Poetry in EFL: a Study of Vocabulary Acquisition Through a Poetry-based Implementation

Pol Vinyeta Muncunill

ACKNOWLEDGEMENTS

I would like to thank my parents for their availability and their patience.

ABSTRACT

This Master's Thesis aims to determine whether poetry can be a useful material for teaching vocabulary in English as a foreign language (EFL) lessons. This is why a study was conducted with a group of EFL adult learners, who completed a poetry-based activity that targeted 12 lexical items. Students were administered two tests to measure the evolution of their knowledge of these items. The hypothesis was that, after the implementation, the test results would show an improvement in their lexicon. Furthermore, the participants answered a questionnaire about their views on the poetry activity. The results show that the participants' knowledge of the items increased by 64%, which confirms the hypothesis. Regarding the questionnaires, most learners found the activity both enjoyable and effective. While learners certainly improved after the implementation, the methodology had some shortcomings, such as the lack of a control group and the polysemy of certain lexical items, which limit its conclusions. Therefore, more thorough research is needed to determine whether poetry can be an aid to vocabulary acquisition or not.

Keywords

English as a foreign language (EFL), adult learners, vocabulary, poetry, vocabulary test, perception questionnaire

RESUM

Aquest Treball de final de màster té com a objectiu determinar si la poesia pot ser un material útil per ensenyar vocabulari en l'àmbit de l'anglès com a llengua estrangera (ALE). Per això es va dur a terme un estudi amb un grup d'estudiants adults d'ALE, que van realitzar una activitat basada en poemes centrada en l'adquisició de 12 unitats lèxiques. Als estudiants se'ls van donar dos tests per mesurar l'evolució del seu coneixement d'aquests ítems. La hipòtesi va ser que, després de la implementació, els resultats dels tests mostrarien una millora en el seu lèxic. D'altra banda, l'alumnat va respondre un qüestionari de percepció sobre la seva opinió en relació amb l'activitat realitzada. Els resultats mostren que el seu coneixement dels ítems va augmentar un 64 %, cosa que confirma la hipòtesi. Quant als qüestionaris, la majoria de l'alumnat va trobar l'activitat agradable i efectiva. Si bé l'estudiantat va millorar després de la implementació, la metodologia tenia alguns defectes, com ara la falta d'un grup de control o la polisèmia de determinades unitats lèxiques, fet que limita les seves conclusions. Per tant, cal recerca més exhaustiva per determinar si la poesia pot ser un ajut a l'hora d'adquirir vocabulari o no.

Paraules clau

anglès com a llengua estrangera (ALE), alumnat adult, vocabulari, poesia, test de vocabulari, qüestionari de percepció

TABLE OF CONTENTS

TABLES

1. INTRODUCTION

As part of a teachers' training program, the present Masters' thesis will try to contribute to the practice of Teaching English as a foreign language (TEFL) by researching a specific topic in this area. More specifically, it will focus on how poetry can be used as a teaching material. The aim of the paper will be to determine if poetry-based activities can help adult EFL learners acquire vocabulary. To this end, a research study has been designed and implemented with a group of adult students in an Official Language School (EOI) in Catalonia. The hypothesis is that there will be an improvement in the students' lexicon after having worked with poems in class. After collecting data from the study, the results should lead to either a confirmation or a refutation of the hypothesis.

While short stories and excerpts from novels are sometimes a part of TEFL classes and textbooks, poetry is hardly ever used by teachers. Its complexity has often led to its dismissal as a superfluous material and, in consequence, foreign language learners rarely encounter poems in the classroom. However, I have decided this material is worth exploring for several reasons.

The first reason, and perhaps the most subjective one, is personal interest. Having studied EFL since my childhood, I have been in many different teaching contexts, yet poetry has seldom been included in any classes. The only time I can remember this happening was during an EOI course. One day, the teacher decided to use three poems in an activity, which we read, listened to, and discussed. Though this was a long time ago, I can still remember enjoying it, especially the videos of famous actors performing the poems. Since then, I have continued to listen to poems, and this is why I believe including poetry in EFL can contribute to the learners' appreciation of this literary genre.

Another reason that is behind the choice of this material is that there is little research about poetry in TEFL. In fact, there are almost no studies that connect vocabulary acquisition to poems in this field. This means that there is virtually no evidence either in favour or against including this material in the classroom. Given this research gap, it is quite evident that this topic deserves some attention, so the present paper could represent an initial attempt at showing that poems do have a place in EFL classes and curriculums.

Additionally, poetry may be a useful TEFL resource because it is culturally relevant. Just like cooking traditions, clothing and dancing are connected to the different groups of people that speak English, so is poetry. Learning languages is, of course, related to internalising certain forms, but it is also connected to culture and tradition. In consequence, learners should be exposed to these cultural forms throughout the learning process, and poetry should be one of them.

Having the aim of the paper in mind, a specific methodology was designed to obtain evidence about whether poetry can help learners improve their vocabulary. This methodology consisted of an activity that required students to work with poetry, and two instruments for collecting data: a vocabulary test and a perception questionnaire. Although, in this type of thesis, the methodology usually includes a learning scenario that is implemented from beginning to end during the internship period, in this case, the teaching context did not allow for such a long implementation. In consequence, the learners were exposed to the material using a short poetry-based activity. More specifically, this activity revolved around two short poems, which students listened to and read. After this, they did a vocabulary exercise and there was a final group discussion about the poems and other connected topics. As for the instruments for data collection, the vocabulary tests assessed the students' knowledge of some specific lexical items from the poetry-based activity. The perception questionnaire was included to obtain information regarding the learners' subjective views on the whole implementation and on poetry more generally.

