are regarded.

Further, corpora are being increasingly used for plagiarism detection. CL is used to analyse texts and reach conclusions in cases in which authorship may be questionable (Hunston, 2002).

2.1.2.6 Language teaching and learning

Corpus-based research began to expand to language teaching and learning. Corpora have been applied both indirectly, in the form of learning material creation, and directly (Römer, 2011). In this last approach to corpus use, corpora are brought into the classroom and accessed directly by students, with the teacher as a facilitator of learning

in a method that has been called Data-driven Learning or DDL (Johns, 1991). This method has brought many benefits and challenges, as is going to be explored in subsection 2.1.4.

Other areas that are being expanded within pedagogical applied linguistics are the creation of learner corpora, the use of corpora for creating and validating language tests, and teacher training (Römer, 2011).

While, in general terms, corpus-based instruction seems to have had an impact on language learners, it is necessary to justify this by looking closely at the varied applications of corpora for language teaching in the next subsection.

2.1.3 Pedagogical applications of corpora to language teaching and learning

The influence of corpora has also extended to the field of second language teaching and learning. This field holds the most importance in this chapter and for this paper, so it is going to be analysed more thoroughly.

Nowadays, researchers and specialists have increasingly valued all the options that CL offers to language pedagogy. Based on these perspectives and the work of Römer (2006), a general distinction can be made between direct and indirect corpus uses in second language (L2 henceforth) learning. This author points out that indirect approaches refer to corpora applied for syllabus design and teaching materials creation, while direct approaches refer to the direct access to corpora from teachers and learners in the language classroom. Finally, the appearance of learner corpora in the scene has also been a useful resource for language instruction.

2.1.3.1 Indirect corpus applications

The indirect approach places the focus on researchers, who use corpus evidence to examine language in use and to study how corpora may contribute to making the process easier for learners (Römer, 2011). Römer (2006) and Conrad & Levelle (2008) distinguish different types of indirect pedagogical corpus applications, which include using corpora to improve course designs and preparing class syllabi, and creating pedagogical materials, like references, dictionaries, grammars and textbooks.

2.1.3.2 Direct corpus applications

According to Römer (2006), following a direct approach implies that “teachers and learners get their hands on corpus data themselves, instead of having to rely on the researcher as mediator or provider of corpus-based materials” (p.124). Corpora applications are available for teachers and students, as presented below.

a) Teacher-corpus interaction: The teacher interacts with corpora as a researcher and practitioner (McCarthy, 2008). When teachers have a language doubt or query and are in need of an explanation, a corpus search can help to obtain an answer (Cobb & Boulton, 2015). What is more, some teachers even compile their own corpora for the specific purpose of examining language that might be troublesome for their students, so that they can work on such issues (Conrad & Levelle, 2008).

b) Learner-corpus interaction: It consists in learners using learner or native corpora in the classroom. They may use them as reference tools along with dictionaries or data sources from which they will infer all knowledge (Cobb & Boulton, 2015). This last use consists in students deriving all knowledge from concordances and corpus analyses, a method also known as Data-Driven Learning or DDL, as mentioned earlier in this chapter. These activities range from exercises with concordance lines previously prepared by the teacher to their own conclusions based on the corpus analysed (Nesselhauf, 2004).

Learner corpora have also been, although little, considered in language teaching up until very recently: the focus relied primarily on native speaker corpora (Nesselhauf, 2004). Like native corpora, they are indirectly used to determine the difficulties in learning for a particular group of students so that these aspects can be studied more profoundly in materials. Further, learner corpora and data are occasionally used directly in the classroom (Nesselhauf, 2004).

As can be observed, teachers are beginning to become more aware of the possibilities of using corpora, but there is still a gap between theory and actual pedagogical implantation and a long way to go (McCarthy, 2008; Römer, 2006). There is still much work to do to bridge this theory-practice gap, and, while in general terms corpus-based instruction seems to provide multiple benefits for language learners, it still

seems to be necessary to justify this by looking at the evidence provided by research to date.

2.1.4 Advantages and disadvantages of using corpora in the language classroom

It has been observed that corpora are increasingly being incorporated into language teaching and learning in both indirect and direct ways (Nesselhauf, 2004). This has brought new benefits for language teachers and students, but the use of corpora has also brought new inconveniences associated with their use. Gilquin & Granger (2010), Conrad & Levelle (2008), Gabrielatos (2005), Meunier (2011), Boulton & Tyne (2015) and Römer (2006) have critically analysed these pedagogical applications. The main advantages and problems that the use of corpora brings along have been studied and compiled in this chapter, so that the whole spectrum of what using corpus linguistics for pedagogical purposes implies can be considered.

Firstly, the advantages of using of corpora in the classroom for students and teachers have been examined:

- **Authenticity:** Corpora make it possible to examine authentic and naturally-occurring language data, produced in real communication situations (Gilquin & Granger, 2010).
- **Representativeness:** Corpus projects aim to make corpora as representative as possible, including a wide range of samples and taking into account the characteristics of the speakers of a language in particular and the different contexts in which it may be present (Conrad & Levelle, 2008).
- **Variety:** A large number of samples of a particular item is available and can be studied (Gilquin & Granger, 2010). Corpora provide the opportunity to explore the different alternatives used by native speakers in different contexts and frequencies (Gabrielatos, 2005). Furthermore, as this author points out, learners work with corpora that represent different varieties and genres, which provide them rich exposure.
- **Empirically-based:** Conclusions reached from corpus examination are not based on individual intuitions about how language is used, but on real and observable data (Conrad & Levelle, 2008). This way, corpora may provide information not found otherwise in a dictionary or grammar book (Römer, 2006).

- **Autonomy-promoting:** Learners have more freedom and become more responsible for their own instruction when accessing corpora (Gilquin & Granger, 2010). Conrad & Levelle (2008) observe that learner autonomy increases as they learn how to make generalizations based on observable data, instead of relying completely on the knowledge presented by their teachers.
- **Motivational:** Following an inductive approach can be appealing for those students with different learning styles or needs instead of the traditional deductive approach for teaching language rules (Conrad & Levelle, 2008). It enhances the discovery factor of learning, in which students take the role of language researchers (Gabrielatos, 2005).
- **Innovative:** Learners explore language through the use of new technologies (Gabrielatos, 2005).

Once the advantages of the use of corpora in the classroom have been analysed, it is important to explore the challenges that corpora imply for both teachers and learners:

- **Teacher reticence:** Teachers do not always share the impressions of linguists about the benefits of corpora and are reluctant to use them in their language classroom. Different authors have provided explanations for this. Meunier (2011) blames this on their lack of awareness on the benefits that corpora can provide, while Gilquin & Granger (2010) believe that they are not trained in this field and do not know enough about corpora to be used in the classroom. On the other hand, these authors suggest as well that a lack of resources is sometimes the explanation, whereas other times they are sceptic about whether this method is effective. As they summarize, using corpora directly implies that the focus moves from teacher-led to learner-led, implying that teachers have a less central role than in traditional methods.
- **Lack of studies:** Many authors, like Meunier (2011) or Conrad & Levelle (2008), suggest that there is a lack of empirical studies that test the efficacy of corpus methods on language learning in terms of outcomes, which is necessary to shed light on the types of activities or skills that would be influenced the most from this approach.
- **A knowledge foundation is required:** Meunier (2011) observes that it takes time and practice for students to become independent language users. In fact, Boulton & Tyne (2015) defend that the inductive processes involved in this approach can be too

demanding for novice learners. Gilquin & Granger (2010) highlight that some areas may present problems for learners, like the annotation of tagged corpora, the Keyword in Context (KWIC) view, discerning the irrelevant hits and the language that the teacher is not interested in students to learn, like swear words or literary phrases. On the other hand, it is necessary to become acquainted with corpora in particular. As Römer (2006) points out, it takes time for corpus users to familiarize with the tool, so basic training would be necessary. Meunier (2011) also claims that teachers need to have a good understanding of how corpora work in order to provide useful aid and to achieve the intended goals.

- Frequency-based problems: In the same way that it would not be useful to elude information about frequency for language learning, it would not be beneficial to completely abide by it (Meunier, 2011). This author elaborates on this adding that higher achievements in learning are related with knowing less frequent words and less common uses, so it is crucial to cover the whole range of frequency and not only those at the top.
- Resources: If learners are to use corpora in the classroom, they need at least one computer for every pair of students, access to corpora and other software. All this costs money, and some schools are not always able to afford them (Gilquin & Granger, 2010). Furthermore, these authors point out that even though some corpora are free, they may have more limited features than those bought, and the creation of one's own material takes time. This is another resource that implies an obstacle for the direct use of corpora. According to Gilquin and Granger (2010), it is time-consuming to prepare the teaching materials, to train students in the use of corpora and to complete a search task.
- Not suitable for all: Even though this approach to learning may be beneficial, it is not appealing for all students, especially those who prefer or perform better with traditional approaches. Moreover, not all learners may feel comfortable working with technologies for language learning (Römer, 2006).
- Representativeness: Corpora ideally represent samples of a language variety. However, generalizations are usually made and corpora are viewed as the language as a whole (Gabrielatos, 2005). As Gabrielatos (2005) suggests, in consequence, there is an over-reliance on corpus data, and it is important to comprehend that corpora cannot capture the entirety of language in use. It is the case of spoken data, for instance.

Tognini Bonelli (2010) argues that spoken data are still scarce because automatic linguistic processing of the speech wave and speech recognition software are still in trial and require skilled and expensive treatment.

- Opposing a communicative language approach: Leńko-Szymańska & Boulton (2015) believe that corpus analysis of language is incompatible with a communicative language teaching methodology, because it is an approach which is more focused on accuracy than fluency. Working with corpora implies that the focus is placed on language patterns and grammatical structures and not on communication, which is considered indispensable for language learning as well.

Overall, in spite of the progress that has been achieved in the use of corpora in the language classroom, there is still room for improvement and some pedagogical considerations need to be taken into account for carrying out any task with this approach. Römer (2006) suggests that in order to improve the teaching practice with corpora, more attention to language teachers, support and assessment of their needs is necessary.

Corpora are, by no means, an ultimate method to solve any teaching problem and are not meant to substitute other teaching methodologies. They are, as Gabrielatos (2005) describes, a good tool to enrich and enhance such methods. Finally, and as Boulton & Tyne (2015) point out, corpora can be an additional technique to improve different aspects of language learning, like language awareness, while also promoting learners' competence in ICT tools and motivating them through a learner-centred approach. Consequently, the knowledge and skills obtained hold the potential of becoming life-long learning. Vocabulary is one of the aspects that may benefit the most from corpus study, a matter that is going to be examined more profoundly in the following chapters.

2.2 Lexis, words, and their role in English language learning

Lexis holds a relevant role within the acquisition of a second language, but before dealing with the processes involved in learning the vocabulary of an L2, it is necessary to identify and make a distinction among the different concepts involved in the matter, such as *lexis*, *lexicon* and *vocabulary*, to approach the topic accurately.

2.2.1 What is the lexis of a language?

Establishing what the vocabulary of a language is and defining the concept of *word* is a complex task that depends on diverse criteria. Research on these topics has resulted in a variety of terms that are used when talking about lexis, and they need to be examined to distinguish what they involve.

Firstly, let us examine how the main English dictionaries describe the term *lexis*. Most of them agree that the word *lexis* has its etymological origin on the Greek λέξις, which means “word” or “speech”. The Collins English Dictionary (Harper Collins, n.d.) defines *lexis* as “the totality of words in a language, including all forms having lexical meaning or grammatical function” in the British English section, and “The full vocabulary of a language, or of a group, individual, field of study, etc.” in the American English section. The Cambridge Dictionary (Cambridge University Press, n.d.) defines it simply as “all the words of a language”, and Merriam-Webster (Merriam-Webster, n.d.) describes *lexis* as “the vocabulary of a language, an individual speaker or group of speakers, or a subject”.

These three reference sources share in common in their definitions the inclusion of the concepts of *full* or *all*, *vocabulary*, and *language*. In fact, *vocabulary* is often used as an interchangeable substitute of *lexis*, but whether these two terms share the same meaning or have different connotations is to be examined.

The Collins English Dictionary (Harper Collins, n.d.) considers that *vocabulary* is “the total number of words you know in a particular language”, “the vocabulary of a language is all the words in it”, and that “the vocabulary of a subject is the group of words that are typically used when discussing it”. Merriam-Webster (Merriam-Webster, n.d.) also distinguishes different meanings for *vocabulary*:

1: a list or collection of words or of words and phrases usually alphabetically arranged and explained or defined.

2a: a sum or stock of words employed by a language, group, individual, or work or in a field of knowledge.

2b: a list or collection of terms or codes available for use (as in an indexing system).

Finally, the Cambridge Dictionary (Cambridge University Press, n.d.) considers that the term *vocabulary* refers to “all the words known and used by a person or all those that exist in a particular language or subject”. It can be observed that these definitions again imply collections of words, and although the definitions of *lexis* and *vocabulary* are very similar, some authors make a subtle distinction between them. As Caro & Mendinueta (2017) point out, people tend to associate the concept of *vocabulary* with words and meanings, whereas *lexis* is broader and engulfs these along with lexemes and other lexical items. In this paper, both terms are used interchangeably as well, but bearing in mind that the concept of *lexis* is more complex.

It is now when the concept of *lexicon* takes part in the narrative associated with notions of mind, concepts, and *lexis*. The term *lexicon* has, like the others, been studied by researchers. English dictionaries, like the Collins Dictionary, distinguish two main meanings of this word, associating it with a dictionary or with *vocabulary* as well. This last sense, the mental *lexicon*, is defined by Baralo Ottonello (2001) as storage for words that becomes available for the speaker to use according to his or her needs. Lipka (1992) expands this by adding that it is not just a collection of isolated elements, but it has a structure in which elements are connected and related.

In their definition of *lexis*, the Collins Dictionary includes not only the words *per se*, but also all their forms containing “lexical meaning and grammatical function”. The Merriam-Webster Dictionary also includes the notion of phrases in their definition of *vocabulary*. All of this implies that some authors abandon the idea that words alone are learnt individually to construct meaning. As Willis (2003) states, much of our production is not composed of individual words, but groups of them that we use as fixed phrases. In such phrases, he proposes, we do not need to analyse each word to work out the meaning. Instead, they operate as a single unit, and they are part of our everyday speech. This notion that words are independent units comes from the fact that words are usually presented separately between spaces in the written language (Almela & Sánchez, 2007). By taking a closer look at different languages, Halliday & Yallop (2007) examine that there is no universal entity to all languages that can be equated with the concept of *word*. However, there is a concept underlying these problems that replaces what *words* vaguely represent, and that is the *lexical unit*.

A *lexical unit*, according to Bogaards (2001), involves much more than an aggragation of letters. For a *lexical unit* to be considered as such, it must contain a semantic component (which may be a word, a phrase or a sentence) that contributes to the overall meaning and one word at least. Bogaards (2001), Willis (2003) and Bybee (1998) recognize various types of lexical units: lexical phrases and their subtypes, collocations, idioms and words. These are further examined hereinafter:

- Lexical phrases: Willis (2003) distinguishes different types of them.
 - Polywords: They are phrases that reappear and do not change their form, like *so far so good*. Many of them are time adverbials (*the day after tomorrow*), place adverbials (*over there*) and sentence adverbials (*in fact*). Others are two or three-part verbs, also called phrasal verbs, such as *carry on* or *take off*.
 - Frames: They are not continuous; they are frames with gaps that can be completed by different words, depending on the context. For instance, “not a matter of... but...”
 - Sentences and sentence stems: These are lexical phrases that are full sentences in themselves. Many of these are social acts, like *How do you do?*
 - Patterns: Patterns are similar to frames, but the words needed to complete them are somewhat predictable because of the meaning of the sentence. For instance, in the case of *read*, it is likely to find nouns that imply communication, like *book, newspaper* or *article*.
- Collocations: Words collocate when they occur together quite frequently. For instance, *drink* and *water* (Willis, 2003).
- Idioms: Idioms are sequences of words that are stored in memory. A proof of this is the fact that many idioms contain words that are no longer in use (Bybee, 1998), like the *fro* in *to and fro*. In this case, *fro* is a remnant of a way of pronouncing *from*.
- Words: Words are still considered the main unit of vocabulary, despite the problems that their concept implies (Willis, 2003). This type of lexical unit is going to be examined further in the following subsection.

As a conclusion, one may agree that the concept of *lexical unit* is useful and precise because it allows examining the process that learners have to undergo to acquire vocabulary in their second language (Bogaards, 2001). Because of the large number of lexical units in a language and so many aspects to learn about each of them, it is necessary to regard positively the high level of competence that many foreign language learners achieve, and consider what may be done for those who do not achieve such competence.

2.2.2 Words

A commonly shared notion is that languages are made up of words. We use them every day, as part of our communicative acts, and we even use them for recreational purposes: they are present on board games, like Scrabble; we play trying to represent their meanings, and we engage in spelling contests. It is not uncommon either to turn to dictionaries for help anytime we are in need of deciphering the meaning of a word.

Being so familiar with words, it should not be difficult to define them. However, the concept of *word* turns out to be a complex term, because its definition depends on various aspects. Ginzburg et al. (1979) describe *words* as the basic lexical unit, which are put together to form phrases or word groups. Jackson (2002), on the other hand, supports the widespread concept of *words* as sequences of letters that are limited by spaces on either side. Counting words is nevertheless a difficult task that depends on different criteria, according to Nation (2000):

- a) Counting tokens: involves counting every word form even if it occurs more than once. It is the total number of words, useful for measuring the reading speed of the subject, for instance.
- b) Counting types: if the same word occurs again in a text, it is not counted.
- c) Counting lemmas: they consist of a base form (*run*) and all its inflections (*runs* or *running*), according to the word class.
- d) Counting word families: it involves a base or root (*govern*), all its inflections (*governed*) and their closely related derived forms as well (*government*, *governable*).