The context in which the research was carried out was the EOI of Terrassa, which is where I did my Master's internship. In Spain, EOIs are public schools where different modern languages are offered to adult learners. There are 45 schools of this kind in Catalonia, and the course contents are organised following the Common European Framework of Reference (CEFR). They allow students to sit certificate exams at the end of most

academic years, though some CEFR levels are distributed in two years. This is the case of the B2 level which, in the case of English, is split into B2.1 and B2.2. In the present study, the activity was implemented with a group of 13 B2.2 students. It is worth bearing in mind that, in EOIs, students tend to be more engaged than in other instructional contexts like high schools. While there are usually between 20 and 30 students enrolled in each group, class attendance tends to decrease as the year goes by, which is why there were only 13 learners in the B2.2 group. The size of the groups makes it easier for the teacher to create an affective connection with students, and in most classes, there is an atmosphere of familiarity.

The remainder of the paper will be organised into three main sections. First, there will be a theoretical framework that will review the most relevant research connected to the topics of poetry and vocabulary. More specifically, it will deal with the reasons in favour of using poetry in TEFL, and with different aspects of vocabulary studies, such as word knowledge and lexical acquisition. The next section will explain the development of the poetry-based activity and detail the methodology that has been followed to collect the research data. Finally, the closing section will present the results in the form of tables and discuss their implications in relation to the initial hypothesis. In addition, there will also be a discussion on the limitations of the design of the research, and some possible solutions will be presented. The reader will also find four annexes at the end of the paper; they contain the worksheet for the poetry-based activity, the model for the vocabulary test, the model for the perception questionnaire, and some answers to the latter.

2. THEORETICAL FRAMEWORK

2.1 Using poetry in the EFL classroom

Literature is the main artistic expression built from language, so it is not surprising that it has often been used as a material for teaching foreign languages. Novels, for instance, are sometimes part of EFL courses, though this is usually done using adapted materials—in the form of graded readers—rather than authentic works of literature. When it comes to the local languages, in our country they are often taught together with literature; in secondary education, for example, "Llengua catalana i literatura" is one course. Foreign languages, however, seem to follow a different rationale, as the priority, in this case, is to develop the learners' communicative and linguistic competences; therefore, literature becomes a secondary pursuit. The question, then, is whether EFL learners' linguistic competences can actually be developed by including unmodified novels, short stories, and poems in the classroom.

There are quite a few arguments that are often used to rule out literature as a teaching material in EFL. McKay (1982), in her seminal article "Literature in the ESL Classroom", mentions three of the most frequent ones, which are the following. First, if the main aim of language teachers is to "teach the grammar of the language, literature, due to its structural complexity and its unique use of language, does little to contribute to this goal" (1982, p. 529). Second, if we want to help our students achieve their academic goals, literature will not help us in any way. Third, because of its "particular cultural perspective" (p. 529), literature will be too complex for foreign language learners, so they will learn nothing from it.

We should not be too quick to dismiss these ideas; in fact, EFL teachers should consider them before including literature in the classroom. Still, there are also good reasons why literary texts can help students develop different language-related competences, which will be explored shortly. Before that, nevertheless, I will provide a short overview of how literature has contributed to the teaching of foreign languages in recent history.

2.1.1 A short history of literature in foreign language teaching

According to Kramsch and Kramsch (2000), "although literature had played a preeminent role in the teaching of foreign languages until World War I, its ascendancy (. . .) waned in five successive stages" (2000, p. 554). In their paper, these five stages are described, and, although they are related mainly to the United States, I believe they are relevant in other countries as well; after all, teaching and learning English is nowadays a worldwide phenomenon, and the cultural history of the US is linked to world history.

As the researchers state, in the 20th century, "the two world conflicts among warring nations have problematized the role that literature, as the cultural heritage of nation-states, has played in the teaching of foreign languages" (2000, p. 553). Therefore, it is worth considering the development of literature instruction as a part of the history of EFL acquisition. Before and during World War I, the materials that were used for teaching foreign languages were almost invariably literary. In the USA, French and German were taught using texts by Victor Hugo, Goethe and other classic authors (2000, p. 555), and literature was not only about grammar and lexicon, but also about internalising moral values such as "duty" and "patriotic hero-worship" (2000, p. 555). However, "the years following WWI showed a disillusionment with the value of both speaking foreign languages and studying literature for literature's sake" (2000, p. 554). During this period, getting the population to achieve literacy became a major priority, and teaching literature became a secondary educational objective. The Coleman report (1929) confirmed this tendency, and in the 1930s literature was partially replaced with "reading education as the primary objective of language teaching" (2000, p. 554). Though literature did not disappear from the classroom, after the emergence of the social sciences and educational psychology, it was often simplified and abridged in the form of graded readers. In addition, regarding the debate over intensive versus extensive reading, the latter was the clear winner (2000, p. 562), which implies that the close reading of literature was considered almost useless in foreign language teaching.