According to these criteria, the sentence “I came I saw I concordanced, I come I see a concordance” would consist of twelve tokens (*I came I saw I concordance, I come I see*

a concordance); eight types (*I, came, saw, concordanced, come, see, a, concordance*); six lemmas (*I, come, see, concord* as a verb, *a, concord* as a noun), and five word families (*I, come, see, concord, a*). This shows that the criteria applied when considering words can affect the result. Another issue that derives from this is the word sense when deciding whether two forms are instances of the same word or not, such as *like* (meaning “similar to”) and *like* (meaning “to be fond of”) (Halliday & Yallop, 2007). This conception comes, in reality, from the written form, in which words are more visually distinguishable. However, in the flow of speech they usually follow each other without spaces or pauses (Jackson, 2002). If written and spoken language are taken into account, this author recognizes also a classification of orthographic words, phonological words, and lexemes (any word in the vocabulary of a language).

But these are not the only ways in which words can be classified. Halliday & Yallop (2007) point out that there are English teachers that distinguish between content words like *concordance* and function words, like *a* or *the*. Finally, Ginzburg et al. (1979) recognize two groups, monosemantic (one meaning) and polysemantic words (more than one meaning), according to the number of meanings a word possesses.

As may be observed, a word can be classified according to a wide range of criteria. Nevertheless, the most common criterion is the conventional classification of “parts of speech” or “word classes”. Jackson (2002) distinguishes among these four large or open classes (nouns, verbs, adjectives and adverbs), which are those to which new words can be added, and four smaller or closed classes (pronouns, determiners, prepositions and conjunctions), which are more fixed, cannot be added new terms and whose purpose is to link the members of the largest classes together.

Up to this point, it has been discussed that the notion of *word* is not fixed. It seems, nonetheless, that the importance of words relies on them acting as triggers for competent speakers (Willis, 2003). This means that one word can suggest instances of others that are likely to be present around it, and it can also provide information about the type of sentence or pattern that is likely to occur. However, achieving this state is an intricate task and the line between total ignorance, partial or full word knowledge is not always clear (Bogaards, 2001). An analysis of how words are formed is carried out in the following subsection.

2.2.2.1 Word formation

Once the concept of *word* and its classification has been examined, their internal structure must be analysed to understand their nature, which is done through the realm of linguistics called morphology.

This analysis must start with the definition of the term *lexeme*. Lexemes are abstract units that carry the basic meaning of a word, and different forms and inflections can be drawn from them (Jackson, 2002).

The smaller units that can be distinguished within a word are called *morphemes*. They are the smallest unit of language that have meaning and cannot be broken down into smaller parts (Tahaineh, 2012). The following types can be distinguished, as Tahaineh (2012) describes:

a) Free morphemes: They stand alone as independent simple words with meaning, and they involve only one morpheme (*boy, cat, read*). They are single root bases by themselves, and additional morphemes can be added to them to build up new words.

b) Bound morphemes: They are those that must appear with another morpheme or word in order to have a meaning. Affixes (i.e. prefixes, suffixes and infixes) are attached to a stem or root, and they can be classified into two categories:

- Derivational morphemes: These can be prefixes, suffixes or infixes. Prefixes attach to the front of a base, like *de-* in *deconstruct*. Suffixes attach at the end of the base, like *-ness* in *kindness*. Infixes, lastly, which are by far less common in English, are inserted within a root, like *passerby* forming its plural as *passersby*. They create complex words.

- Inflectional morphemes: They are always suffixes like the plural-forming *-s*, and the *-ing* or *-ed* that are added to verb stems.

Morphemes offer English speakers a wide array of resources to create words as they need them in a particular context. However, words are also formed through different processes. Tahaineh (2012) establishes a classification of different word formation processes in English. Some of the most common processes are described by this author (p.1108) as the following:

- Compounding: Two or more words that are joined produce a new single one, like *handbag*.
- Borrowing: Loanwords that are picked from other languages, like *bazaar* from Persian, meaning *market*.
- Conversion or zero derivation: A lexical item is changed from one grammatical class to another without affixation, like the noun *bottle* to the verb *to bottle*.
- Stress shift: When pronounced, the word stress is moved from one syllable to another, like *transport* (/ˈtrænsˌpɔːrt/) to *transport* (/trænsˈpɔːrt/), changing the grammatical class of the word (noun and verb, respectively).
- Clipping: Words of more than one syllable are reduced in casual speech, like *flu* from *influenza*.
- Acronym formation: Words are formed from the initials of a group of words, like NASA (National Aeronautics and Space Administration).
- Blending: Instead of morphemes, two parts of already-formed words are joined to create a new one, like *brunch* (*breakfast* and *lunch*).
- Backformation: A suffix is removed from the base, and this base is used as a word (like *babysit* from *babysitter* or *burger* from *hamburger*).
- Coinage: Invention of brand new terms, most of them from a company's product that becomes the generalization, like *Kleenex*.
- Onomatopoeia: Words that sound like the sound they name, like *buzz* or *crack*.

Further, derivation is one of the most frequent processes involved in word formation in English (Tahaineh, 2012). This implies that a distinction must be made between the two processes involved with lexemes: derivation and inflection. Derivation deals with creating new lexemes, whilst inflection is involved with creating different forms of such lexemes (Booij, 2006). For instance, creating *walker*, a noun derived from the verb *walk*, would involve a process of derivation, because it has changed its word category (a noun) (Booij, 2006). On the other hand, creating the verb forms *sings*, *singing*, *sang* and *sung* from the verbal lexeme *sing* would involve a process of inflection, which provides information about tense, number and person (Jackson, 2002).

Finally, Tahaineh (2012) points out that affixation is another common process by which words are formed. Although both affixation and derivation involve the intervention of affixes, there is a subtle difference between them. Affixation consists in combining affixes with roots. Derivation, as was previously distinguished, consists in joining together affixes with already existing words to create new ones that belong to a different grammatical category, like *refuse* (verb) to *refus-al* (noun) (Tahaineh, 2012).

Word derivation and affixation take place only in the category of open word classes, those that allow forming new words (nouns, verbs, adjectives and adverbs) (Jackson, 2002), as was examined in the previous subsection. A characterization of these two processes is provided hereinafter:

▪ **Derivation:** Words take suffixes to change their category, as prefixes rarely take part in changing the category of a noun (Blevins, 2006). This involves forming nouns, verbs, adjectives and adverbs from any of these categories. Some examples of these processes are the following, based on Comrie & Thompson (1985) and Gillett (2020):

- **Nouns:** Some examples of noun derivation are the suffix *-ion* (forming *direction* from the verb *direct*), *-er* (forming *astronomer* from the noun *astronomy*) or *-ness* (forming *rudeness* from the adjective *rude*). These suffixes change the word class of the noun, and affect their meaning as well. For instance, the suffixes *-ity* or *-ness* imply “a state or quality of”.

- **Verbs:** Some examples of verb derivation are the suffix *-ise/-ize*⁴ (forming *symbolise* or *symbolize* from the noun *symbol*), *-ate* (forming *liquidate* from the noun *liquid*) or *-en* (forming *shorten* from the adjective *short*). These suffixes change the word class of the verb, which takes the meaning of “causing to be”.

- **Adjectives:** Some examples of adjective derivation are the suffix *-ful* (forming *beautiful* from the noun *beauty*), *-less* (forming *hopeless* from the noun *hope*) or *-able* (forming *countable* from the verb *count*). These suffixes affect the word meaning as well, implying “full of”, “without”, and “fit for”, respectively.

- **Adverbs:** As stated in Collins Easy Learning Grammar (n.d.), Adverbs are generally based on adjectives, and formed by adding the suffix *-ly* to the

⁴ *-ise* is prescribed in British English whilst *-ize* is the North American English spelling variant.

adjective. They usually denote manner or degree. For instance, forming *calm* from *calmly*, or *complete* from *completely*.

▪ Affixation: By adding prefixes, the meaning of words is affected, but they rarely change the category they belong to (Blevins, 2006). Some examples of this process are provided below, based on Comrie & Thompson (1985) and Gillett (2020):

- Nouns: *co-*, as in *co-owner*; *mal-*, as in *malnutrition*; or *re-*, as in *reassessment*. By taking these prefixes, these words acquire the connotation of “together”, “bad/wrong”, and “again” respectively.

- Verbs: *un-* as in *undo*, *pre-* as in *prefabricate*, *out-* as in *outperform*. By taking these prefixes, these words acquire the connotation of “opposite of”, “before”, and “better than”, respectively.

- Adjectives: *im-/in-/ir-/il-*⁵ as in *impatient*, *insufficient*, *irrational* or *illegal*, *non-* as in *non-neutral*, *dis-* as in *dishonest*. By taking these prefixes, these words acquire the connotation of “opposite of”.

This classification suggests that there are recognizable and predictable patterns involved in word building. Many language teachers assume that these are not in need of explicit learning, because students will end up, at some point, inferring them while paying attention to other processes (Tahaineh, 2012). However, teaching these mechanisms is an area worthy of attention in English language teaching. Kim (2013) suggests that morphological awareness, which involves being aware of the meaning and structure of morphemes, has a close relationship with vocabulary knowledge. Nation (2000) also claims that bringing learners’ attention to word parts and word formation processes is a useful strategy for learning vocabulary, because they would be more likely to identify affixes and interpret the meaning of the whole word, especially when encountering new ones. Moreover, Nation (2000) states that this knowledge will help reduce the difficulty of learning new words, particularly if the stems are already recognized from other languages. Finally, and, as is going to be explored in the next section, identifying the processes by which words are formed and their internal components is one of the factors that are involved in and promote word knowledge (Nation, 2000).

⁵ Negative prefixes whose spelling depends on the beginning consonant of the root (that is, they are allomorphs of the same morpheme)

2.2.3 Word knowledge

A conclusion derived from the multiple attempts to describe *words* is that they are not independent units with a single dimension. There are many things to know and many degrees of knowing any given word (Nation, 2000), and using the language fluently depends on both knowing plenty of words and much information about them (Willis (2003). A distinction is usually made between two dimensions of word knowledge: size or breadth and depth.

2.2.3.1 Vocabulary size or breadth

Vocabulary breadth refers to the number of words a person knows (Caro & Mendinueta, 2017). According to Meara (1996), the questions that remain regarding vocabulary size is how many words people know, how fast their vocabularies grow and how these factors may influence all other areas of a user's linguistic competence.

In consequence, lack of evidence influences teachers and learners of a language, as there is no universal agreement on the amount and type of items that students of a language should learn. This matter is going to be examined more thoroughly in chapter 4.

2.2.3.2 Vocabulary depth

Knowing a word involves much more than knowing how it is spelt or pronounced; there are multiple dimensions to recognize (Caro & Mendinueta, 2017). Nation (2000, p.40) distinguishes the following features that are involved in knowing a word:

Form		
Which form?	Receptive skills	Productive skills
spoken	What does the word sound like?	How is the word pronounced?
written	What does the word look like?	How is the word written and spelled?
word parts	What parts are recognizable in this word?	What word parts are needed to express the meaning?

Table 1. Word knowledge dimensions: form. Adapted from Nation (2000)

Following the example proposed by Nation (2000, p. 41) of the word *underdeveloped*, knowing the form of the word *bubbly* would imply:

- Recognizing it when it is heard and producing it with correct pronunciation, including the stresses /'bʌbli/.
- Familiarizing with the written form. This involves recognizing it when reading and spelling it correctly when writing.
- Accepting that it is built by the parts *bubble* and *-y*, adding them, and being able to relate these parts to its meaning.

Meaning		
Which aspects?	Receptive skills	Productive skills
form and meaning	What meaning does this word form signal?	What word form can be used to express this meaning?
concept and referents	What is included in the concept?	What items can the concept refer to?
associations	What other words does this make us think of?	What other words could we use instead of this one?

Table 2. Word knowledge dimensions: meaning. Adapted from Nation (2000)

In the example proposed of the word *bubbly*, being familiar with its meaning implies:

- Knowing that *bubbly* signals a particular meaning and being able to produce the word to express it. It can take the form of an adjective, referring to a drink that is full of or produces bubbles, or describe a person as lively and cheerful. On the other hand, it can take the form of a noun to refer to champagne.
- Knowing what the word means in the particular context in which it occurs and adapt to that context to produce it with the intended meaning, either as an adjective referring to an object, a person, or as a noun.

- Knowing that there are related words like *fizzy*, *effervescent* or *energetic*, and being able to produce synonyms and opposites like *still* or *apathetic*.

Use		
Which use?	Receptive skills	Productive skills
grammatical functions	In what patterns does the word occur?	In what patterns must we use this word?
collocations	What words or types of words occur with this one?	What words or types of words must we use with this one?
constraints on use (register, frequency...)	where, when, and how often would we expect to meet this word?	Where, when, and how often can we use this word?

Table 3. Word knowledge dimensions: use. Adapted from Nation (2000)

In the case of the word *bubbly*, being familiar with its meaning implies:

- Recognizing the correct use of the word in a sentence and using it appropriately when producing an original one.
- Being able to recognize that words like *personality*, *water* and *bottle* are typical collocations of the word and producing words that commonly occur with it.
- Knowing that *bubbly* is not an uncommon or pejorative word, and adapting the term to the degree of formality of the situation, knowing that *bubbly* in the form of a noun to refer to champagne is an informal use.

Knowing a word is, consequently, the result of a process that learners have to undergo (Bogaards, 2001). The process implies that before knowing a particular word, the learner has to become familiar with it in different contexts, this author suggests. This means that teachers must ensure that learners are presented with vocabulary in a variety of situations and forms. They must also become aware of their students' current lexical knowledge to provide the best instructional decisions (Caro & Mendinueta, 2017). Finally, as Bogaards (2001) highlights, word knowledge is consolidated over time, and learners are likely to encounter themselves within the spectrum of total

ignorance and total knowledge, which indicates that teachers must also provide them with the strategies that can help them cope with difficult situations.

2.2.4 Lexis and the development of language proficiency

The previous subsection is proof that it is possible to describe, at least theoretically, which elements of a word are necessary for achieving full word knowledge. Nonetheless, how language learners use such words and their lexical competence is an aspect that has not received much attention (Meara, 1996). According to this author, there has been a shift in the paradigm that dominates L2 education, and the communicative approaches have taken over the spot. As a result, lexical competence is regarded as a component of communication, and this has affected the systematic way in which it was studied before.

Under this paradigm, Canale & Swain (1980) define a theory of basic communication skills as “one that emphasizes the minimum level of (mainly oral) communication skills needed to get along, or cope with, the most common second language situations the learner is likely to face” (p.9). This implies that conveying the intended meaning is the main goal, whilst using all the language resources the learner has acquired. Since the focus is on meaning, these authors compare the language acquisition process in the L2 with that in their first language, as the main objective lies more on being understood than on speaking grammar correctly. They suggest that, subsequently, language teachers must assume a similar role if they intend to provide a natural context for communication in their language classroom.

Lexical competence holds for these authors an important role within the communicative paradigm (Canale & Swain, 1980). It has been considered, for a long time, that grammar and lexis were two separate matters, the former dealing with sentences and the latter dealing with words, which learners used to insert in the gaps left by the structures they learnt (Willis, 2003). Nonetheless, it is beginning to be proved that they share a close relationship in which lexis would determine the final shape of the sentence that is produced (Willis, 2003). Barcroft (2004) has added on this by noting that usually, grammar errors do not impede a successful meaning transmission (for example, when a learner forgets to add the third person -s), while vocabulary errors may lead to the incomprehensibility of a particular sentence (for example, a Spanish native saying *perr* instead of *dog*).

There is a general assumption that the lexical knowledge of L2 learners will increase automatically as they increase their proficiency level as well (Zareva et al., 2005). For some learners this may be true, but in the case of the others, their lack of progress in the lexical dimension may hinder their acquisition process. Both the quantity and the quality of the vocabulary knowledge are good indicators that may set the difference between a learner at the intermediate level and another at the advanced level (Zareva et al., 2005). Intermediate learners usually present vocabulary of about 6,000 words, and consequently present few connections among words. Subsequently, those learners with larger vocabularies have richer connections, quantitatively and qualitatively (Zareva et al., 2005). Finally, Zareva et al. (2005) suggest in their study that even though vocabulary size and word knowledge may depend on proficiency development, the knowledge and use of metacognitive abilities and strategies do not, so this must be taken into account for teaching practice.

Within this discussion on the effect of lexical competence on different aspects of language proficiency, August et al. (2005) have contributed by observing that those English language learners that present a slow development of their vocabulary are less skilled at comprehending any text, performing poorly on assessments. Poor comprehension results, partly, from this limitation, as vocabulary is critically important to comprehension (August et al., 2005).

This influence extends to other aspects, like the emotional factors involved in the language learning process. First, if learners are presented with limited knowledge of the L2 vocabulary, it could lead to feelings of frustration and demotivation, as they would not be able to express themselves properly when producing in the target language (Caro & Mendinueta, 2017). Furthermore, a foundation of lexis is necessary to reach higher levels of development in the other basic communication skills (that is, reading, writing, speaking and listening) and therefore its negligence can negatively impact the development of their communicative competence (Caro & Mendinueta, 2017). Willis (2003) also expands on this idea that a lack of vocabulary knowledge can hinder a learner's progress by adding that children's intellectual capacity gradually demands more complex meanings and knowledge as they grow older, so it is necessary to cater the developmental needs of learners as well.

Measuring the importance of vocabulary acquisition in an L2 allows researchers and educators to observe how it influences other aspects of language, with communication as the ultimate goal. During the past decades, teaching and learning lexis has achieved its relevance within the field, but before this, the main idea was that acquiring vocabulary consisted only in learning new words.

2.2.5 The role of lexis in the Spanish Secondary Education curricula

Before analysing the approaches teachers commonly use to teach vocabulary in a second language, it is necessary to establish the framework in which the didactic proposal presented in this paper is based on: the Spanish Secondary Education curricular framework. Understanding the educative background in which the learning unit will be carried out is necessary in order to understand the implications of bringing corpora into the language classroom within this context.

The curricular guidelines for Secondary Education in Spain, divided into the stages of Educación Secundaria Obligatoria (ESO henceforth) and Bachillerato, are established in *Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato* (Ministerio de Educación, Cultura y Deporte. «BOE» núm.52, de 1 de marzo de 2014). This document sets the objectives, competences, abilities and contents that students must reach in each stage and subject. The aim of this decree is to provide learners with the necessary skills to participate in society and access higher education, and being able to express in one or more languages is a key component for it.

The subject of First Foreign Language, which is usually English, is incorporated into the curriculum as a basic subject in learner formation, and it is grounded on The Common European Framework of Reference for Languages (CEFR henceforth). Learners are then expected to be able to apply the acquired knowledge and skills in real interaction processes, with communication as the final purpose.

The subject is divided into four main blocks according to each communicative skill: oral comprehension or listening, oral production or speaking, written comprehension or reading, and written production or writing. Each of these blocks presents the contents, assessment criteria and evaluable learning standards necessary for each stage, that is, first cycle of ESO, fourth grade of ESO, first grade of Bachillerato

and second grade of Bachillerato. As learners face a new grade, they must also meet an increased level of competence in the language, preparing themselves for more complex tasks. Consequently, the criteria in each level vary according to the abilities students are expected to acquire, but in broad terms, the same processes are involved, especially when it comes to lexis.

In every level of the stage of Secondary Education, the amount of lexis, the different aspects of lexis that must be known or the specific lexical items that must be taught are not explicitly stated. Instead, lexical knowledge is included and involved with other processes, depending on each communicative skill.

In the areas related with comprehension (oral and written), learners are expected to be able to understand both the common and more specialized vocabulary on topics ranging different aspects of the students' personal lives (education, activities, events...) and use inferencing meaning as a technique to hypothesize about the purpose of the text. Furthermore, it is included the description of physical and abstract qualities of people, objects, places and procedures. In terms of evaluative criteria, applying the strategies that allow students to understand the general meaning and the details of texts is established. Finally, within the section of syntactic-discursive strategies, the curriculum includes the familiarization with inflections (through verbal conjugations) and expressing time, quantity and manner.

Similarly, in the areas associated with production (oral and written), students must be able to use the common and more specialized vocabulary, and compensate their gaps using linguistic, paralinguistic and paratextual features as a guide. The same occurs with the syntactic-discursive strategies included. However, in this case, they are also expected to modify words that share similar meanings and search for those expressions that may ease the communicative situation, like prefabricated language, to convey the same meaning. The description of physical and abstract qualities is included as well. In the evaluative criteria, it is stated that texts should be composed of frequent vocabulary (common and specific), which should be adapted to the context of the communicative situation in terms of register and adequacy. Lastly, in the upper stages of Secondary Education, students are also expected to use the target language for humoristic or aesthetic purposes.

Guidance about how to proceed with the teaching of lexis is not explicitly stated, except when setting the common topics of vocabulary and the fact that it must be recognized and used properly. Although in this curriculum it is embedded within other competences, lexis still holds great importance in the communicative situation, as all tasks involving meaning, comprehension, or inferring place a lot of weight on lexis. Furthermore, those tasks that require using adjectives, adverbs or verb conjugations are involved with morphology. Finally, it should be noted that word usage and cultural aspects of the word are involved in tasks in which selecting the appropriate term and adapting to the communicative context is necessary. As can be observed, diverse parameters of word knowledge and word formation are important if the objective is to communicate the intended meaning with ease.

The fact that these contents are not explicit gives teachers the freedom to select those contents that they consider necessary and adapt them to the needs of their students. However, this can also be a problem. In many cases, leaving the choice to educators and creators of educative material may lead to a wide difference in lexis knowledge among groups of students. For instance, one teacher may consider studying affixation necessary while others may not. As a result, the amount, knowledge of different word aspects and the strategies students know and use to cope with gaps in their vocabulary may vary greatly in this stage of language learning in which lexical richness should increase.

2.3 Teaching and learning vocabulary

2.3.1 Which words should we teach?

As was examined in subsection 2.2.2, determining what words are and counting them is subject to the criteria of the counter. The same happens when deciding how much and which vocabulary should be taught to English learners, and the first question that arises is how many words there are in the English language and whether learners should aim to know the whole range. Nation & Waring (1997) turned to the largest existing dictionaries to answer this question, and after excluding some items like abbreviations, proper names or alternative spellings, they concluded that there existed around 54,000 word families, including base words, their inflections and derivations. Setting the goal of learning all the words of the language, they consider, is beyond what second language learners and most native speakers could achieve.

It is not uncommon that teachers, then, turn to other standards. Nation & Waring (1997) claim that a vocabulary size of 2,000 to 3,000 words is a good basis for language use for teenagers. But even if the learner acquired a large number of words, not all of them are equally useful in language use. The most common criterion to select words according to their usefulness is frequency, and Nation (2000) distinguishes four kinds of vocabulary in a text based on their occurrence:

- High-frequency words: They take about 80% of the running words in a text. They occur in all uses of the language.
- Specialised vocabulary: Nation (2000) distinguishes two types:
 - a) Academic words: Words common in academic texts.
 - b) Technical words: They cover about 5% of the running words in a text. They are words closely related to the subject area the text deals with. They are those that are common in a particular topic, but not elsewhere. High-frequency words with specialized meanings also fall under this category.
- Low-frequency words: They take 5% of the words in an academic text and a small proportion of any text. They are the largest group of words, and they consist of technical words for other subjects, proper nouns, and those that occur rather infrequently.

Nation & Waring (1997) highlight the feasibility of placing high-frequency words at the top of the priority list in vocabulary teaching by claiming that “if a learner knows these words, that learner will know a very large proportion of the running words in a written or spoken text” (p.6). Nation (2000) argues that both teachers and learners should spend considerable time on these words. Specialized vocabulary should be treated like high-frequency words (Nation, 2000). Nonetheless, much of the technical uses and meanings will make sense only in the context of the subject matter that is being studied, and those connections and variances in meaning should be paid attention to. Finally, and on the opposite side, Nation (2000) considers that low-frequency vocabulary should not be devoted to large amounts of practice time. This author claims that teachers should best concentrate on training learners in the use of strategies to deal with this type of vocabulary, like using dictionaries, guessing from a particular context, or focusing on word parts. These strategies, Nation (2000) states, will allow learners to continue increasing their vocabulary.

2.3.2 How do we learn vocabulary?

Learning new words and having a vast vocabulary is considered a crucial process in mastering a language, but the attitude towards vocabulary acquisition research has not received the same treatment by researchers. Although research on this area has increased during the last decade, vocabulary is studied as an isolated matter apart from mainstream Second Language Acquisition (SLA henceforth) theories. As a result, there is no unified model of how vocabulary is acquired in a foreign language, so scholars resort to extrapolating from what is already known, that is, their knowledge about general SLA theory and the organization of the mental lexicon in the first language (Haastrup & Henriksen, 2000).

2.3.2.1 How does passive vocabulary turn into active?

A matter that has generally been agreed upon is the fact that learners of a second language tend to know more words than they can actually use (Fan, 2000). Nation (2000) distinguishes for this purpose between receptive and productive vocabulary. Receptive vocabulary would deal with receiving language input through listening or reading and trying to comprehend it, whilst productive vocabulary would be involved in expressing meaning through speaking or writing, an act in which the speaker would need to retrieve and produce the appropriate word form. In some cases, the terms “passive” would refer to receptive vocabulary (listening and reading) and “active” would refer to productive vocabulary (speaking and writing) (Nation, 2000). Fan (2000) claims that students who show a higher proficiency level in passive vocabulary are thus more proficient in active vocabulary. Nation (2000) supports this by claiming that broadly speaking, receptive learning is easier than productive, although the reasons why this happens are still not clear.

Nation (2000) suggests a three-step model of word integration into the learners’ active vocabulary, which begins by noticing a particular word:

- Noticing: It consists in focusing on an item by arousing the attention of the learners. It involves decontextualizing the word through meaning negotiation or providing clear definitions.

▪ Retrieval: It involves strengthening the memory of a word by retrieving its meaning during a task. Such memory depends on recalling the previous meeting, which is determined by two factors:

a) The learner's vocabulary size: a learner with a large vocabulary needs to process larger quantities of language to meet a new word again, which occur less frequently.

b) The length of time word memory lasts: the learner must perceive that the word, in fact, repeats. In general terms, word memory can last for weeks.

▪ Creative or generative use: It occurs when words that had been met previously are used in new, different ways from the previous meetings. Consequently, the learner reconceptualises the knowledge of that word.

Finally, Ellis (1994) agrees that those who read more know more vocabulary, and that vocabulary acquisition is affected by reading. He claims that it is the ideal environment to acquire new words, especially those that are less frequent, as they are more likely to appear in print than in common speech.

2.3.2.2 The role of the mental lexicon

There are many indicators that words are likely to be well arranged in the mind, like the high number of words native speakers know or the fact that they can be recognized and located at a fast speed (Aitchison (2012)). This organization system is conceived as the mental lexicon, which is defined by Aitchison (2012) as a "human word-store" or "mental dictionary" (p.3). Although the mental lexicon and dictionaries have been related metaphorically, Aitchison (2012) claims that there are more differences than similarities between them. For instance, she exemplifies that words are not arranged alphabetically in the mental lexicon, because otherwise, speakers would choose an alphabetically adjacent term when making a mistake. Consequently, there is evidence that words are arranged in the human mind, but not in the same fashion as in a dictionary.

In the field of second language vocabulary acquisition, researchers have tried to devise a universal model for the structure of the L2 mental lexicon and analyzed in which ways it is related to that of the L1 (Wolter, 2001). Ameel et al. (2009)

differentiate two types of hypotheses that have been developed over this matter. The two-pattern hypothesis distinguishes two different sets of word forms and references for each language, attributed to monolinguals. On the other hand, the one-pattern hypothesis holds that bilinguals do not have two separate mappings of words. Instead, there is pattern interaction between the two languages, and are not isolated from one another. While researching this matter, it is usual to consider in which aspects the L1 and L2 mental lexicons are similar to each other, and in which ways they separate and create concept mappings from referents derived directly from the real world (Pavlenko, 2009).

Most models conclude that the phonological and morphosyntactic components differ from one language to another, while meanings and concepts are shared in all of them, at least partially (Pavlenko, 2009). In consequence, Pavlenko (2009) highlights that, because of this, bilingual speakers can translate most words from one language to another. Furthermore, if the L1 and the L2 mental lexicons worked as two completely separate systems, learners would have to comprehend the differences between the naming patterns for an object in the two languages, acquire these patterns, and maintain them separately over time, to achieve a nativelike proficiency in both languages (Ameel et al., 2009). This is quite impractical, as it would not be possible to avoid that there are interconnections between concepts and keep two separate patterns of word forms and referents (Ameel et al., 2009). Dong et al. (2005) also support this view by claiming that vocabulary is stored in almost the same brain area for both languages.

For Pavlenko (2009), it is a matter of conceptual equivalence in the L1 and the L2. She claims that in the early stages of L2 learning, students may resort to acquiring explicit definitions of a particular word in their L1. If the concept is equivalent in both languages, the word will be linked to an already existing linguistic category. If, on the contrary, the concept is not equivalent or only partially equivalent, learners will be led to inaccurate performance when trying to use the word in context. Consequently, this author supports the notion that vocabulary teaching should bear these differences in mind along with the proficiency level of learners, and suggests following this approach, depending on their degree of equivalence L1-L2:

- a) Concepts are equivalent: Instruction should be oriented to creating stronger links between words and their translation through production and recalling tasks. Focusing on metaphorical uses at more advanced levels is also suggested.
- b) Concepts are partially equivalent or non-equivalent: Instruction should highlight similarities and differences in both languages through awareness-raising activities, and how such concepts are referred to by native speakers. Language corpora are a good tool for this purpose.
- c) Concepts are not equivalent: Instruction should aid in developing new concepts through tasks that help learners familiarize with them in both languages.

These insights have implications for second language instruction. As Pavlenko (2009) points out, it is common that materials aimed for L2 learners overlook the differences in meaning, relying on the basis that concepts are always equivalent across languages.

2.3.2.3 The conscious vs unconscious process debate

Another topic that is generally agreed upon is that the outcome of vocabulary acquisition must be to be able to comprehend and produce words in different contexts of communication quickly. In order to do that, it is necessary to form a steady cognitive representation of the word that is easily accessible in such a variety of contexts (Schwartz et al., 2008). Nevertheless, how these representations are created is not a simple matter, and researchers' positions on how vocabulary is acquired often range from those that support that learners acquire vocabulary unconsciously to those that hold that learners should be taught vocabulary explicitly and consciously (Ellis, 1994).

Krashen's Input Hypothesis (Krashen & Terrell, 1983) is considered the main representative of the unconscious position on vocabulary acquisition. Through this hypothesis, Krashen & Terrell (1983) state that students will progress more easily to the next stage of learning by understanding language input with structures situated on the next level with the aid of contextual and extra-linguistic information. They suggest focusing on oral and written comprehension, and production will emerge on its own when the learner has developed enough competence. Therefore, there is no need to teach speaking or writing explicitly. Consequently, Krashen (1989) claims that vocabulary is best acquired in the same manner, through comprehensible input, and more specifically, in the form of reading. According to Krashen & Terrell(1983),

children who report more free voluntary reading tend to perform better on vocabulary tests, and those whose environments are print-rich also have better vocabularies. Laufer (1991) supports Krashen's theory by stating that even though the acquisition of language in the L1 and the L2 is not exactly equal, they share some similarities, and she claims that if natives can learn new words through mere exposure, foreign learners might learn new vocabulary in the same way, especially if that exposure comes from reading.

Krashen & Terrell (1983) mention as well the Skill-Building Hypothesis (SBH) to support their claim that most language learning takes place unconsciously. According to the SBH, we learn new rules or items that gradually end up becoming automatic, but first we must learn them consciously and practice them through drills and exercises. This means that vocabulary is learnt in the same manner, by learning words one at a time, paying attention to morphology and practising through exercises. However conscious language learning does not seem to be as efficient as acquiring language from the input (Krashen et al., 1983).

Despite these theories, current research does not allow proving whether learning words in certain contexts, like reading, are a reflection of implicit, incidental or a case of explicit learning without instruction (Ellis, 1994).

Although these theories are a brief approximation to the different models and suggestions that researchers have contributed to the matter of vocabulary acquisition, it can be noted that it is a complex process in which both conscious attitudes and unconscious processes are involved. In addition to these, some factors have captured the attention of researchers, like the easiness or difficulty of learning a word in the L2, and no theory accounting for second language vocabulary acquisition is complete without referring to these factors, which are explained more profoundly in the following subsection.

2.3.3 Approaches for teaching vocabulary

L1 speakers succeed in mastering their language if they are exposed to enough amounts of input and they do not suffer from physical or mental impairments that prevent them from doing so. However, L2 learners show different attainment levels even after being exposed to the language for many years (Hulstijn, 2005). One possible explanation for

this is that such differences in success in the acquisition may rely on cognitive psychology and the study of implicit and explicit learning (Ellis, 2009).

For this reason, it has been studied whether adults can acquire a language following the same implicit mechanisms employed by children learning their native language and whether grammar is best taught explicitly (DeKeyser, 2003). Other investigators, on the other hand, have tried to examine which processes are involved in implicit and explicit learning, which aspects they share in common and how they can be influenced through instruction (Ellis, 2009). Understanding which aspects of L2 domains are influenced by either type of learning plays a relevant role for all language education professionals and provides a clue about how L2 learners may benefit from both.

During the last decades, researchers in cognitive psychology and SLA have attempted to define what implicit and explicit language implies, but no consensus has been reached on their definitions, especially on implicit learning. Ellis (2011) defines implicit learning as the acquisition of knowledge “by a process which takes place naturally, simply and without conscious operations” (p. 38). This author compares the process of acquiring knowledge about a language to that by which we acquire our L1: without being aware of the fact that we are acquiring the rules and mechanisms involved in the language. Nevertheless, there is controversy around the notion of awareness and what is meant by it. Authors like DeKeyser (2003) have defined implicit learning as “learning without awareness of what is being learnt” (p.314), that is, without reflecting upon the language content.

This kind of instruction, according to Ellis (2009), consists in providing learners with samples of a rule which they would internalize, as their attention would be drawn to meaning instead of the pattern. Rules, therefore, would be inferred without awareness, by masking the learning target. This notion has been often equalled with “incidental learning”, that is, the mode of learning in which information is picked up unintentionally (Hulstijn, 2005). Nonetheless, the efficiency of this type of instruction is often debated. Some authors, like Ellis (2015), consider that although L2 learning takes place in a language-rich environment, not all of the information is grasped and is, therefore, much less successful than L1 acquisition.

Contrariwise, explicit learning has been defined by Ellis (2015) as a “conscious operation where the individual makes and tests hypotheses in a search for structure” (p.3). That is, learners intentionally look for the rules that create the language. It takes place consciously and intentionally, and learners are aware that they are learning something and can verbalize it (Ellis, 2009). Explicit learning is easier to measure, and, according to Ellis (2015), this type of instruction can speed the process of language acquisition, being more effective than the implicit type. Furthermore, by explicitly learning a language, the different formulas, rules and drills create language productions that, used subsequently, promote implicit learning and automatization, considering that the two types intervene in the process of acquiring a language (Ellis, 2015).

It may be concluded, therefore, that the ability of a learner to produce language systematically is the result of the interaction of both conscious and unconscious learning processes (Ellis, 2015). These methods have been classified in two ways, distinguishing deductive and inductive instruction (DeKeyser, 2003).