Then, after World War II, speaking the language—rather than reading it—became the main priority. "The ability to read literary texts was seen as following from the ability to speak the language" (2000, p. 563). This was the beginning of the scientific study of second language acquisition (SLA), and it was also when the audiolingual method began to be applied. The final stage corresponds approximately to the 1980s and 1990s. During this period, literature played an extremely marginal role, almost as a pretext "for providing an "authentic" experience of the target culture" (2000, p. 568).

In Kramsch and Kramsch's view, though literature has appeared in many different forms throughout the 20th century, "the poetic function of language that undergirds these various manifestations of the literary (. . .) has remained largely invisible up to now" (2000, p. 569-570). This is why they contend that future research on the topics of literature and foreign language learning should also be concerned with the essentialness of the poetic function of language for both learners and teachers (2000, p. 570).

After this brief summary of how literature has been approached in foreign language education, I will move on to the specific case of poetry, a genre that has only occasionally been explored in the field of TEFL research. There are several reasons for using this material in foreign language contexts, which will be presented in the following subsection. Then, I will comment on two of the studies published so far on literature in EFL and, finally, I will return to the main arguments against using poetry as a teaching material and present some possible counterarguments.

2.1.2 Reasons for using poems in EFL

When we consider the different arguments in favour of including poetry in class, perhaps the most obvious one is its length. Even though some poems are certainly long (e.g. *The Waste Land*), lyrical poetry tends to be short. In fact, it is quite easy to find examples with an extension of less than ten lines, such as haikus and limericks. This allows students to read an entire poem several times in one session, which means they can work with its content and its language, interpret it, talk about it, etc. As Khatib et al. (2011) point out, while "novels are good for extensive reading purposes (. . .) the best literary text for intensive reading purposes can be poetry" (p. 202). Of course, any passage from a novel or an essay could be used for intensive reading, and this is already done in textbooks, yet poems have the advantage that there is no need to abridge them, which means students can see the entire piece and consider it for themselves.

A second reason for using poems in TEFL is that they can be listened to as well as read. In many cultures, poetry was originally written to be performed aloud in front of other people, rather than silently read from books. Nowadays, thanks to online platforms such as YouTube, we can listen to poetry being read and performed. It is not uncommon to find videos of well-known actors like Anthony Hopkins reading classic poems. Furthermore, the more modern phenomenon of spoken verse may also be used to update the experience of listening to poetry and to make it seem less elitist. Poems, thus, can be appreciated through more than one sense; in other words, they can become a multimodal experience. As Khatib et al. (2011) put it, "for listening purposes, the learners can be exposed to the audio versions of the poems" (p. 203), and they can also practise pronunciation by "adhering to the principles of rhythm, rhyme, and intonation" (p. 203). A possible pronunciation activity would be for learners to choose a poem from a list and to practise it at home; they would have access to some videos in which celebrities read the poems, and they would have to imitate their pronunciation as closely as possible. The following week, students would read their chosen poems aloud in front of the class, and they would get feedback about their pronunciation.

A third argument is that poetry can contribute to developing learners' critical thinking. This is connected to a specific characteristic of poetry, that is, its "structural complexity and its unique use of language" (McKay, 1982, p. 529). Though this complexity might, at first sight, seem a drawback, it may also be understood as an advantage, because this text type also challenges readers to find their own interpretation. One of the characteristics that almost all poems have in common is the presence of figurative language. Although understanding metaphors, metonymy and other rhetorical devices does require reflection and patience, in the long run, learners will benefit

from the comprehension effort. Langer (1997), a scholar specialised in literacy, writes that, "when students read literature, "horizons of possibility" come to mind, moving them to reflect on and interpret ideas at hand; students raise questions, recognize problems, seek causes and solutions, and make connections" (1997, p. 607). These verbs describe higher-order thinking skills, which the experience of understanding poetry can help develop.

Also in support of the use of poems in the EFL classroom is the fact that they are authentic materials. In other words, they are texts that have been produced to fulfil a social purpose in the culture where the language is spoken. In the case of poetry, this purpose is most often either entertainment or instruction (or both). Therefore, because poems are an authentic material, students may be able to enjoy themselves and learn from the reading experience, as long as the chosen texts are adequate for their linguistic level. This enjoyment connects with the aesthetic experience that reading poetry can induce in readers, which McKay (1982) believes should be the main priority when reading literature, even in an EFL classroom (p. 533).

In relation to this last point, McKay (1982) makes a distinction between efferent reading and aesthetic reading. In the former, "the reader is concerned with what she will carry away" (p. 532), that is, with extracting useful information from the text; in the latter, what really matters is "the enjoyment attained by interacting with the text" (p. 533). Even though McKay contends that literature in EFL should be focused mainly on the aesthetic experience, she concedes that language usage is also part of it, but "only when it impedes or highlights that experience" (p. 533). Having such an aim is extremely commendable, and teachers should try to encourage aesthetic reading, yet, in a real EFL context, it is almost impossible to avoid focusing on forms completely. In other words, efferent reading will always be a necessary part of the literary experience.