. Once this distinction has been explored, the focus of this study will turn to explicit deductive and inductive methods, which are the backbone of the learning unit presented in this paper, as students are encouraged to pay attention to particular language features to develop metalinguistic awareness.

The difference between inductive and deductive learning resides at the moment in which such rule is presented. Inductive learning, according to Hulstijn (2005), takes place before rules are presented. It is defined by Mallia (2014) as a bottom-up approach in which rules are not given, but instead, learners induce rules from language in use. This author states that students discover the target language and its rules by themselves thanks to previously selected materials that illustrate the use of a particular feature of the target language.

Deductive learning, on the other hand, consists on providing specific language rules at the beginning of a lesson, which are then demonstrated and practised (Mallia, 2014). In this top-bottom approach, rules are dictated and then the particular examples are given (Alzu’bi, 2015). About both approaches, Gollin (1998) points out that while inductive reasoning implies inferencing general facts from particular examples, deductive reasoning deals with applying general rules to particular examples.

The difference between both techniques is how the goal is reached. For Seliger (1975), as the result is the same, the main concern should be the efficiency of the approaches. Acknowledging their benefits and constraints will allow teachers to choose the most convenient according to the learning situation.

On the favourable side of the inductive approach, which is more student-centred, it helps learners in becoming more involved in their learning process and in engaging actively in the lesson (Alzu'bi, 2015). Furthermore, Alzu'bi (2015) points out that it increases learners' motivation, confidence and enthusiasm towards language learning. However, Seliger (1975) claims that with this method it cannot be guaranteed that a learner will induce a concept correctly or that the student has actually discovered such rule. In the case of the deductive approach, learners tend to feel more comfortable when learning with this methodology (Mallia, 2014), and it provides more certainty of the grammatical knowledge acquired (Fischer, 1979). Nevertheless, many language teachers avoid expressing support for this method as it is associated with the negative criticisms aimed towards the grammar-translation method and previous generations of language teaching even though it may be more effective (Seliger, 1975).

Controversy still exists for both approaches, and the common ground for both methods is that it requires from students to be mentally active, which leads to increased motivation and more thorough learning (Gollin, 1998). Further, teachers may switch approaches in their lessons, depending on what is needed at the moment: if memorization and comprehension is the priority, students are more likely to remember those features they have worked out for themselves, but if the priority is time or less intricacy, a deductive approach would be more suitable (Gollin, 1998).

2.3.4 Learning vocabulary through electronic corpora and Data-Driven Learning

Current society is characterized by people's immediate access to several resources in which ICTs play an important role. The influence of digital tools has permeated in the same manner in language teaching and learning. The improvements in the quality of tools, software and connectivity have led to the creation of new approaches towards vocabulary learning, and have, at the same time, optimized the already existing.

Computer-assisted language learning (CALL) is one of the areas that experienced rapid growth with this technology development. Although not universally

popular among teachers and learners (Murphy, 1996), it paid special attention to vocabulary learning, with early programs including activities for this purpose like gap-filling, vocabulary games or text reconstruction (Ma & Kelly, 2006). Nowadays, online resources like apps or social networks have taken the lead. For instance, learners have access to many digital reference tools, like *Lingua.ly*; games for testing vocabulary, like *Quizlet* or *Kahoot!*; social networks for interaction like *Facebook* or *Twitter*; communication software for conferences like *Skype*, or language learning programs, like *Busuu* or *Duolingo* (Elgort, 2018). The choices available have been increased significantly, and, as was explored in subsection 2.1, a central key in this issue has been the ability of computers to store and process large amounts of language data efficiently. The study of language, and especially vocabulary, through concordances or corpus linguistics plays thus an important role in this discussion.

Although the efficiency of corpora or concordance for vocabulary learning has been a disputed issue, they offer a wide spectrum of possibilities for analysing vocabulary. First, the data presented by corpora fit communicative teaching approaches, as they represent real language used in authentic contexts of communicative situations (Murphy, 1996). Further, as psycholinguistic research has proven, language processing is sensitive to the frequency of usage and statistical knowledge (Ellis, 2015), and corpora may be helpful indicating which forms occur more frequently in a variety of contexts. This is another feature that benefits vocabulary learning, as Ma et al. (2006) claim, because through corpora, vocabulary is accessed in context instead of presenting isolated words. Thanks to this, by analysing patterns through options like Keywords in Context or KWICs, the learner might be able to observe facts about terms not easily accessible otherwise, like semantic relations, conceptual fields and collocations (Murphy, 1996).

Access to millions of words in a variety of genres and formats has created new ways for teachers and learners to explore real patterns of vocabulary use (Elgort, 2018). Nevertheless, learning through corpora has brought in new troubles for teachers to be aware of. The action of merely presenting learners with corpora and language data, as happens with dictionaries, does not guarantee the acquisition of knowledge (Boulton, 2009). The direct application of corpora in the classroom entails the need for guidance in their use, usually presented through lessons before working with the tools and a more practical session to explore the functionalities of these tools (Pérez-Paredes et al., 2011).

However, in general, discovery learning through corpora usually leads towards more self-managed study (Murphy, 1996). This potential has drawn increasing attention in the past few years in the form of Data-Driven Learning (DDL henceforth) (Boulton, 2009).

DDL is defined by its coiner, Johns (1991), as a computer-based approach to language learning in which the students “discover the foreign language”. According to Boulton (2009), DDL is based on the premise that learners discover patterns on their own when examining naturally-occurring language. Johns (1991) describes that the role of the teacher is relegated to fostering an environment in which the learner can develop strategies for discovering such patterns. According to this author, the language learner is the protagonist and would thus turn into a researcher, deriving knowledge through access to linguistic data, a notion that would name this approach.

At the core of the approach, the computer and corpora would perform as the informant, according to Johns (1991). The role of the concordance is not to provide answers about the language per se, but to provide data so that learners can infer such knowledge by making sense of the data produced while integrating it with what they already know. Johns (1991) considers that, by this approach, “we simply provide the evidence needed to answer the learner’s questions, and rely on the learner’s intelligence to find answers”. As may be examined, DDL is considered to offer advantages like increasing awareness about the language, improve the ability to manipulate it, offer authentic language data, or fostering learner autonomy (Boulton, 2009). It is this last advantage that is pondered more beneficial, as it is considered that by allowing learners to engage directly with the evidence, speculation and enquiry are stimulated, allowing learners to generalize from particular instances of the target language in use (Johns, 1991). These characteristics that allow moving from data to generalization can be of special value in the process of language learning and vocabulary learning paying attention to morphology because, as was examined in previous chapters, there is some regularity to be found in word formation processes.

However, Lee et al. (2019) point out some limitations of this approach, like the fact that it might be costly in terms of time, because students who are less accustomed with inductive learning methods may require great amounts of time to make inferences. Breyer (2009) also claims that teachers are left with some challenges they need to overcome for the method to succeed, like considering which materials are appropriate

for the learners, their proficiency level and how these materials can be integrated into the curriculum. Furthermore, this author holds that teachers need to present certain degrees of literacy in working with corpora, so that learners are able to fully implement this type of learning. Another issue, according to Ma et al. (2006), is that the amount of autonomy assigned to learners might be unfavourable: too much freedom may affect the learning outcome. As Römer (2011) argues, even the complexity of the data shown may intimidate learners, especially those who still show a limited vocabulary.

To address the matter of working with an inductive approach which may discourage beginner and more teacher-centred students, Lee et al. (2019) suggest combining DDL with existing or more traditional teaching approaches to reduce the cognitive load involved. This author claims that both inductive and deductive approaches entail different methods of reasoning, worth applying through DDL. Further, research conducted by Lee et al. (2019) showed that both approaches are equally effective in promoting vocabulary acquisition and retention, and, despite the criticism on this issue, Boulton (2009) claims that DDL could benefit both advanced learners trained in corpora as well as intermediate students.

What may be concluded is that these resources have exerted an influence on language education which cannot be ignored, and bringing them into the classroom is a practice that teachers must consider if the ultimate goal is to facilitate learning in an environment that accommodates students' needs.

3. Didactic proposal

3.1 Context

This learning unit is addressed to a group of 16 students of the subject of English as First Foreign Language in the educative stage of 2nd grade of Bachillerato (Secondary Education) in a state high school in Spain with a bilingual program. They are 17 years old.

They have been enrolled in a bilingual program since the compulsory stage of E.S.O. (Educación Secundaria Obligatoria), and they have studied English as a Foreign Language lessons since the stage of Primary. Since Primary, they all have been in contact with conversation assistants from different countries, and most of them have

participated in abroad programs offered by the school or their extracurricular language centres. Overall, they all have been in contact with different variants of the language. Furthermore, as they have been enrolled in the bilingual program, they have a solid foundation of classroom language and engage easily in conversation with others whenever the tasks require so.

Taking into account that individual differences and learning styles are present, their linguistic level at present ranges between intermediate B1-B2 based on the CEFR standards.

As English as Foreign Language is a compulsory subject at this stage, students belong to different modalities of Bachillerato (Science, Humanities and Social Science or Arts), but it is a cohesive group; most of them know each other since earlier stages and new students have been easily integrated. Most of them plan on taking part in exams to access university, as well as on continuing studying English or take official exams to obtain official language certifications.

Students at this stage take three sessions of 60 minutes (three hours in total) per week of English as a Foreign Language, which is considered a general and compulsory subject. The total amount of instruction they receive each week is 30 hours, thus the subject of English as Foreign Language comprehends the 10% of the weekly study load.

The learning unit will be carried out throughout one scholar term (September through December) and will consist of 14 sessions, distributed in the following way:

- 2 sessions of corpora training.
- 11 practice sessions, arranged by content:
 - Three sessions on affixation.
 - Three sessions on word derivation.
 - One session on zero derivation and stress shift.
 - One session on acronyms and onomatopoeias.
 - One session on coinage and loans.
 - One session on clipping and backformation.
- One session on compounds and blends.
- One last session for content review.

This learning unit will be carried out a session per week, during the time assigned to work in the ICTs room, and it will have a study load of a 30% of the subject content of the term. Each session will last 60 minutes.

Regarding the materials available, the educative centre has an ICTs room equipped with 25 computers. Furthermore, tablets and laptops are available for student loaning at the school library, in case students need them for personal study. All sessions will be carried out at the centre, so no extracurricular time is needed to complete the activities in the learning unit. Nevertheless, students are encouraged to practice on their own and research.

Students have taken part in language studies through ICTs in previous courses, with at least an hour per week being devoted to this type of learning. They have worked with and are acquainted with online reference tools, like dictionaries or thesauri; they usually work with text processors and engage in activities involving multimedia, apps or games, as well as document-sharing and cloud storing services.

Nevertheless, students have never approached electronic corpora and have never encountered the concept of corpus linguistics. This learning unit will be their first encounter with the concept and the tools.

3.2 Methodology

Activities have been designed for students to work individually, in small groups, and to participate in whole-group discussions.

The learning unit will be carried out in the ICTs classroom of the educative centre, so students will perform all their work with a computer. For this reason, they will use different webpages and applications, text processors, reference tools suggested by the teacher (like dictionaries), and the electronic corpora. As it is the first time that learners use electronic corpora, two sessions will be devoted, at the beginning of the learning unit, to ensure that the concept of corpus linguistics is grasped and that students acquire basic knowledge about corpora searches and become familiar with language data analysis. In this case, the training received will be operated on the electronic corpora BNC and COCA.

These two corpora have been selected because of their simple and easy interface, the fact that they represent a large amount of authentic native speaker data and their free access. Both corpora have been selected to represent different varieties of English so students can critically analyse language based on parameters like usage, form or adequacy.

The texts used in this unit (which have been previously selected by the teacher), are presented both through oral and written mediums, and the activities have been designed so that at least more than one language skill is worked in each. Further, the texts have been selected so that authentic language input is provided to the students in a varied and rich way, including texts from different genres. This unit allows working in all competences established for this educative stage as well. On the other hand, these texts have been selected according to their genre, vocabulary variety, and features of interest, size and difficulty. After this, these texts have been examined using the tool Text Inspector, which provides information about text content in aspects like word frequency (based on corpora), lexical diversity or metadiscourse tagging. The purpose of this analysis is to ensure that the texts are rich in terms of language content, adequate for the language level of the students and suitable in terms of size.

The teacher will, at times, step back on their role of traditional instructor, and will act as a guide for students in their use of corpora. The teacher will be in charge of managing timing in the classroom, confirming the rules examined and directing the group debates. Furthermore, the teacher will aid students that may need it in their corpora searches.

Concerning the pedagogical approach, it is aimed towards a combination of Data-Driven Learning inductive learning and more traditional approaches. Through DDL and inductive work, students become protagonists of their learning, they become aware of the language feature studied and it enhances their autonomy while promoting task engagement. Through traditional work, students that are more accustomed to teacher-oriented methods will feel more comfortable, while reducing the difficulty and less positive aspects involved in inductive learning. This way, learners will benefit from both methods. The fact that students may not have enough experience with this type of learning has been taken into account, thus, in order to aid them, activities have been designed so that they have enough support to carry them out.

On the other hand, students will work on their portfolio in every session, which will serve as an assessment tool and as a classroom journal in which they will record the key elements to remember, the features studied and their conclusions reached, their reflections and thoughts on the lesson and attitudes. They will be able to display, with the rest of the educative centre and families, the work they have done throughout this unit; it will be a useful tool for the teacher to analyse student development, difficulties and needs, and it will be useful for students, as it may help them during the development of the sessions.

3.3 Competences and aims

3.3.1 Competences

The competences for the educative stages of Secondary Education and Bachillerato, as established in Article 3 “Curricular competences” in *Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato* (Ministerio de Educación, Cultura y Deporte. «BOE» núm.52, de 1 de marzo de 2014). are the following:

- a) Linguistic communication.
- b) Mathematical competence and basic competences in science and technology.
- c) Digital competence.
- d) Learning to learn competence.
- e) Social and civic competences.
- f) Sense of initiative and entrepreneurial spirit.
- g) Cultural awareness and expressions.

Through the development of this learning unit, students will become involved in all of them in the following manner:

- a) Linguistic communication: This learning unit is integrated with the curricular guidelines stipulated for the development of the first foreign language (English), and its ultimate goal is to help students learn vocabulary, which is essential for communication and interaction. Furthermore, the four basic communicative skills (listening, speaking, reading, and writing) are worked through this unit, and students will engage in communicative situations and activities that favour interaction among classmates and

with the teacher. Finally, different types of texts in diverse modalities and formats are presented, so that the language stimuli received are authentic and varied.

b) Mathematical competence and basic competences in science and technology: Language is, in this unit, accessed through language corpora in the form of language data, which implies that students will need to apply reason and logical thinking to be able to observe, describe, and interpret the data they observe. They will also need to infer, create hypotheses and prove them, following an inductive approach to language that is related to the scientific method of hypothesis making and testing.

c) Digital competence: Learners will work with electronic corpora to access Information and Communication Technologies (ICTs) and to integrate these tools in their language learning process as a means to access language in authentic contexts and as reference tools for consultation.

d) Learning to learn competence: For the development of this learning unit, learners will become more autonomous in their learning process, as the teacher will act as a guide. This will help them become protagonists and will provide them with a high level of responsibility in organizing their task and time management when working in the classroom. Furthermore, by learning to establish connections with other word aspects and by paying attention to morphology as a means to help decipher meaning, learners are acquiring strategies that will help them in the future in their vocabulary acquisition process.

e) Social and civic competences: This unit involves working collaboratively with other classmates to participate in activities, exchange ideas and solve problems. Furthermore, it involves engaging in debates, sharing opinions, valuing those of others and learning to respect them.

f) Sense of initiative and entrepreneurial spirit: Problems and situations to be solved imply that students will have to become aware of the matter, plan and manage their knowledge. They will also need to consider the steps necessary to solve such problems, so they can achieve the desired objective and reach a solution.

g) Cultural awareness and expressions: In this unit, and through the study of the English language, students will acquire knowledge about different cultures while learning to

understand, value and appreciate such cultures, throughout cultural manifestations like literature. Furthermore, they will be able to appreciate the similarities and differences between these cultures and that of their own, with a critical eye and respect.

3.3.2 Aims

Learning unit general aim

- To identify and apply the different processes by which words are formed with the help of electronic corpora, and to establish links with other dimensions of words to promote word knowledge.

Specific aims

- To become acquainted with corpus-based language analysis and use the COCA and BNC electronic corpora as both linguistic data sources and reference tools.
- To examine different word formation processes in authentic language contexts and pay attention to their social and cultural aspects, like register, adequacy and language variant.
- To practice the different processes of word formation through different combinations of the four language skills: listening, speaking, reading and writing.
- To participate in class discussions with the teacher and other classmates about the studied language feature, sharing the own findings, hypotheses, ideas or opinions.
- Sessions on corpus training: To become acquainted with the concept of corpora and corpus-based language analysis, and to acknowledge the basic search functions in BNC and COCA.
- To recognize common roots and affixes in English, their connotations and how they affect word meaning.
- To distinguish affixes and how they combine in order to create new words and change their class.
- To know and identify different processes of word formation and apply the same criteria for the suggestion of new terms.

- Portfolio work: To record their own practical activities, attitudes and contents studied during each session in order to reflect on the learning process and as a tool to display personal work.

3.4 Resources

In the following subsections, the materials necessary for the development of this learning unit are described. Further, the corpora selected for working in the classroom will be examined as well along with the criteria employed in this choice.

3.4.1 Material resources

The use of ICTs is essential for the correct development of these activities, as consultation through electronic corpora require available devices and Internet access. Despite this, the choice concerning the devices may depend on the resources available at the centre, as corpora can be accessed through computers, tablets or smartphones.

In case no Internet connection is available, the teacher can adapt activities by providing students with result lists extracted and printed out from the corpora.