Yet another advantage of poems is that, because of their diversity of themes and subgenres, they can contribute to develop the learners' intercultural competence. Though many poems are too linguistically complex for language learners, on the Internet there is such a vast number of poetry-related resources that it is not too difficult to find appropriate texts in terms of vocabulary, syntax, extension and so on. There are websites, such as Poetry Foundation (2025) and Hello Poetry (n.d.), which allow users to refine their search using different filters, something that can help teachers find whatever materials they need for any specific purpose. In addition to its availability, contemporary poetry in English is also extremely diverse when it comes to country of origin, the authors' gender and race, the historical context, etc. Thus, this extreme diversity of voices in poetry can help learners see cultural differences as a blessing and value their own cultural backgrounds in relation to other people's origins. In other words, it can foster their intercultural competence.

Finally, a more traditional reason for making poetry part of EFL instruction is that it can be used as a material to create form-focused exercises, as it is often done in textbooks using newspaper articles and short stories. In a study carried out by Gómez Rodríguez (2018), a group of students worked with a number of poetry-based exercises in a university EFL context. Among other aspects, the author assessed their improvement in language skills after the poetry implementation, and the results were rather positive. In the conclusions section, he states that, "according to the findings, poetry enhanced their [i.e. students'] lexis" (2018, p. 374). This is relevant for the present study, as it implies that, apart from developing the students' intercultural and communicative competences, poems may be a powerful tool for the improvement of their vocabulary as well.

2.1.3 Two case studies on using literature in EFL

A relevant study that connects the intercultural competence with literature is the one conducted by Rodríguez and Puyal (2012). In it, they implemented a teaching unit based on literature—which included both narrative and poetry—to show how this material could contribute to students' intercultural competence. After implementing the unit and analysing the results, the authors concluded that literature is a powerful resource for developing this specific competence (Rodríguez & Puyal, 2012, p. 118). As they write, "the fact that the students' perception of their own linguistic, communicative and intercultural competence had improved supports our initial idea that the use of literary texts in an English Language classroom can enrich students from an intellectual and affective perspective" (2012, pp. 118–119). In other words, when students are exposed to other cultures and ideas through

the experience of literature, they relativise their own values and, in consequence, become more empathetic and open-minded.

Rodríguez and Puyal (2012) also collected data regarding their students' personal views about using literature in EFL. According to the results, most learners expressed a high level of enjoyment using the material. In fact, "the majority of students agreed and strongly agreed with all 25 items, which confirms the value they place on [literary] activities" (p. 115). This response is quite relevant, because, if students had a positive subjective experience while reading narrative and poetry, this implies that their motivation must have been quite high, which is a fundamental factor in SLA. The above-mentioned reasons for including poetry (its short extension, its aesthetic qualities, etc.) are certainly relevant, yet the students' views on the experience of using this material are equally important. This is precisely why the present thesis, apart from considering the use of poetry for the improvement of vocabulary, will also take into account student perceptions. As it will be specified in the next section, I have employed a perception questionnaire to find out what learners thought about the enjoyableness and effectiveness of the poetry-based activity.

A second researcher who has explored the potential of poems in EFL teaching is the already-mentioned Gómez Rodríguez. In his study (2018), several poems were included as teaching materials, and he used perception questionnaires as well as written essays to research the students' personal views. As in Rodríguez and Puyal's study (2012), Gómez Rodríguez's results are quite encouraging, because "Colombian learners changed their minds about the widespread idea that poetry was unsuitable for their EFL learning" (2018, p. 375); in fact, "despite the challenges they encountered (. . .), they enjoyed poetry and gave positive comments about its inclusion in language learning" (2018, p. 375). The researcher concludes that, considering his students' positive response, teachers should try to include poetry as a complement to more traditional materials like textbooks and videos (2018, p. 374).

2.1.4 The drawbacks of including poetry in EFL and some possible solutions

After this description of the main reasons in favour of using poetry, I will return to the three arguments against using literature enumerated at the beginning of this section. The first reason refers to the "structural complexity and unique use of language" (McKay, 1982, p. 529) of literature, which are even more marked in the case of poetry. As it has been stated, though this unique use of language can certainly be problematic, teachers can use it to develop their students' critical thinking skills. Furthermore, they can use its structural complexity to their own advantage to have students focus on the forms that appear in each specific poem. Khatib et al. (2011) do acknowledge that "poems are usually written in a form deviant from the norms of speaking or even writing" (p. 203-204), though they also argue that "this complexity itself can become a source for practice" (p. 204), especially for more advanced learners. In this way, poetry could be used to create exercises to practise the vocabulary and grammar aspects present in each specific text.

The second reason is that literature is not a good material to help learners achieve their academic and occupational goals (McKay, 1982, p. 529). This may be the case with English for specific purposes, in which learners are expected to focus on the language necessary for a specific profession, but in most other cases it is not true. If we are teaching teenagers or adult learners in a more generic context, literary texts can play a role in the learning process. As it has been argued, this text type can foster intercultural awareness and aesthetic appreciation, something more traditional materials cannot always do successfully. Furthermore, it is evident that the inclusion of literature in the classroom does not imply excluding other textual genres; EFL classes should include materials from as many different sources as possible, as students learn better when they experience language in different forms. In consequence, being part of the cultural heritage of the English-speaking world, poetry should be included in EFL just like any other material.

The last reason is that, "on a conceptual level, it may be quite difficult for students" (McKay, 1982, p. 529) because it is a culturally specific product. As it has been argued in the previous subsection, this specificity can be seen as a learning opportunity. The fact that literature can reflect very different cultural contexts is an opportunity to

raise the students' intercultural awareness by exposing them to influences beyond traditional materials. Nowadays, poetry is being written in English in many countries all over the world. Therefore, learners need not be confined to the USA and the UK. Of course, introducing literature produced in countries such as India and South Africa will probably lead to comprehension difficulties, but this can also become a motivating factor for students.

If we step back and look at these three drawbacks together, there is one possible common solution, that is, text selection. Any problems related to excessive complexity can be tackled by choosing the adequate poems for each level of language proficiency. In this vein, Khatib (2011) states that, in his opinion, "the major reasons of students' failure in understanding and appreciating English poetry are (a) inappropriate selection of texts and (b) ineffective teaching methodology" (p. 165). Similarly, McKay (1982) writes that "the key to success in using literature in the ESL class seems to me to rest in the literary works that are selected" (p. 531). Although she mentions that a frequent solution is "the simplification of the text" (p. 531), she argues that this is not desirable, because altering the original text may reduce its cohesion, which is precisely one of the most interesting aspects of using literary materials. If modifying the texts is not an option, the only remaining possibility is using simple texts as they are, without any adaptations. This is why McKay (1982) recommends choosing literature aimed at young adults, because these texts deal with relatable themes like "the problem of personal growth and development" (p. 532), and because they "tend to be stylistically less complex" (p. 532). Although finding poetry written for young adult readers may not be easy, there is a great deal of poetry aimed at children. Nursery rhymes, for instance, may be appropriate and even culturally relevant for lower proficiency levels.

In sum, teachers who want to use poetry in the EFL class should find works that are not too complex—lexically, semantically and syntactically—, because readable texts may be the first step in learning to appreciate literature. If the chosen poems are too difficult, learners will probably feel discouraged and may even develop a dislike for anything connected to literature. Conversely, if the chosen texts are excessively simple, learners will obtain few linguistic benefits from the reading experience. It is therefore essential to strike a balance between these two extremes.

2.2 Vocabulary in SLA

2.2.1 The three main areas in vocabulary research

The present paper, apart from focusing on poetry, will also be concerned with lexical learning. Thus, this subsection will explore some theories of vocabulary studies in connection to SLA. In order to better understand what vocabulary is, a good starting point may be the Oxford English Dictionary (2024). This source lists as many as 11 different definitions for the word *vocabulary.* The one that best captures the meaning of the word as it will be used in this thesis is the following: "the words and phrases in a foreign language taught or used as part of a language-learning course or exercise" (Oxford English Dictionary, 2024). I have chosen this particular entry because, though some of the other definitions are also appropriate, this one makes reference to lexical acquisition within foreign language learning. Furthermore, it acknowledges that words and phrases may be "taught", but also "used"; that is, it considers both the direct instruction of vocabulary and the use that learners make of it.

According to Webb (2020), all research that has been carried out on vocabulary so far falls into three different categories: pedagogy, description, and assessment. Though the three categories are relevant in EFL, it seems reasonable to expect teachers to be interested primarily in the first one; in other words, teachers will probably be concerned with the pedagogical implications that research has for classroom activities. As Nation (2020) writes, "the main reason for a teacher to be interested in what is involved in knowing a word is so that the focus and balance of a language course ensures the development of well-rounded, usable vocabulary knowledge" (p.

15). In other words, for a teacher, being concerned with research is justified as long as it is applied in some way to improve their learners' use of vocabulary.

The other two categories of vocabulary research are description and assessment. Description consists mainly in "understanding what is involved in knowing a word or sequence of words, (. . .) how vocabulary fits into existing theories of language learning" (Webb, 2020, p. 4) and so on, and it is the basis of most vocabulary research. Assessment is connected to the different methods that are used to measure how much vocabulary learners know (and how well they know it). In EFL, vocabulary tests can "indicate the extent of lexical development within a course (. . .) and reveal which words students know and which words they need to learn" (2020, p. 4). Consequently, assessment research can also inform how vocabulary is taught in the classroom. Having explained the three main areas of vocabulary research, I will now deal with a few concepts that are particularly relevant for vocabulary learning within the field of EFL.

2.2.2 Nation's Four Strands and language-focused learning

Within vocabulary studies, the most widely recognized name is probably that of Paul Nation. His volume *Learning Vocabulary in Another Language* (2013) is considered to be "the most important work on the subject to date" (Webb, 2020, p. 4). Throughout his career, Nation has researched vocabulary learning from numerous points of view, though he is particularly well-known for his division of language learning into four strands (2007). According to him, in order to plan how learners should spend their time, curriculum developers and teachers should consider four distinct strands, which are the following: (1) meaning-focused input, (2) meaning-focused output, (3) language-focused learning, and (4) fluency development. Nation (2007) believes that foreign language courses should "provide a roughly equal balance of the four strands" (p. 11), though he also remarks that both EFL teachers and students often consider the third strand, which focuses on language forms, to be more important. This is because it allows instructors to target specific linguistic features and to explain them, something that becomes even more evident when they are teaching vocabulary explicitly. Nation writes that "there is plenty of evidence, *certainly in vocabulary learning*, that deliberate learning can make a very useful contribution to a learner's language proficiency" (2007, p. 6; emphasis added). In other words, direct vocabulary instruction can be part of EFL classes, though teachers should not forget about the other three strands. Otherwise, their students will spend too much time developing their knowledge on linguistic forms, and not enough on fluency, extensive reading, and other necessary skills.