In sum, the material required will be:

- 16 computers or the other electronic devices mentioned above.
- A projector.
- Corpus of Contemporary American English (COCA) and British National Corpus (BNC).
- Activity work pages and game cards (see appendices II-IV).
- Portfolio work page (see appendix III).

3.4.2 Corpora

After considering the different types of English corpora available on the web, the corpora chosen for the learning unit are the Corpus of Contemporary American English (COCA) by Mark Davies and the British National Corpus (BNC) by Oxford University Press.

The Corpus of Contemporary American English (COCA) is a native speaker corpus available at <https://www.english-corpora.org/coca/>. It currently contains more than one billion words. Since its publication in 1990, these numbers have kept increasing yearly. It is estimated that it contains over 485,000 texts belonging to a variety of genres, including blogs, webpages and TV and movie subtitles.

The British National Corpus (BNC) available at <https://www.english-corpora.org/bnc/> represents a wide sample of British English from the later 20th century. It currently contains nearly 100 million words. These belong to a wide variety of texts, arranged in different genres. It shares the same interface as COCA, with the exception of some characteristics. The genres of non-academic and miscellaneous are not included in the American Corpus. On the other hand, this corpus does not include the recently added sections of television and movie subtitles, blogs and websites.

The reasons why these corpora have been chosen are the following:

- They are free and available online (only previous registration through email is required).
- Their interface is simple and user-friendly.
- Multiple search options are available.
- Allow performing advanced searches with wildcards (*), part of speech (*PoS*) and keywords (*Key Word in Context or KWIC*).
- They are large and updated corpora.
- Instructions and help are available.
- Allow using both to compare English variant differences.

Because their interface and options are simple but appropriate for this work, their functions adapt to what the activities in the didactic proposal require.

They allow performing through five search options: *list*, *chart*, *collocates*, *compare* and *KWIC*. In all of them the *PoS (Part of Speech)* or grammatical category can be specified.

- *List*: it shows the frequency and a list of the contexts in which the word/phrase appears, allowing one to examine each one.

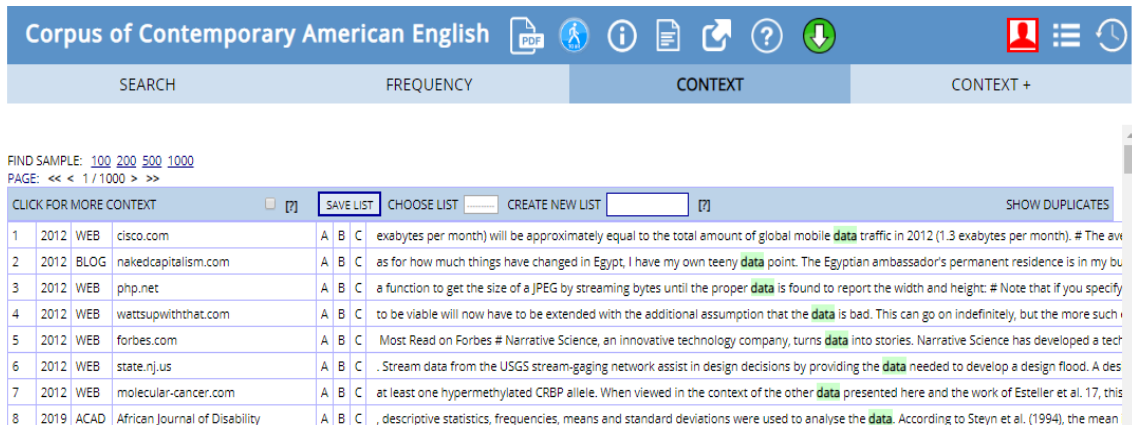


Figure 1. Example of a List search in COCA

- *Chart*: Performs a search of the term and allows comparing their frequency in each genre section.

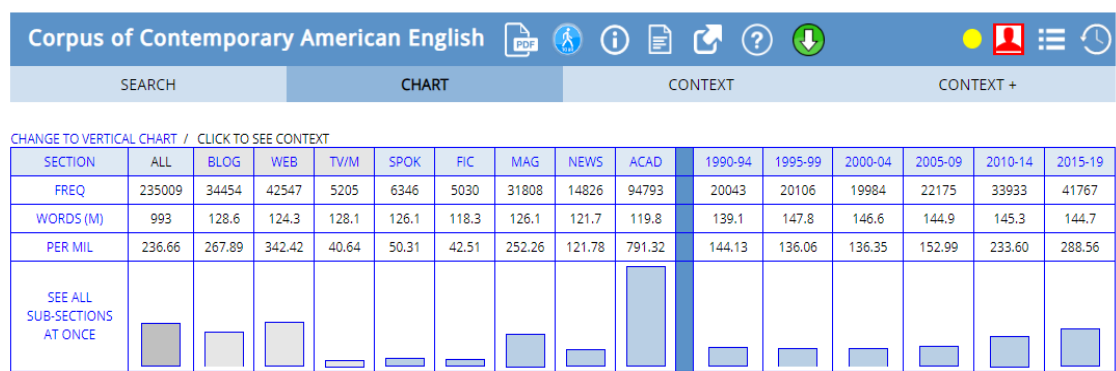


Figure 2. Example of a Chart search in COCA

- *Collocates*: Allows observing which words occur more frequently next to another. The corpus offers the option to introduce the word or phrase of interest paying attention to a part of speech in particular, and to search the collocates according to the part of speech they present as well.

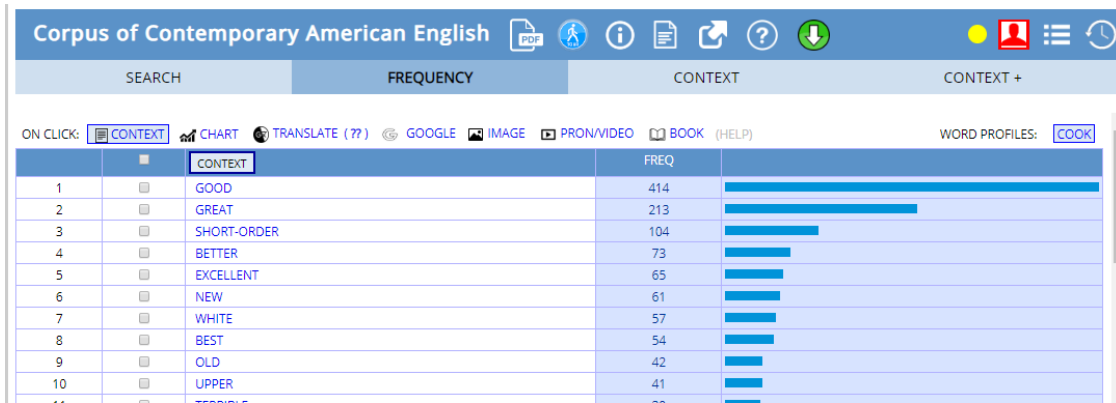


Figure 3. Example of a Collocates search in COCA

- *Compare*: Allows comparing two terms to identify a pattern of occurrence.



Figure 4. Example of a Compare search in COCA

- *Keyword in context (KWIC)*: Shows the patterns of occurrence in a word by sorting them to the left and/or right. Each word in the text is labelled with a colour code:

Noun: cyan

Verb: pink

Pronoun: blue

Adverb: orange

Proper noun: light blue

Preposition: yellow

Adjective: green

Corpus of Contemporary American English

SEARCH FREQUENCY **CONTEXT** CONTEXT +

L - - - 1 2 3 R + RE-SORT ?

CLICK FOR MORE CONTEXT [?] SAVE LIST CHOOSE LIST CREATE NEW LIST [?] SHOW DUPLICATES

1	2001	MAG	Redbook	A	B	C	Add tomatoes , black beans , hominy , and broth ;	cook	10	minutes	Stir	in ciantro . Sprinkle each serving with 1
2	2009	MAG	GoodHousekeeping	A	B	C	olive oil on medium 1 minute . Add salsa onion and	cook	15	minutes	or	until very soft , stirring occasionally . Stir in
3	2007	MAG	GoodHousekeeping	A	B	C	golden . 4 . Stir in broccoli rabe ; cover and	cook	5	minutes	or	until broccoli rabe begins to wilt , stirring
4	2011	MAG	GoodHousekeeping	A	B	C	shimmers and is almost smoking , add steaks ; 4 .	Cook	2	minutes	With	tongs , lift each steak from pan and
5	2007	MAG	Prevention	A	B	C	Cover Dutch oven and place on middle rack of oven ;	Cook	2	hours	or	until meat and beans are tender .
6	2000	MAG	SouthernLiv	A	B	C	longer pink . STIR in crawfish and next 6 ingredients ;	cook	3	minutes	or	until thoroughly heated . Spoon mixture into
7	2012	MAG	VegTimes	A	B	C	until reduced by half . Stir in cheese mixture , and	cook	2	minutes	or	until cheese is melted , stirring
8	1998	MAG	Redbook	A	B	C	3 . Add salsa , broth , and remaining salt and	cook	2	minutes	or	until thickened . Return pork pieces and

Figure 5. Example of a KWIC search in COCA

The COCA also presents two other different options, *word* and *browse*:

- *Browse*: Allows searching for examples to the word form, preferred part of speech, frequency range, pronunciation and rhymes, and number of syllables and word stress.

Corpus of Contemporary American English

BROWSE/RANDOM WORD CONTEXT CONTEXT +

HELP Find random words

Click on any of the search types below for more information and examples. (more...)

Word form	<input type="text"/>
Part of speech	<input checked="" type="checkbox"/> NOUN <input checked="" type="checkbox"/> VERB <input checked="" type="checkbox"/> ADJ <input checked="" type="checkbox"/> ADV <input checked="" type="checkbox"/> OTHER <input type="checkbox"/> ALL
Range	<input type="text"/> - <input type="text"/>
Pronunciation	Rhymes with <input type="text"/> Type <input type="text"/> EXACT
Syllables / stress	○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ×

Show all words Reset

Figure 6. Example of a Browse search in COCA

- *Word*: Different information about a word is presented at a glance, like definition, images, pronunciation, chart, or collocates.

Corpus of Contemporary American English

SEARCH WORD CONTEXT CONTEXT +

however (ADV) #333

BLOG WEB TV/M SPOK FIC MAG NEWS ACAD

1. however, D M O C G

PlayPhrase YouGlish Yarn

Translate: choose language

SYNONYMS (more)

though however, nevertheless, nonetheless, still, though, yet

CLUSTERS (more)

however »	however i » however it » however you » however they » however much » however we » however there » however he
» however	for however » it however » do however » did however » not however » can however » you however » will however
however » »	however you want » however i do » however if you » however i have » however long it » however i think » however long it » however i would
» » however	i do however » it is however » i did however » there is however » i am however » there are however » i will however » i would however
however » » »	however i do n't » however long it takes » however it is not » however long it takes » however i do not » however you look at » however i would like » however i am not
» » » however	should be noted however » i do n't however » in the end however » in this case however » i do think however » i do not however » it does not however » to any discussion however

TEXTS / VIRTUAL CORPORA (more)

WEB:supermemo.com » WEB:starwars.wikia.com » WEB:...ooks.adelaide.edu.au » ACAD:AmerIndianQ » WEB:...heball.wordpress.com » WEB:malegislature.gov » WEB:ada.gov » BLOG:wattsupwiththat.com » WEB:...lytrinitymission.org » WEB:starwars.wikia.com » WEB:plato.stanford.edu » BLOG:wattsupwiththat.com » BLOG:blogs.suntimes.com » BLOG:stallman.org » ACAD:MarineFish » ACAD:AmerIndianQ » WEB:...portance.corante.com » BLOG:reakynews.com » WEB:dailywritingtips.com » BLOG:wattsupwiththat.com » WEB:w3.org » WEB:ada.gov » ACAD:AmerIndianQ » WEB:...pranos.wordpress.com » FIC:Analog »

CONCORDANCE LINES (more)

1	BLOG: 2012: ...omiccollapseblog...	it up this way.Neither Romney nor Obama was a choice however he did so with Romney we would have bought some more time
2	NYTimes: 1997: NYTimes	testine the children with only the Commissioner however work best for the children. # " I

Figure 7. Example of a Word search in COCA

3.5 Activity sequence

As mentioned earlier, the learning unit will be carried out throughout one scholar term (September through December) and will consist of 14 sessions, distributed in the following way:

- 2 sessions of corpora training.
- 11 practice sessions, arranged by content:
 - Three sessions on affixation.
 - Three sessions on word derivation.
 - One session on zero derivation and stress shift.
 - One session on acronyms and onomatopoeias.
 - One session on coinage and loans.
 - One session on clipping and backformation.
 - One session on compounds and blends.
- One last session for content review.

All sessions will follow the same structure, except the final activity, which varies depending on the word formation process featured in each session. The structure consists of the following activities:

Introduction and organization (5'-10')	The teacher activates students' prior knowledge about the topic, asking questions and providing examples. They also inform about the class' structure and timing.
Activity 1. Text analysis (10')	A text is presented to students, who must select the target vocabulary and classify it. (Appendix I)
Activity 2. Mind map (20')	Students are asked to complete a mind map (Appendix II) based on their predictions and hypothesis and then check them on the corpora. They will analyse the root or affix by exploring: <ul style="list-style-type: none"> - Meaning: examples, synonyms - Collocations - Variant and register differences - Part of speech - Pronunciation - Topics or clusters
Activity 3. Sharing and explaining (5-10')	Rules examined will be shared and confirmed by the teacher, along with an explanation. It will also be an opportunity for students to debate and share their theories and hypotheses.
Activity 4. Practice (10')	*
Work in the portfolio (5')	Students will record the knowledge they have learnt, and reflect on their practice and attitude. It will be filled out in every session (Appendix III).

Table 4. Activity sequence structure.

* The activities presented to practice the language feature examined in the text are the following (Appendix IV). Students will be arranged in pairs or small groups:

Session 1) Password game. One of them will choose a card from a deck and others will have to guess it, but only one word can be given, as a hint. For instance, a synonym. (Appendix 4.1)

Session 2) Matching card game. Learners, at random, will be given cards with roots and affixes. Each student will aim to create as many words as possible, by asking others for cards. (Appendix 4.2)

Session 3) Word chain game. Students will say a word at random, and the next student will continue the chain by producing another word starting with the last letter of the one that was previously said.

Session 4) Part of speech switch. Students will select a card from a deck in which a sentence containing a derived word will appear. They have to change the selected word into the categories specified (adjective, verb, noun or adverb). (Appendix 4.3)

Session 5) Register change: In this activity, students receive texts with small fragments, in a formal or informal tone. Their purpose is to transform the message and adequate it to the context, paying attention to underlined words. Then, they will share it through a role-play representation. (Appendix 4.4)

Session 6) The telephone: A small text will be given to the students, arranged in groups of three. In turns, they will have to summarize the contents of the text, trying to be more concise each time. (Appendix 4.5)

Session 7) News headline: Students select a card, and they have to create two news headlines in two manners. First, including two uses of the word, and secondly, substituting one of the terms with a synonym. (Appendix 4.6)

Session 8) New word entries: Students will make up new word suggestions and create dictionary entries for them.

Session 9) Web search: Students will debate on the origin and meaning of the words in the list, and prove their hypotheses through a web search. (Appendix 4.7)

Session 10) Web search: Students will debate on the original of clipped words in the list, and prove their hypotheses through a web search. (Appendix 4.8)

Session 11) Possible or impossible? In this activity, learners will select two cards at random from the deck and combine them to create a word. They will write down the word and whether it is a real word or not, their meaning, and then prove their suggestions through a web search. (Appendices 4.9 & 4.10)

On the other hand, two special sessions have been planned before dealing with corpora, which follow two different structures:

Session 1: Corpus training (Appendix V)

- Activity 1. Manual corpora

The main purpose of this activity is for students to comprehend the notion of corpus linguistics, as it is the first contact they have with corpora. To fulfill this aim, a selection of text fragments is going to be handed out to each two of students. Learners, in pairs, will have to read the texts and highlight, using colours, a word that is repeated in all texts.

Students will then have to count this word and analyse it in terms of frequency, part of speech it belongs to, and suggest some collocations and synonyms.

After performing this task, the teacher will reveal a faster way to do all this, and will introduce the notion of electronic corpora and corpus linguistics.

- Activity 2. Guided search in electronic corpora

With the guidance of the teacher, and using the worksheet provided, students will conduct a guided search in the established corpora. In this search, learners will explore the basic features of a corpus search, like list, chart, collocates, or compare and answer a series of questions.

- Activity 3. Autonomous search practice

In order to apply the knowledge about the corpus acquired in the previous activity, students will be invited to perform a search on a term of their preference with regard to the functions examined in the previous exercise.

- Portfolio work.

Session 2: Parts of speech and corpus training II (Appendix VI)

- Activity 1. Jabberwocky: reviewing parts of speech

In order to activate previous knowledge about grammatical categories or parts of speech and to perform more efficiently in corpora searches, learners will read the poem by Lewis Carroll and be explained that the words in the poem are nonsensical and do not exist. Their task is to recreate the poem by adding suitable words of their choice, from the indicated grammatical category, and share it with the class.

- Activity 2. Guided search in electronic corpora

With the guidance of the teacher, and using the worksheet provided, students will conduct a guided search in the established corpora. In this search, learners will explore the basic features of search involving parts of speech, KWIC searches and wildcard uses, like * and capital letters. They will answer a series of questions as they search.

- Activity 3: Data analysis activity

Learners will be presented with a series of situations related to language use, and will have to perform a search in the corpora to reach a conclusion. The purpose of this activity is to introduce students to language data analysis through corpora, as they will need to do in future sessions.

- Portfolio work.

Finally, and concerning the last session in which contents will be reviewed, students will take part in the same activities as the rest of the sessions. Nevertheless, students will analyse a text of their choice.

All activities will be carried out in the computer or electronic devices. The worksheets for the sessions and the portfolio pages will be uploaded in the class document sharing folder, so that materials are easily accessible and work is shared.