2.2.3 Nation's classification of the different aspects of vocabulary knowledge

Nation has also researched what he calls "vocabulary knowledge". In the chapter "The Different Aspects of Vocabulary Knowledge" (2020), he explains what we mean when we say we know a word, and divides this knowledge into nine different categories or aspects. He presents these in a table, which has been reproduced below (see **Table 1**). Before dealing with the table, however, I will present some concepts that are relevant in relation to vocabulary acquisition: receptive and productive knowledge, vocabulary breadth and depth, intentional and incidental learning, and the learning burden.

As it may be seen, each of the nine aspects in the table is subdivided into receptive (R) and productive (P) knowledge. The former is "the kind of knowledge needed for listening and reading" (Nation, 2020, p. 16), and it requires, among other things, to recall the meaning that corresponds to a specific linguistic form, be it written or spoken. Productive knowledge, on the other hand, is necessary to speak or write; we use it to recall the word form that is linked to whatever meaning we want to convey. This distinction is relevant in teaching because using vocabulary requires a combination of both receptive and productive word knowledge. In addition, teachers should remember that, according to research, "it seems that receptive learning and use is easier than productive learning and use" (Nation, 2013, p. 50), which means that learners will have to practice more to develop their productive skills.

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognisable in this word?
		P	What word parts are needed to express the meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	constraints on use (register, frequency ...)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

Note: In column 3, R = receptive knowledge, P = productive knowledge.

Table 1. The nine aspects of word knowledge according to Nation (2013, p. 49)

Another distinction that is closely linked to vocabulary knowledge is breadth and depth. Breadth of vocabulary knowledge corresponds to the quantity of words that a language learner knows (Webb, 2020, p. 6), even if these have not been entirely mastered. When we say someone has a good vocabulary, we often mean that they know the form-meaning link for a lot of lexical items. In other words, we are referring to their vocabulary breadth. Depth, on the other hand, corresponds to the quality of the knowledge; if a learner has a very deep vocabulary knowledge, this means they "know aspects of knowledge, such as collocation, multiple meanings of words, and derivations" (Webb, 2020, p. 6). Both types are certainly important for EFL teachers and learners, but it is frequent to consider only breadth, while forgetting that vocabulary knowledge involves many other interrelated aspects. Additionally, Nation (2020) points out that depth of knowledge "gradually develops in a variety of ways and *teaching is only one of the contributors* to this knowledge" (p. 26; italics added). Put another way, though explicit instruction does help acquire lexis, some aspects of vocabulary knowledge will have to be developed outside the classroom. As Schmitt (2017) writes, teachers need "both *explicit instruction*, to fast-track the initial learning of vocabulary, and *large amounts of exposure*, (. . .) to cover the word knowledge aspects that we cannot teach" (p. 6).

The latter distinction corresponds to two other concepts that are relevant in vocabulary studies: intentional and incidental learning. This distinction is particularly important for teachers because, depending on the aspect of word knowledge that they want to develop, one approach will be more effective than the other; as Schmitt (2017) states "some word knowledge aspects, such as the form-meaning link, are relatively responsive to explicit teaching and intentional learning" (p. 6). Others, however, like collocation knowledge, are "less responsive to explicit attention" and more adequately developed through incidental learning (p. 6).

Finally, another of Nation's contributions to lexical studies is the notion of the learning burden, which measures how much effort a learner has to make to learn a specific lexical item (Nation, 2013, p. 44). Its most fundamental principle is quite intuitive: "the more a word represents patterns and knowledge that the learners are already familiar with, the lighter its learning burden" (Nation, 2013, pp. 44–45); therefore, a considerable similarity between the L2 word to be learned and its L1 counterpart will make it easier to acquire. This is especially relevant in the context of EFL, in which instructors should be able to estimate the burden of the vocabulary their students are working on. This will allow teachers to reduce the learning burden "by drawing attention to systematic patterns" and "by pointing out connections between the second language and the first language" (Nation, 2013, p. 45).

I will now move on to some of the aspects that constitute word knowledge according to Nation (2020). In the first column of **Table 1**, these aspects are classified into three groups: form, meaning and use. When we think about learning vocabulary, it is common to think exclusively about the former two, though use is equally important for most students. Still, there are some exceptions to this, and, for certain learners, learning about form and meaning may be sufficient to cover their needs. According to Schmitt (2017), "if [learners] only want to read or listen, then the form-meaning link might be enough" (p. 5), so perhaps we need not worry if they do not learn the aspects related to use. The students who want to communicate fluently in English (i.e. achieve productive use), nonetheless, will need to develop as many aspects of their knowledge as possible. This is because they will have "to recall a word that is appropriate not only for the desired meaning, but also for the particular grammatical, collocational and contextual situation" (Schmitt, 2017, p. 5).