3.6 Assessment

Once the sessions have been carried out, it is necessary to evaluate the whole learning unit. The assessment will be developed by the students and the teacher. Students will assess themselves daily through the portfolio work, regarding the following aspects:

- Goal of the session. This way, students will become more engaged in the session and more aware of its purpose.
- Summary of the work done and useful facts to remember.
- Achievements in the session, like work completed or new words learnt.
- Rating the perceived difficulty of the session.
- Rating the perceived behavior, interest and attitude in the session.
- Aspects of the unit the student is confident with.
- Areas the student might need help with.

Through this portfolio, the worksheets and the observations made during the development of the unit, the teacher will be able to assess students in a final rubric (Appendix VII). The parameters included in the rubric are based on the general and specific objectives set for the lesson, as well as other attitudinal components. The teacher will rate students through a scale of 1-4, based on the following criteria:

- 1) The student performs poorly and struggles. Guidance is needed.
- 2) The student performs correctly, but only when help is provided.
- 3) The student performs correctly, needing only occasional help.
- 4) The student works independently, showing a high level of confidence.

On the other hand, the portfolio will serve as a daily follow-up of the work of the students, and through their reflections and the observations made in the classroom, the teacher will be able to identify their general perception of the difficulty, the most problematic areas, and identify those students that might be struggling the most or oppositely, might need a new challenge. It will help students as well, by prompting

critical judgement of their own work, raising self-consciousness. This will help address any difficulties that may occur during the implementation of the unit.

4. Conclusion

The main question this work has attempted to answer is how English teachers in Secondary Education can introduce corpus linguistics as a tool to learn word formation and establish relations with other aspects of word knowledge to promote a better and more complete acquisition of lexis. Investigating this matter was possible thanks to a theoretical review and the creation of a didactic proposal.

Among all the areas that corpus linguistics has permeated to, language learning seems to have benefitted from it the most. With the new technological advantages and the uprising of ICTs, learning through electronic corpora has become an invaluable resource for the acquisition and consolidation of language. Nevertheless, they are still far from becoming a common tool in the language classroom due to factors like teacher reticence. Some of these factors might be overcome by getting more acquainted with corpora and their benefits and constraints. By knowing the most problematic areas for students, teachers can adapt their lessons to provide the most suitable learning experience for the group.

This theoretical review has focused as well on the matters related with teaching vocabulary. With special regard it is considered Nation's (2000) model of the different dimensions of word knowledge; a knowledge that is partial in most cases. This is due to the fact that most language learning approaches disregard lexical competence and vocabulary is subject to translation activities, lists or left by its own. The proposed learning unit, based on the teaching of various word formation processes, addresses words from multiple points of view, establishing links with diverse aspects of word knowledge, like meaning, pronunciation or collocates through corpora. Further, the resources and materials implemented allow placing vocabulary at the center of the learning focus and subject to communicative goals. This way, learners can explore new depths of knowledge about terms they already know, and integrate new ones in their vocabulary.

To finish with the theoretical framework, explicit approaches to language teaching were examined: deductive and inductive methodologies. It was observed that both shared in common that the ultimate goal is the acquisition of the rule, and that overall, both promote analysis and critical thinking. Of special interest was the method of Data-Driven Learning or DDL, but introducing these methods in the traditional language teaching methodology implying that the manner of work and the role of teachers and students may change involves some difficulties and may not suit the needs of all. Because of this, we propose introducing corpora through a combination of inductive and more traditional methodology with activities that are familiar for students.

In this work the stage of 2nd grade of Bachillerato in the Spanish curricular context is considered an ideal scenario for the implementation of a learning unit based on corpus linguistics. The unit is intended to suit the needs of a group of learners whose ultimate goal is accessing superior education and continuing studying English as a first language. The designed activities present diverse processes of word formation in which learning is developed through the use of electronic corpora. These activities are based on a contextualized practice of word formation processes through the promotion of all language skills, in which a place for communicative situations has been granted.

Lexical contents are also related with cultural contents in an attempt to increase students' intercultural competence through exercises that motivate students to elaborate hypotheses that require a significant use of language. Finally, the ability to reflect and self-evaluate is put into practice with the work on a personal portfolio in which class practice and performance is self-assessed.

The intention is not only to introduce corpora for this learning unit alone, but to promote their use so that students continue consulting them autonomously during their learning process. Word formation, the lexical component and the learning unit presented in this paper are just some of the many possible ways in which electronic corpora can be introduced in the language classroom.

To finish with the theoretical framework, explicit approaches to language teaching were examined: deductive and inductive methodologies. It was observed that both shared in common that the ultimate goal is the acquisition of the rule, and that overall, both promote analysis and critical thinking. Of special interest was the method of Data-Driven Learning or DDL, but introducing these methods in the traditional language teaching methodology implying that the manner of work and the role of teachers and students may change involves some difficulties and may not suit the needs of all. Because of this, we propose introducing corpora through a combination of inductive and more traditional methodology with activities that are familiar for students.

In this work the stage of 2nd grade of Bachillerato in the Spanish curricular context is considered an ideal scenario for the implementation of a learning unit based on corpus linguistics. The unit is intended to suit the needs of a group of learners whose ultimate goal is accessing superior education and continuing studying English as a first language. The designed activities present diverse processes of word formation in which learning is developed through the use of electronic corpora. These activities are based on a contextualized practice of word formation processes through the promotion of all language skills, in which a place for communicative situations has been granted.

Lexical contents are also related with cultural contents in an attempt to increase students' intercultural competence through exercises that motivate students to elaborate hypotheses that require a significant use of language. Finally, the ability to reflect and self-evaluate is put into practice with the work on a personal portfolio in which class practice and performance is self-assessed.

The intention is not only to introduce corpora for this learning unit alone, but to promote their use so that students continue consulting them autonomously during their learning process. Word formation, the lexical component and the learning unit presented in this paper are just some of the many possible ways in which electronic corpora can be introduced in the language classroom.

Bibliography

- Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon* (4th ed.). Wiley-Blackwell.
- Allen, J. D., Anderson, D., Becker, J., Cook, R., Davis, M., Edberg, P., ... & Jenkins, J. H. (2014). *The Unicode Standard*. Unicode Consortium.
- Almela, M., & Sánchez, A. (2007). Words as “lexical units” in learning/teaching vocabulary. *International Journal of English Studies*, 7(2), 21-40.
- Alzu'bi, M. A. (2015). Effectiveness of inductive and deductive methods in teaching grammar. *Advances in Language and Literary Studies*, 6(2), 187-193.
- Ameel, E., Malt, B. C., Storms, G., & Van Assche, F. (2009). Semantic convergence in the bilingual lexicon. *Journal of Memory and Language*, 60(2), 270-290.
- August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research and Practice*, 20(1), 50-57.
- Baralo Ottonello, M. (2001). El lexicón no nativo y las reglas de la gramática. *Estudios de Lingüística Universidad de Alicante*, Anexo I, 23-38.
- Barcroft, J. (2004). Second language vocabulary acquisition: A lexical input processing approach. *Foreign Language Annals*, 37(2), 200-208.
- Biber, D., Douglas, B., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Blanco Suárez, Z., Gallardo del Puerto, F., & Gandón Chapela, E. (2020). The Primary Education Learners' English Corpus (PELEC): Design and compilation. *Research in Corpus Linguistics*, 8(1), 147-163.
- Blevins, J.P. (2006). English inflection and derivation. In B. Aarts & A. McMahon (Eds.), *The handbook of English linguistics* (pp. 507-536). Blackwell Publishing.
- Bogaards, P. (2001). Lexical units and the learning of foreign language vocabulary. *Studies in Second Language Acquisition*, 23, 321-343.

- Booij, G. (2006). Inflection and derivation. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (Vol. 5, pp. 654-661). Elsevier.
- Boulton, A. (2009). Testing the limits of data-driven learning: language proficiency and training. *ReCALL*, 21(1), 37-54.
- Boulton, A., & Tyne, H. (2015). Corpus-based study of language and teacher education. In M. Bigelow & J. Enns-Kananen (Eds.), *The Routledge handbook of educational linguistics* (pp. 301-312). Routledge.
- Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning*, 22(2), 153-172.
- Bybee, J. (1998). The emergent lexicon. *Chicago Linguistic Society*, 34(2), 421-35.
- Cambridge University Press. (n.d.). Lexis. In Cambridge Dictionary. Retrieved from <https://dictionary.cambridge.org/dictionary/english/lexis>
- Cambridge University Press. (n.d.). Vocabulary. In Cambridge Dictionary. Retrieved from <https://dictionary.cambridge.org/dictionary/english/vocabulary>
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied linguistics*, 1(1), 1-47.
- Caro, K., & Mendinueta, N. R. (2017). Lexis, lexical competence and lexical knowledge: A review. *Journal of Language Teaching and Research*, 8(2), 205-213.
- Chamot, A. U. (2005). Language learning strategy instruction: Current issues and research. *Annual Review of Applied Linguistics*, 25, 112-130.
- Cheng, W. (2012). Exploring corpus linguistics: Language in action. *Routledge introductions to applied linguistics*. Routledge.
- Cherifi, H. (2019). The essential contributions of corpora in language research. *New Trends in Social and Liberal Sciences (NETSOL)*, 4(2), 62-73.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. The M.I.T. Press.

- Cobb, T. & Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber & R. Reppen (Eds.), *The handbook of corpus linguistics* (pp.478-497). Cambridge University Press.
- Collins Easy Learning Grammar. (n.d.). Form of adverbs. Retrieved from <https://grammar.collinsdictionary.com/easy-learning/form-of-adverbs>
- Comrie, B., & Thompson, S. A. (1985). Lexical nominalization. *Language Typology and Syntactic Description*, 3, 349-398.
- Conrad, S. & Levelle, K. (2008). In B. Spolsky & F. Hult (Eds.), *The handbook of educational linguistics* (pp. 539-556). Blackwell Publishing.
- DeKeyser, R. (2003). Implicit and explicit learning. In C.J. Doughty & M.H. Long (Eds.), *The handbook of second language acquisition* (pp. 313-348). Blackwell.
- Dong, Y., Gui, S., & MacWhinney, B. (2005). Shared and separate meanings in the bilingual mental lexicon. *Bilingualism: Language and Cognition*, 8(3), 221-238.
- Elgort, I. (2018). Teaching/developing vocabulary using ICTs and digital resources. *The TESOL Encyclopedia of English Language Teaching*, 1-15.
- Ellis, N. C. (1994). Consciousness in second language learning: Psychological perspectives on the role of conscious processes in vocabulary acquisition. *AILA Review*, 11, 37-56.
- Ellis, N. C. (2011). Implicit and Explicit SLA and Their Interface. In C. Sanz & R.P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 35-47). Georgetown University Press.
- Ellis, N. C. (2015). Implicit and explicit language learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 1-24). John Benjamins.
- Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*, 42, 3-25.

- Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal*, 31(2), 105-119.
- Gabrielatos, C. (2005). Corpora and Language Teaching: Just a Fling or Wedding Bells? *The Electronic Journal for English as a Second Language TESL-EJ*, 8(4).
- Gillett, A. (2020). *Vocabulary building*. Retrieved from <http://www.uefap.com/vocab/build/building.htm>
- Gilquin, G. & Granger, S. (2010). How can data-driven learning be used in language teaching? In M. McCarthy & A. O’Keeffe (Eds.), *The Routledge handbook of corpus linguistics* (pp. 359-370). Routledge.
- Ginzburg, R. S., Khidekel, S. S., Knyazeva, G. Y., & Sankin, A. A. (1979). *A course in modern English lexicology*. Higher School Publishing House.
- Gollin, J. (1998). Deductive vs. inductive language learning. *ELT Journal*, 52(1), 88-89.
- Granger, S. (2019). Learner corpora. In C.A. Chapelle, *The Encyclopedia of applied linguistics* (pp. 1-8). John Wiley & Sons.
- Haastrup, K., & Henriksen, B. (2000). Vocabulary acquisition: Acquiring depth of knowledge through network building. *International Journal of Applied Linguistics*, 10(2), 221-240.
- Halliday, M. A. K., & Yallop, C. (2007). *Lexicology: A short introduction*. Continuum.
- Harper Collins. (n.d.). Lexis. In Collins English Dictionary. Retrieved from <https://www.collinsdictionary.com/dictionary/english/lexis>
- Harper Collins. (n.d.). Vocabulary. In Collins English Dictionary. Retrieved from <https://www.collinsdictionary.com/dictionary/english/vocabulary>
- Harper, D. (n.d.). Corpus. In Online Etymology Dictionary. Retrieved from <https://www.etymonline.com/search?q=corpus>

- Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition*, 27(2), 129-140.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.
- Jackson, H. (2002). *Lexicography: An introduction*. Routledge.
- Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research*, 4, 1-16.
- Kim, C. (2013). Vocabulary acquisition with affixation: Learning English words based on prefixes & suffixes. *Second Language Studies*, 31(2), 43-80.
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, 73(4), 440-464.
- Krashen, S. D. & Terrell, T.D. (1983). *The natural approach: Language acquisition in the classroom*. Prentice Hall Europe.
- Laufer, B. (1991). The development of lexis in the production of advanced L2 learners. *The Modern Language Journal*, 75(4), 440-448.
- Lee, D. Y. W. (2010). What corpora are available? In M. McCarthy & A. O’Keeffe (Eds.), *The Routledge handbook of corpus linguistics* (pp. 107-121). Routledge.
- Lee, P., & Lin, H. (2019). The effect of the inductive and deductive Data-Driven Learning (DDL) on vocabulary acquisition and retention. *System*, 81, 14-25.
- Leńko-Szymańska, A., & Boulton, A. (Eds.). (2015). *Multiple affordances of language corpora for data-driven learning* (Vol. 69). John Benjamins.
- Lipka, L. (1992). *An outline of English lexicology: lexical structure, word semantics, and word-formation* (Second Edition). Max Niemeyer Verlag.
- Llamazares, M. V. (2008). Lingüística con corpus. *E.H. Filología*, (30), 329-349.

- Ma, Q., & Kelly, P. (2006). Computer assisted vocabulary learning: Design and evaluation. *Computer Assisted Language Learning*, 19(1), 15-45.
- Mallia, J. G. (2014). Inductive and Deductive Approaches to Teaching English Grammar. *Arab World English Journal*, 5(2), 221-235.
- McCarthy, M. & O’Keeffe, A. (2010). What are corpora and how have they evolved? In M. McCarthy & A. O’Keeffe (Eds.), *The Routledge handbook of corpus linguistics* (pp. 3-13). Routledge.
- McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching*, 41(4), 563-574.
- McEnery, T. & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Meara, P. (1996). The dimensions of lexical competence. *Performance and Competence in Second Language Acquisition*, 35, 33-55.
- Merriam-Webster. (n.d.). Lexis. In Merriam-Webster.com Dictionary. Retrieved from <https://www.merriam-webster.com/dictionary/lexis>
- Merriam-Webster. (n.d.). Vocabulary. In Merriam-Webster.com Dictionary. Retrieved from <https://www.merriam-webster.com/dictionary/vocabulary>
- Meunier, F. (2011). Corpus linguistics and second/foreign language learning: exploring multiple paths. *Revista Brasileira de Linguística Aplicada*, 11(2), 459-477.
- Murphy, B. (1996). Computer corpora and vocabulary study. *Language Learning Journal*, 14(1), 53-57.
- Nam, J. (2010). Linking research and practice: Effective strategies for teaching vocabulary in the ESL classroom. *TESL Canada Journal*, 127-127.
- Nation, I. S. P. (2000). *Learning vocabulary in another language*. Cambridge University Press.

- Nation, P. & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt and M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge University Press.
- Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. *How to Use Corpora in Language Teaching*, 12, 125-156.
- Pavlenko, A. (2009). Conceptual representation in the bilingual lexicon and second language vocabulary learning. In A. Pavlenko (Ed.), *The bilingual mental lexicon: Interdisciplinary approaches* (Vol. 70). Multilingual Matters.
- Pérez-Paredes, P., Sánchez-Tornel, M., Alcaraz Calero, J. M., & Jiménez, P. A. (2011). Tracking learners' actual uses of corpora: guided vs non-guided corpus consultation. *Computer Assisted Language Learning*, 24(3), 233-253.
- Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato. Ministerio de Educación, Cultura y Deporte. «BOE» núm.52, de 1 de marzo de 2014.
- Römer, U. (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik*, 54(2), 121-134.
- Römer, U. (2010). Using general and specialized corpora in English language teaching: Past, present and future. In M. Campoy-Cubillo, B. Bellés-Fortuño & M. Gea-Valor (Eds), *Corpus-based approaches to English language teaching* (pp. 18-35). Continuum.
- Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205-225.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.
- Schwartz, A. I., Yeh, L. H., & Shaw, M. P. (2008). Lexical representation of second language words: Implications for second language vocabulary acquisition and use. *The Mental Lexicon*, 3(3), 309-324.

- Seliger, H. W. (1975). Inductive method and deductive method in language teaching: A re-examination. *IRAL: International Review of Applied Linguistics in Language Teaching*, 13(1), 1.
- Sinclair, J. (1999). The computer, the corpus and the theory of language. *Transiti Letterari e Culturali*, (2), 1-15.
- Stubbs, M. (2004). Language corpora. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 106-132). Blackwell Publishing.
- Tahaineh, Y. (2012). The Awareness of the English word-formation mechanisms is a necessity to make an autonomous L2 learner in EFL context. *Journal of Language Teaching & Research*, 3(6), 1106- 1113.
- Tognini Bonelli, E. (2010). Theoretical overview of the evolution of corpus linguistics. In M. McCarthy & A. O’Keeffe (Eds.), *The Routledge handbook of corpus linguistics* (pp. 14-27). Routledge.
- Willis, D. (2003). *Rules, patterns and words: Grammar and lexis in English language teaching*. Cambridge University Press.
- Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition*, 41-69.
- Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, 27(4), 567-595.

Appendices

Appendix I. Texts

Appendix 1.1

Edinburgh is crowned the 'Capital of Adventure' in Scotland

New survey finds Scots feel more appreciative of their home country following Covid-19 travel restrictions

Scots have crowned Edinburgh Scotland's 'Capital of Adventure' after lockdown.

A poll conducted by The Scots Magazine, asked people to share their favourite places to visit and Edinburgh took the top spot, for its accessibility and choice of activities. Scots described the capital as a place where you can just as easily walk in the footsteps of the Outlander cast, take a trip to Cramond Island, or get a slice of countryside at Gorgie City Farm, all in the same day, Edinburgh had the kind of broad appeal that many Scots love.

Isle of Skye, Glasgow, Loch Ness and Fort William also placed in the top five for the wide range of outdoor, cultural and food and drink attractions close by.

42 per cent of those surveyed say that they have discovered new places to visit and things to do close to home. And almost two-thirds of Scots admitted they had found a new love for their home country, saying they feel more grateful to live in Scotland following Covid-19 related travel restrictions.

The survey was commissioned as part of its launch of an interactive map that it says will help people find the best 'big and wee adventures' on their doorstep.

Robert Wight, Editor of The Scots Magazine, said: "While Scots have crowned Edinburgh 'Capital of Adventure' - which is a very worthy winner - we have launched our fantastic 'Adventure Is On Your Doorstep' map to show people that wherever you are in Scotland, and whatever you're interested in, there are loads of adventures for you close by.

"This is something we bring to you each month in The Scots Magazine. Whether it's a thrilling weekend in the mountains, an exciting cultural experience or discovering amazing local food and drink, we love helping people find their next adventure."

Daily Record (2020)

Notes from Underground: Subways of New York

In America's most populous city, life teems not only on the streets, but below them as well. The New York City subway opened for business on October 27, 1904 and since then it has become more than a way to get around, but a place in which the city lives, standing clear of the closing doors, in chunks of a half-hour at a time (or longer, depending on delays).



It's also a place where you can run into just about anyone—and not just the famous musicians who've been busking incognito with Jimmy Fallon in recent years. While the photos in this collection are heavy on famous faces, when pass through the turnstiles you are admittedly more likely to see commuters on the way to work, or school kids on the way to school, or a man with a parrot, or tourists on the way to one of the city's unending list of attractions. A subway ride can contain its hardships (wi-fi is spotty at best, so bring a book), especially so if the machinery breaks down. But it's also a way to beat the traffic, and as the photos show, noted New Yorkers such as John F. Kennedy Jr., Meryl Streep and Bernard F. Gimbel were not above going underground. They knew that this enduring monument to mass transit was a smart way to get where they were going.

Life (2020)

Where Are Stars Made? NASA's Spitzer Spies a Hot Spot

The nebula known as W51 is one of the most active star-forming regions in the Milky Way galaxy. First identified in 1958 by radio telescopes, it makes a rich cosmic tapestry in this image from NASA's recently retired Spitzer Space Telescope.

Located about 17,000 light-years from Earth, in the direction of the constellation Aquila in the night sky, W51 is about 350 light-years – or about 2 quadrillion miles – across. It is almost invisible to telescopes that collect visible light (the kind human eyes detect), because that light is blocked by interstellar dust clouds that lie between W51 and Earth. But longer wavelengths of light, including radio and infrared, can pass unencumbered through the dust. When viewed in infrared by Spitzer, W51 is a spectacular sight: Its total infrared emission is the equivalent of 20 million Suns.

If you could see it with your naked eye, this dense cloud of gas and dust would appear about as large as the full Moon. The Orion Nebula – another well-known star-forming region and a favorite observing target for amateur astronomers – occupies about the same size area in the sky. But W51 is actually much farther from Earth than Orion and thus much larger, and it's about 75 times more luminous. While Orion contains four known O-type stars – the most massive stars in the universe – W51 contains over 30.

"Star factories" like this one can operate for millions of years. The cavernous red region on the right side of W51 is older, evident in the way it has already been carved out by winds from generations of massive stars (those at least 10 times the mass of our Sun). The dust and gas in the region are swept around even more when those stars die and explode as supernovas. On the nebula's younger left side, many stars are just beginning to clear away the gas and dust in the same way the stars in the older region have done. It's apparent that many of these young stars are in the process of forming bubbles of empty space around themselves.

This image was taken as part of a major observation campaign by Spitzer in 2004 to map the large-scale structure of the Milky Way galaxy – a considerable challenge because Earth lies inside it. Called the Galactic Legacy Infrared Mid-Plane Survey Extraordinaire (GLIMPSE), the survey also turned up valuable data on many wonders within the Milky Way, including images of multiple stellar factories like W51 that were hidden by dust from visible-light observatories.

"The really spectacular images provided by Spitzer via the GLIMPSE survey – in concert with data from many other, complementary telescopes – give us insight into how massive stars form in our Milky Way, and then how their powerful winds and radiation interact with the remaining ambient material," said Breanna Binder, an assistant professor of physics and astronomy at California State Polytechnic University, Pomona, who studies the life cycles of massive stars. "We can't observe star-forming regions in other galaxies in anywhere near the level of detail that we can in our own galaxy. So regions like W51 are really important for advancing our understanding of star formation in the Milky Way, which we can then extrapolate to how star formation proceeds in other, nearby galaxies."

NASA: Jet propulsion laboratory (2020)

Appendix 1.4

Invisible people

There are many spaces we do not often think about unless, or until, we have to go there. These are spaces that are somewhat hidden from our daily lives, such as hospitals, nursing homes, mental institutions, prisons, and cemeteries, in other words, places for the ill, the criminal, the dying, the dead, and anything else we do not like to contemplate. Hospitals are, perhaps, the least hidden of these, located not too far from living areas to be accessible, but not too close to be unavoidable.

Carol Delaney. *Investigating Culture: An Experiential Introduction to Anthropology*.



Chapter Five. Weasleys' Wizard Wheezes



"What are you working on?" said Harry.

"A report for the Department of International Magical Cooperation," said Percy smugly. "We're trying to standardize cauldron thickness. Some of these foreign imports are just a shade too thin — leakages have been increasing at a rate of almost three percent a year —"

"That'll change the world, that report will," said Ron. "Front page of the Daily Prophet, I expect, cauldron leaks."

J.K. Rowling. *Harry Potter and the Goblet of Fire (2000)*

Five ways to check shortened links for safety

Verify a destination site is safe and suitable before you click a shortened URL.

Link shortening services are great for squeezing long URLs into tiny places. A shortened link is a URL with fewer characters that redirects to a longer destination URL. Many reasons exist for wanting to shorten a link. They take less time to type, are easier to remember, and are a method for tracking clicks. Shortening a link most often masks the original link, and its true destination becomes unknown. The need to check shortened URLs for safety and suitability arises from this condition.

Short links are easier to share than long links. Web developers often append URLs with query strings. A query string added to a URL can extend the length of a link by hundreds of characters. That is displeasing from an aesthetic standpoint. Links of that size are often truncated or divided when they reach the recipient. Such links can point to useless, invalid locations. Shortening services remedy the long link issue by reducing links into 10 to 30 character URLs.

James White (2017). *Medium*. <https://medium.com/@itsjameswhite/five-ways-to-check-shortened-links-for-safety-31e8e0dc1865>

Glen Carbon mayor displeased with trash hauler's pick-up postponement

GLEN CARBON – Glen Carbon Mayor Rob Jackstadt is not happy with Republic Services' decision to postpone its spring large item pick-up that was scheduled from April 13 through April 17.



During the village trustee meeting on April 14, Jackstadt discussed his thoughts after Village Administrator Jamie Bowden read off a list of canceled or postponed activities by the village.

"You mentioned, 'we stopped.' That's Republic Services' decision to stop," Jackstadt said. "I did have some concerns with that. We are all at home and we're running out of things to do, some of us. One of those things is you're wanting to clean your basement, your garage, your attic, your room, your office, your house.

"I have received a lot of questions about, 'Why can't we have large item pick-up or to transition to making the call to Republic Services ...'," he continued. "They're reducing the risk of touching items but they're still coming out, picking up garbage, picking up recycling, through mechanical and still offering pay-per-use yard waste."

"I will be talking with Republic Services to see if they get that service back sooner than later because we have no end in sight when this will end," he added, conceding maybe it is a safety issue for the company.

Republic Services replied to a reporter's questions with a statement.

"These temporary changes will help keep our routes running smoothly and our communities clean. Customers are being notified of service changes by their municipality and/or directly from Republic Services.

"We recognize the coronavirus (COVID-19) situation is evolving rapidly and we are continuously evaluating the situation to ensure we can help keep our employees healthy and safe while working hard to ensure the highest quality customer service and minimal service disruption.

"Our top priority is the health and safety of our employees, and that is at the forefront of every decision we make. We are confident that we have the necessary plans and protocols in place for protecting employees from COVID-19."

Charles Bolinger (2020). The Edwardsville Intelligencer

How Do I Get My Students Over Their Alternative Conceptions (Misconceptions) for Learning?

When teachers provide instruction on concepts in various subjects, they are teaching students who already have some pre-instructional knowledge about the topic. Student knowledge, however, can be erroneous, illogical or misinformed. These erroneous understandings are termed alternative conceptions or misconceptions (or intuitive theories). Alternative conceptions (misconceptions) are not unusual. In fact, they are a normal part of the learning process. We quite naturally form ideas from our everyday experience, but obviously not all the ideas we develop are correct with respect to the most current evidence and scholarship in a given discipline. Moreover, some concepts in different content areas are simply very difficult to grasp. They may be very abstract, counterintuitive or quite complex. Hence, our understanding of them is flawed. In this way, even adults, including teachers, can sometimes have misconceptions of material (Burgoon, Heddle, & Duran, 2010).

In addition, things we have already learned are sometimes unhelpful in learning new concepts/theories. This occurs when the new concept or theory is inconsistent with previously learned material. Accordingly, as noted, it is very typical for students (and adults) to have misconceptions in different domains (content knowledge areas). Indeed, researchers have found that there is a common set of alternative conceptions (misconceptions) that most students typically exhibit. There is one class of alternative theories (or misconceptions) that is very deeply entrenched. These are "ontological misconceptions," which relate to ontological beliefs (i.e., beliefs about the fundamental categories and properties of the world).

Alternative conceptions (misconceptions) can really impede learning for several reasons. First, students generally are unaware that the knowledge they have is wrong. Moreover, misconceptions can be very entrenched in student thinking. In addition, students interpret new experiences through these erroneous understandings, thereby interfering with being able to correctly grasp new information. Also, alternative conceptions (misconceptions) tend to be very resistant to instruction because learning entails replacing or radically reorganizing student knowledge. Hence, conceptual change has to occur for learning to happen. This puts teachers in the very challenging position of needing to bring about significant conceptual change in student knowledge. Generally, ordinary forms of instruction, such as lectures, labs, discovery learning, or simply reading texts, are not very successful at overcoming student misconceptions. For all these reasons, misconceptions can be hard nuts for teachers to crack.

Joan Lucariello, City University of New York & David Naff, Virginia Commonwealth University

Definition of healthy eating in the Spanish adult population: a national sample in a pan-European survey

Abstract

A national survey was carried out to find out how the Spanish adult population defined 'healthy eating'. Consumers were asked to describe in their own words what 'healthy eating' means to them. The sample included 1009 Spanish subjects over 15 y of age selected by a multistage procedure. This study belongs to the Spanish partnership in a pan-European survey about attitudes to food, nutrition and health coordinated by the Institute of European Food Studies of Dublin.

M A Martinez-Gonzalez, I Lopez-Azpiazu, J Kearney, M Kearney, M Gibney, J A Martinez (1998)

Fire Rises

"Help, Gabelle! Help, every one!" The tocsin rang impatiently, but other help (if that were any) there was none. The mender of roads, and two hundred and fifty particular friends, stood with folded arms at the fountain, looking at the pillar of fire in the sky. "It must be forty feet high," said they, grimly; and never moved.

The rider from the chateau, and the horse in foam, clattered away through the village, and galloped up the stony steep, to the prison on the crag. At the gate, a group of officers were looking at the fire; removed from them, a group of soldiers. "Help, gentlemen—officers! The chateau is on fire; valuable objects may be saved from the flames by timely aid! Help, help!"

Charles Dickens (1859) A tale of two cities

Sample St. Croix's Local Cuisine

Don't forget to sample the locally made beverages if you get a chance. There is a fantastic variety of local cuisine to feast on while you are here on St. Croix, so try something new. Stop at the farmers markets to sample and shop for locally made preserves, hot sauces, seasonings, and fresh produce. Or, support the local restaurants while you try the fresh seafood and local fusion cuisine. However you choose to indulge, make sure you savor the flavor of St. Croix!

Jennie Ogden. <https://www.gotostcroix.com/st-croix-blog/sample-st-croixs-local-cuisine/>

Chapter 4

*D*arcy, on the contrary, had seen a collection of people in whom there was little beauty and no fashion, for none of whom he had felt the smallest interest, and from none received either attention or pleasure. Miss Bennet he acknowledged to be pretty, but she smiled too much.

Mrs. Hurst and her sister allowed it to be so—but still they admired her and liked her, and pronounced her to be a sweet girl, and one whom they would not object to know more of. Miss Bennet was therefore established as a sweet girl, and their brother felt authorized by such commendation to think of her as he chose.

Jane Austen (1813) *Pride and Prejudice*

Appendix 1.8

HOW TO EAT CANDY LIKE A SWEDISH PERSON



At New York's newest Swedish candy store, Bon Bon, on the Lower East Side—there's another one, in the West Village, called Sockerbit, which means "sugar lump"—one of the scoop-your-own bins contains a confection labelled Socialcandy. Socialcandy are slightly sticky gummies, in opaque pastel shades ranging from yellow to pink to seafoam green, with vaguely tropical flavors (tutti frutti, you might say), and different shapes, most of which take the form of a word, acronym, or symbol of the Internet age. There's a LOL, a yolo, a hashtag, a thumb's-up sign that looks like the one on Facebook. There's an O.M.G., a SELFIE, an @, and a <3. The only outlier is a squishy stack of words that I had to squint at to make out: CANDY PEOPLE. The longer I looked at it, the stronger my desire was to eat it. In Internet parlance, "it me"—which is to say, I am a Candy Person.

Nordic countries, in general, are crazy for candy. But if any one particular country knows its candy, it's Sweden. A friend of a friend named Danielle, who is married to a Swedish man, marvelled that her "thin, thirtysomething" sister-in-law's nightly routine, at her home in Copenhagen, involves sneaking to the store on the ground floor of her apartment building after she's put her kids to bed, gathering a large bag of pick-and-mix candy, then plopping in front of the TV to eat it. "It struck me as totally shameless behavior," Danielle said. "The kind of thing you're not supposed to do after the age of twelve."

Hannah Goldfield. (2018). The New Yorker

Are Lizards as Silent as They Seem?

WHERE I LIVE in Florida, I can't open the door without causing a gecko or another lizard to skitter away. They've never complained. But maybe I haven't been listening.

While it is true that most lizards are mute, many make sounds of various kinds," Robert Espinoza, a biologist at California State University, Northridge, explains via email. Geckos are the gabbiest, and some produce "a variety of chirps, clicks, and squeaks, some inaudible to humans," Espinoza says.



The chirping, sometimes called 'barking,' of geckos is either a territorial or courtship display," to ward off other males or attract females, Peter Zani, a biologist at the University of Wisconsin-Stevens Point, says via email.

Some noteworthy noisemakers, Zani says, are Mediterranean house geckos, which squeak during fights and flirtatiously click to draw females. The turnip-tailed gecko of Central and South America makes territory-marking clicks thought to mimic insects. And the New Caledonian gecko, the largest gecko at 14 inches (36 centimeters), has a growl that earned it the local nickname of "the devil in the trees."

The sound of one gecko is even embedded in its name: Male tokay geckos, from Asia, make a loud, persistent mating call, "tokay-tokay!"

Liz Langley. (2015). National Geographic

Appendix 1.9

Eclipse Chocolate Bar & Bistro

We are delighted to offer custom chocolate, meal & mimosa kits, brunch boxes, frozen meals, groceries, & much more! Brunch is served every day on a half-shade / half-sun garden patio where we grow fresh basil, sage, lavender, jasmine, & butterfly friendly California blossoms.



We're delighted to make access to delicious food relaxing, easy, & safe. Hand sanitizer is provided to all visitors, social distancing is observed, & guests must wear masks unless seated at their sparkly turquoise table. A health survey is also conducted with all employees before every shift, including a temperature scan.

You are welcome order in advance for curbside pickup, or delivery. We offer contactless QR code menus, as well as laminated copies. All shared surfaces are cleaned frequently with bleach and commercial cleaners approved to help stop the spread of COVID-19. We're excited to safely serve you!

About us. Eclipse chocolate bar & bistro. San Diego, California, USA.

FOREVER 21[™] Women's Jackets & Coats



Brrrrr! It's getting pretty cold, isn't it? Even if it's not the winter season, there's always a chilly night out-and-about that gives you a reason to shop our collection of jackets and coats. If you just need a lightweight jacket, try our blazer, cropped jacket, or denim jacket. Given all of the options in a denim jacket, it really is a closet staple that everyone should have in their closet. If you want a jacket that's more tailored and versatile for dressier occasions, opt for a darker wash in a slimmer cut that sits closer to the body. Anoraks, windbreakers, and track jackets are also other options if you just need something light. The benefit of these jackets is that they are water resistant. So if there is a minor chance of rain, it doesn't hurt to throw these on top to prevent getting completely drenched.