As it may be seen in **Table 1**, the first aspect of use is being aware of the grammatical functions of a certain lexical item. Nation (2020, p. 20) admits that, in some cases, grammatical knowledge may overlap with collocational knowledge, though they are not exactly the same. If we take a verb such as *like*, some of the grammatical facts about it that a learner should know are its part of speech, that it is used transitively, that it is sometimes followed by a verb, and that this verb may be in its infinitive form (*I like to sing*) or in the present participle form (*I like singing*). Teaching these facts may help learners self-monitor their vocabulary use, though "it is likely that the systematic knowledge of grammatical features which underlies normal language use is ultimately the result of incidental learning from large amounts of meaningful comprehensible input" (Nation, 2020, p. 20). Hence, the importance of extensive listening and reading in order to internalize the aspects of word knowledge that are more intuitive, although Nation (2020) concedes that memorizing one or two example sentences that are representative of the most common uses of a word may be a useful learning strategy technique (p. 21).

The second aspect is collocational knowledge, which refers to the behaviour of words in contact with other words, and which is an essential part of vocabulary knowledge. In collocations, "the meanings of the parts make an obvious contribution to the meaning of the whole" (Nation, 2020, p. 21) and the combination of these parts follows patterns that "are not arbitrary groupings of words but are typically regular predictable combinations" (Nation, 2020, p. 21). Therefore, looking at the component parts, rather than simply memorizing the combinations mechanically, can be a good teaching/learning strategy.

A possible classification of collocations is Grant and Nation's (2006, pp. 8–10), who divide them into core idioms, figuratives and literal sequences. Each of these types has different teaching strategies. Core idioms are the most opaque category, as we cannot deduce what they mean by looking at the individual words (as is the case of *by and large*). A good technique for learning what they mean is creating false etymologies, "such as *cats and dogs* in the expression *raining cats and dogs*" (Nation, 2020, p. 21). Figurative collocations are less opaque in most cases, and they belong to a specific topic area, in which they originated with a literal sense; the original meaning of *play it close to the vest*, for example, is connected to playing cards. These collocations can be learned "using the obvious strategy of relating the literal meaning to the figurative meaning" (Nation, 2020, p. 21). Finally, literals are usually quite transparent, so they are easily understood. According to Nation (2020), students can learn collocations such as *like ice cream* incidentally, which means that, in the case of literals, instruction is more dispensable than in other cases.

Yet another way of thinking about collocational knowledge is what is often called formulaic language, which includes idioms, phrasal verbs and collocations (Schmitt, 2017). It has been calculated that "between one third and a half of all English discourse consists of formulaic language" (Schmitt, 2017, p. 6). This implies that, for learners, learning formulaic language in chunks can be extremely useful, because "this makes production easier, as students merely need to remember a formulaic phrase" (Schmitt, 2017, p. 6), rather than having to process language word by word.

The last aspect of word knowledge listed in **Table 1** is related to the pragmatic constraints on the use of certain words. There is a considerable number of words that have restrictions in where, when and how native speakers tend to use them. These are connected to politeness, datedness, frequency of use and oral vs. written register,

among other restrictions (Nation, 2020, p. 22). Teachers should be aware that the knowledge of these constraints is quite difficult to teach deliberately; Nation (2020) recommends providing feedback, and says that some words should be learned with a "warning"—something that is especially true of swear words, which should not be used in many situations.

These are the three knowledge aspects related to how lexical items are used in context. Because they are quite advanced, it usually takes students longer to develop them. Nevertheless, it is worth remembering that "the various aspects of word knowledge are not equally important" (Nation, 2020, p. 25). For beginners, the spoken form and the form-meaning link should be enough, while "other aspects of knowledge can become focuses of attention as proficiency develops" (Nation, 2020, p. 25). Put another way, word associations, collocation knowledge and constraints on use should not be a priority until students have reached a certain linguistic competence.

2.2.4 Is the direct teaching of vocabulary useless?: Nation's perspective

Another relevant issue that connects EFL and lexical studies is how useful it is to teach vocabulary in the classroom, something Nation (2021) has also addressed. In his view, spending a lot of time on direct vocabulary instruction is not very efficient, at least in comparison with other tasks the teacher should be concerned with. He defines these tasks as—in order of importance—"to plan, organize, train, test, and teach" (Nation, 2021, p. 2). However, it is worth noting that Nation's (2021) definition of teaching is quite narrow, as he believes that "teaching occurs when the teacher is the source of information, (. . .) the focus of attention, and (. . .) determines the pace of the learning" (p. 2). This excludes situations that might also be considered teaching—such as using flashcards—, which he defines as vocabulary learning. We should not conclude that teachers must never spend time on vocabulary; however, it is true that good language teachers, apart from teaching vocabulary explicitly, should be concerned about "supporting learners in taking control of their own learning" and about helping them develop learning strategies like "word card use, using word parts, and guessing from context" (Nation, 2021, p. 8).

The four main reasons that Nation (2021) cites for the limited usefulness of direct vocabulary teaching are the following: the size of the task, the rate of teaching, the amount of learning, and the efficiency of word card learning (pp. 3-4). According to Nation (2006), in order to read a text comfortably, we need to reach a lexical coverage of 98% of the words. This means that "a vocabulary of 8,000 to 9,000 words is needed to read a novel, and even then, 1 word in 50 will be unfamiliar" (Grant & Nation, 2006, p. 74). In other words, the vocabulary that students learn cannot come only from explicit teaching because, if this were the case, it would take an extremely long time to have the necessary vocabulary to read and understand texts.