But if you do need something more heavy-duty, for a ski trip or Alaskan getaway perhaps, then opt for our bombers, trench coats, puffers, and pea coats. Bomber coats look so good in an oversized fit. You can toss it over any outfit to bump up your street edge. Accessorize it with gold jewelry for a real street-inspired look. The same goes for puffer jackets which share the same vibe. They both boast a nylon shell with some kind of down filling held in place with quilted stitching. They do look really great though with fitted knit sweaters and leather boots. It's so New England cool. For a contemporary update, you can style an oversized trench with a graphic tee tucked into skinny jeans with a pair of heels.

Forever 21. Women's jackets and coats.

Travel in Liverpool



By Bike: Liverpool is a compact city centre so is easy to get around by bike and you'll be able to park even more into your day. There's cycle parking nearby most venues which is shown on the Liverpool Cycle Map.

By Bus: Bus travel in and around Liverpool is a green, convenient and cost-effective way to explore the destination.

Ferry: Take the world-famous 'Ferry Cross the Mersey' with Mersey Ferries, owned and managed by Merseytravel. Cross over to Wirral Peninsula while enjoying the on-board commentary before visiting the quaint Port Sunlight Village and Lady Lever Art Gallery.

Taxi: There's no shortage of taxis in Liverpool, making it quick and easy to grab a lift, whether you need to get to an attraction outside the city centre or just quickly whizz across town. Liverpool Lime Street Station has a taxi rank just outside if you wish to jump into a black cab, these can be hailed across the city just look for the orange light and you'll know if they're taking passengers.

Car Hire: If you'd prefer to be your own boss and drive yourself, there's several car hire companies in Liverpool offering visitors the choice to hire a vehicle for their stay.

ACCOMMODATION

[Hotels](#) | [Self-Catering B&Bs, Inns & Guest Houses](#) | [Budget Accommodation](#) | [Spas](#)

FOOD & DRINK

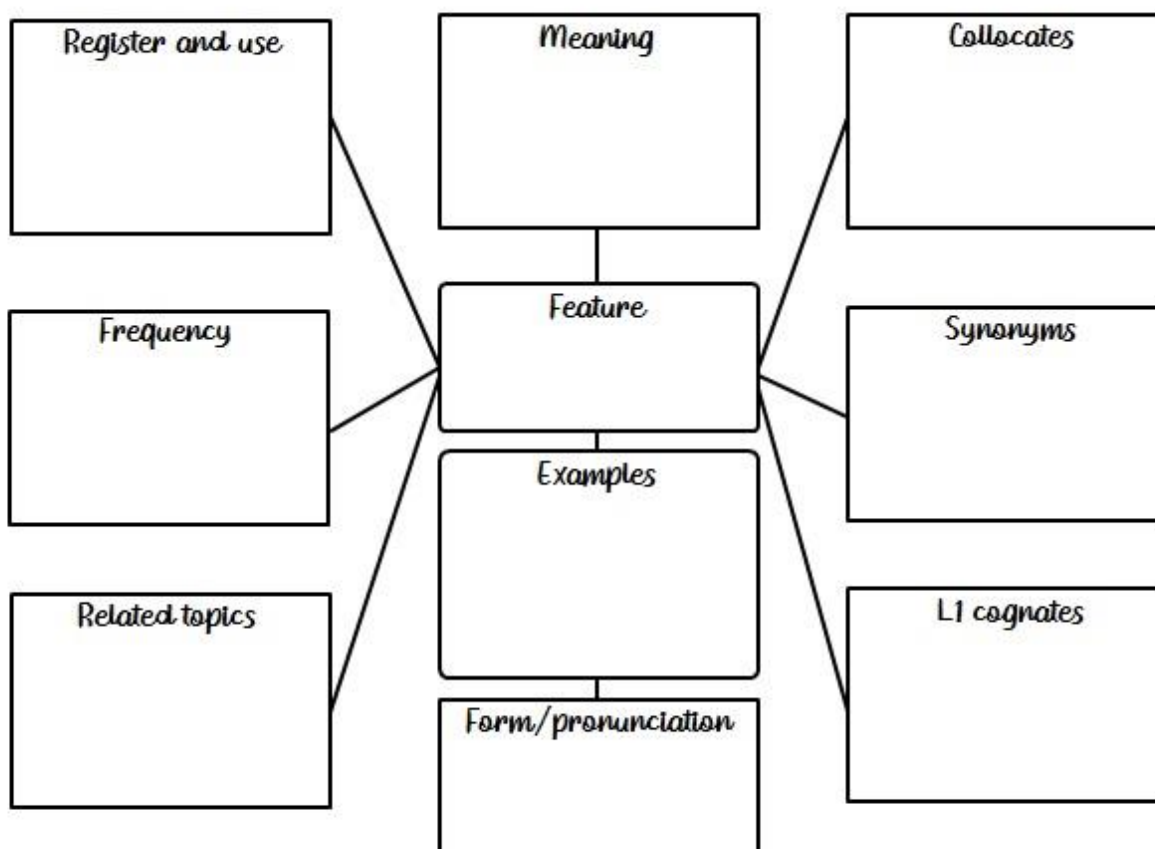
[Restaurants](#) | [Traditional Pubs](#) | [Cafes](#) | [Bars & Nightlife](#)

EXPLORE THE CITY

[Film Locations in Liverpool](#) | [Outdoor Spaces in Liverpool](#) | [Neighbourhoods](#) | [Beyond the City](#)

VisitLiverpool.com. Travel in Liverpool

Appendix II. Concept map model



Appendix III. Portfolio model

Session:

My goal for this session is:

Notes:

Achievements:

Comment on the difficulty of this session:

Comment on your attitude in class:

I think I'm confident with:

I think I will need help with:

Appendix IV. Activity materials

Appendix 4.1

unsafe	logical
dislike	cheerful
entertainment	dishonest
teacher	untidy
invisible	hopeless

Appendix 4.2

I have... <i>play</i>	I have... <i>comfort</i>
I have... <i>-able</i>	I have... <i>dis-</i>

Appendix 4.3

<p><i>The weather made our journey <u>impossible</u>.</i></p> <p>[noun]</p>
<p><i>All <u>employees</u> are on vacation.</i></p> <p>[adjective]</p>

Appendix 4.4

Stephen,

Thank you so much for your kind gift the other day! You know how much I love chocolate! I'm extremely happy to report that it will go to good use! I appreciate that you thought of me on my birthday!

With Love,

Christine

Appendix 4.5

At first, I thought one of my fuzzy, orange socks disappeared in the dryer, but I could not find it in there. Because it was my favorite pair, nothing was going to prevent me from finding that sock. I looked all around my bedroom, under the bed, on top of the bed, and in my closet, but I still could not find it. I did not know that I would discover the answer just as I gave up my search. As I sat down on the couch in the family room, my father was reclining on his chair. I laughed when I saw that one of his feet was orange and the other blue! I forgot that he was color-blind. Next time he does laundry I will have to supervise him while he folds the socks so that he does not accidentally take one of mine!

Appendix 4.6

light

present

fire

conflict

bank

record

work

contrast

water

protest

Appendix 4.7

Bazaar

Kindergarten

Aficionado

Buckaroo

Poltergeist

Tangram

Samurai

Wanderlust

Rucksack

Ketchup

Bamboo

Taboo

Divan

Boss

Buffalo

Cosmonaut

Cashew

Ukulele

Aspirin

Band-Aid

Laundromat

Google

Kerosene

Frisbee

Kleenex

Hoover

Sandwich

Coke/Cola

Polaroid

Scotch tape

Vaseline

Appendix 4.8

Ad	Memorandum	Edit
Gator	Situation comedy	Automate
Exam	Facsimile	Donate
Gas	Influenza	Brainwash
Gym	Cafeteria	Enthuse
Lab	Professor	Typewrite
Math	Kilogram	Televise
Photo	Promotion	Housekeep
Rep	Exposition	Lase
Phone	Parachute	Diagnose

foot

proof

spoon

fog

water

blue

electronic

hungry

print

sound

smoke

fork

teacher

way

biology

anger

air

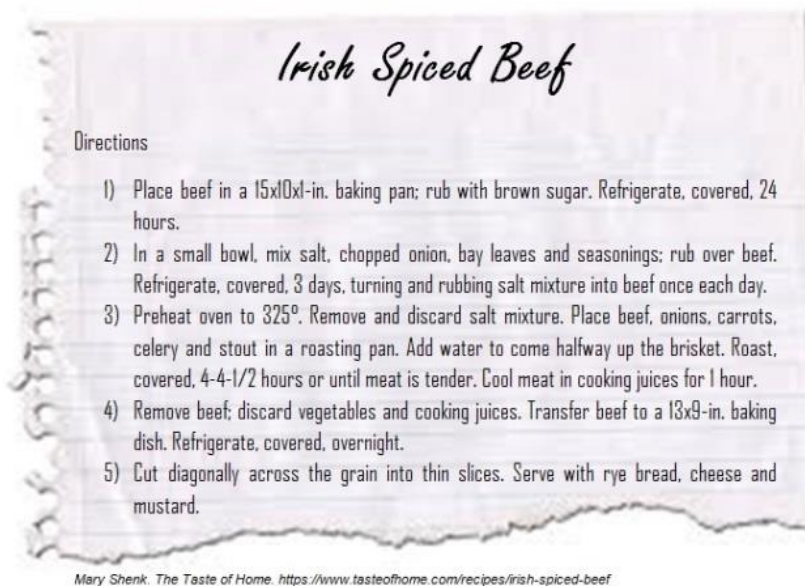
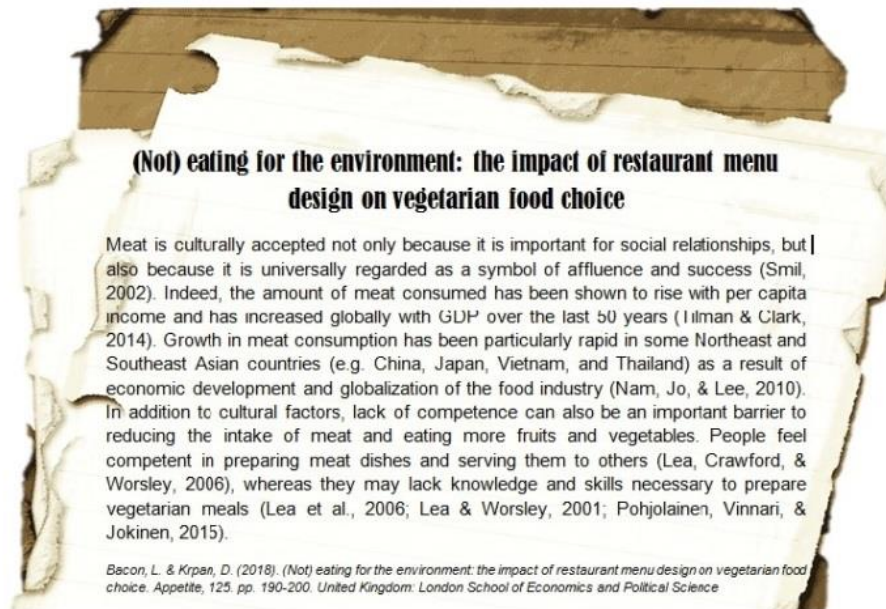
box

chill

relax

Appendix V. Corpus training (session 1) materials

Appendix 5.1



From fish to bacon: A ranking of animal protein in order of healthfulness

As an average consumer, you probably have a vague awareness of the nutritional value of your meats - fish being better than red meat, for example. The issue can be complicated, because all meats have pros and cons, research can come up with conflicting results, and studies can surprise us. For example, research suggests that in terms of cholesterol alone, eating white meat chicken is as bad for you as eating beef.

Still, there's a generally agreed upon hierarchy of nutritional value when it comes to meat, and small shifts in your diet might have greater effects than you realize. In a study of the Danish population, researchers found that Danes could gain more than 7,000 years of healthy life annually if they ate the recommended quantity (12 ounces per week) of fish while replacing red and processed meats in their diet.

Jenna Birch (2019). The Washington Post.

Hold this

School was over for the year. Silver Spring was having a drought and everyone's yards were brown. But the lightning bugs blinked happily anyway. The neighborhood smelled vaguely like grilled meat. We didn't have plans for the summer like other kids. We weren't going to camp or to the beach or to Disney World. We weren't going to visit relatives in other states. We weren't that kind of family. Anyway, our father had pretty much left us on our own. His latest girlfriend wasn't much of a kid person, he told us. She had a Little Mermaid tattoo on her ankle and she smelled like coconuts. But she wasn't a kid person, he insisted, and so she never came over, which meant he stayed at her apartment a lot, which meant we had the house to ourselves.



Eve Gleichman. (2018). Hold this.

Appendix 5.2

Date:

Classmate:

Session 1

Learning to use electronic corpora

Guided corpora search

- ❶ Function: LIST. Search for the term “library” in the British National Corpus (BNC). Now search for the same word in the Corpus of Contemporary American English (COCA). Would you say that “library” is frequent in each variant? In which one is it more frequent?

- ❷ Function: CHART. Search for the term “digital” in the BNC. In which genre is the term used more often? In which subject of non-academic publications is this word used the most?

- ❸ Function: WORD (COCA). You are reading a text and you come across the word “apron”. Find out what it means by searching for this term in COCA. How many meanings does it have? How is it pronounced? Are there any synonyms for this word? Can you name some collocates? Topics? Clusters?

- ❹ Function: COLLOCATES. Think about the term “brilliant”. Can you suggest other words that collocate with this term? Search for one collocation to the right of “brilliant” and see if they match. Now think about the word “job”. Can you suggest any collocations? Search for one collocate to the left of the word. Do they match?

- ❺ Function: COMPARE. Do you think the words “attractive” and “beautiful” are synonyms? Why/why not? Search for these terms in BNC or COCA. Which differences in meaning can you find based on their collocations?

Appendix 5.3

Free corpora search

Select a term of your choice and analyze it with the help of the corpora paying attention to the following aspects:

Definition(s):	Top 5 collocates (1 left)
Pronunciation:	Top 5 collocates (1 right)
Synonym(s):	
Genre(s):	Topic(s):
Subject matter(s):	Cluster(s):

PRACTICE: JABBERWOCKY



Jabberwocky

'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe;
All mimsy were the borogoves,
And the mome raths outgrabe.

"Beware the Jabberwock, my son!
The jaws that bite, the claws that catch!
Beware the Jubjub bird, and shun
The frumious Bandersnatch!"

He took his vorpal sword in hand:
Long time the manxome foe he sought--
So rested he by the Tumtum tree,
And stood awhile in thought.

And, as in uffish thought he stood,
The Jabberwock, with eyes of flame,
Came whiffling through the tulgey wood,
And burbled as it came!

One two! One two! And through and
through
The vorpal blade went snicker-snack!
He left it dead, and with its head
He went galumphing back.

Jabberwocky

'Twas [ADJ], and the [ADJ] [NOUN]
Did [VERB] and [VERB] in the [NOUN];
All [ADJ] were the [NOUN],
And the [ADJ] [NOUN] [VERB].

"Beware the Jabberwock, my son!
The jaws that bite, the claws that catch!
Beware the Jubjub bird, and shun
The [ADJ] Bandersnatch!"

He took his vorpal [NOUN] in hand:
Long time the [ADJ] foe he sought--
So rested he by the [ADJ] tree,
And stood awhile in thought.

And, as in [ADJ] thought he stood,
The Jabberwock, [ADV],
Came whiffling through the [ADJ] wood,
And [VERB] as it came!

One two! One two! And through and
through
The vorpal [NOUN] went snicker-snack!
He left it dead, and with its head
He went [VERB] back.

Appendix 6.2

Date:

Classmate:

Session 1

Learning to use electronic corpora

Guided corpora search

❶ Function: BROWSE (COCA). This function allows us to search for words that rhyme with the term that we have introduced. Search for words that rhyme with the adverb “slowly”. Are there any other adverbs?

❷ Function: KWIC (Keyword in context). Search for the words “happen” and observe one collocation to the right. Which word classes or part of speech are frequent?

Now search for the term “wrong”. Which verbs does it usually collocate with? Examine collocations to the left.

❸ Function: LIST. “Head” is a word that has different meanings, and which can be a noun or a verb. Search for “head” as both selecting “noun.ALL” and “verb.ALL”. Which are these meanings?

Now search for collocations by part of speech. Search for prepositions that collocate with the verb “make”.

❹ Function: LIST. Any word, like “break” or “green” can take many different forms. Examine them by searching for these words in capitals. Example: BREAK, GREEN.

❺ Function: LIST. There are “wildcards” that we can use in our corpus searches, like *. Search for different words that begin with the prefix “pre*”, and words that end in the suffix “*-y”.

Appendix 6.3

Data analysis

❶ “Say” and “claim” have a similar meaning. However, these words are not exactly the same word and they may have different uses and connotations. Paying attention to their usage in the different genres, in which context would you use them?

❷ What are the differences between “labor” and “labour”? When would you use each of them?

❸ Is “series” a singular or a plural word? Pay attention to the verbs it collocates with.

Appendix VII

	Parameter	Rating (1-4)
Work in the corpora	Performs correctly in guided searches.	
	Performs correctly in free consultations.	
	Analyses the language data obtained, drawing conclusions.	
	Uses corpora as reference tools and a means to verify hypotheses.	
Word knowledge and word formation	Recognizes the internal structure of words, identifying roots and affixes.	
	Comprehends and applies the processes of affixation and derivation.	
	Comprehends and applies other processes of word formation (zero derivation, stress shift, acronyms, coinage...)	
	Applies studied features to create or suggest new terms.	
	Analyses words in relation with the cultural aspects, like register.	
	Establishes links with other aspects of words, like synonyms, by hypothesizing.	
	Establishes links with the L1.	
Attitudinal	Engages and participates in class debates.	
	Works autonomously.	
	Works collaboratively with others.	
	Respects materials and the members of the group.	
	Self-reflects through the portfolio.	