Even though Nation's ideas need not be taken as the absolute truth, there is some value in what he proposes. In his ideal vocabulary learning program, extensive reading and listening would have an essential role; speaking and writing activities would also have to contribute to vocabulary acquisition, and fluency development of vocabulary use would be essential too. Although the teacher "would do small bits of teaching", he or she would "largely be involved in planning, organizing, training, and monitoring" (Nation, 2021, p. 8).

Despite all that has been said so far, there are also arguments in favour of direct vocabulary instruction. According to Nation (2013, p. 94), the number of high-frequency words in English is between 2,000 and 3,000, so "it is practical and feasible to directly teach a substantial number of them" (p. 94). Furthermore, EFL learners rarely encounter input that corresponds to their proficiency level in everyday communicative situations; therefore, Nation (2013) believes that direct vocabulary teaching could "bridge the gap between second language learners' present proficiency level and the proficiency level needed to learn from unsimplified input" (p. 94). When this level has been reached, learners can start learning vocabulary by using the unmodified input that is available to them through books, streaming platforms, the Internet and so forth.

2.2.5 Learning vocabulary: input type, word lists and modality

Having talked about the relative usefulness of teaching vocabulary, I will now deal with SLA research so far and how it explains some aspects of vocabulary acquisition. Knowledge about this research area is relevant, as it can help language instructors in choosing the most adequate type of input for their learners, or in teaching vocabulary effectively in explicit instruction.

As for the choice in the kinds of input that students should be exposed to, according to Kormos (2020), all kinds of input (digital or paper-based, authentic or adapted, spoken or written, etc.) can help learners, "but their effectiveness can vary depending on the target vocabulary, students' level of proficiency, existing lexical knowledge, and the instructional context" (p. 208). Thus, depending on the context, one type of input may be more effective than another. In the case of vocabulary teaching, two types of form-focused teaching are especially relevant: intentional learning and explicit instructional contexts. These two teaching methodologies are often performed together. That is, intentional learning, which is dependent on "the individual's motivation to learn" (Kormos, 2020, p. 208), is often paired with explicit instructional contexts, which are "associated with a high level of attention control, targeted and sustained attentional focus" (Kormos, 2020, p. 208). This implies that, if students do not have the motivation to focus on the materials they are given, these two types of instruction will be ineffective.

Another frequent method used to foster explicit vocabulary learning are word lists. Kormos (2020, p. 208) groups them in two categories: decontextualized lists, which include the definitions of the words or the L1 counterparts, and contextualized lists, which are integrated into sentences. It is worth noting that, according to research by Qian (2002) decontextualized word lists result in a higher retention of vocabulary and contextualized ones. Kormos (2020) believes that this may be because, in early stages of learning EFL, having to process the context of a new word in a list may be more demanding cognitively than learning it in isolation; however, she is unsure if "depth of word knowledge, such as syntactic information, collocations, and associations, can be acquired if L2 learners encounter words devoid of context" (p. 208). This seems to indicate that both contextualized and decontextualized word lists have their uses.

Vocabulary input can also be presented by means of data-driven learning (DDL) approaches. These follow usage-based theories of SLA, "which argue that learning happens through the probabilistic extraction of regularities from the input" (Kormos, 2020, p. 209) and not from the internalization of explicit rules. DDL approaches follow a procedure similar to input enhancement, "whereby the salience of a lexical item is increased by both providing several examples and highlighting them in some way" (Kormos, 2020, p. 209). This directs the learners' attention to the targeted words, so that they can notice usage patterns. This type of enriched input is particularly useful "in developing not only form-meaning links but also depth of lexical knowledge" (Kormos, 2020, p. 209), which means it can help learners improve their collocational knowledge. Nevertheless, we should take into account that DDL approaches may require teachers to spend more time creating materials than traditional word lists, which are less time-consuming.

Another way of classifying input is modality, as materials may be presented as audio, as written text or as both at the same time—that is, as multimodal texts. The modality of the materials is an important factor for vocabulary learning. According to research, it seems that the reading mode is more effective in terms of vocabulary retention, especially due to working-memory limitations on the listening mode (Kormos, 2020, p. 210). When we listen to a text, "language processing usually takes place under time pressure" (Kormos, 2020, p. 210), and, therefore, we have little time to reflect on the vocabulary. On the other hand, when we read, we can focus on any new lexical items, which makes retention easier. Brown et al.'s (2008) study compared the three input modes—reading only, reading-while-listening and listening only—in relation to incidental vocabulary learning. The results showed that, in order to learn vocabulary in the listening mode, learners need to encounter the new words more times than in the reading mode. The researchers found it "rather obvious that the listening-only mode should be the most difficult to acquire new vocabulary from" (Brown et al., 2008, p. 148); moreover, it is worth noting that "when asked which input mode they preferred, 0% of the subjects chose listening-only" (Brown et al., 2008, p. 148).

In the mentioned study by Brown et al. (2008), the researchers measured incidental learning of new words with each of the three modes of input. The test scores for reading and reading-while-listening were almost the same, which seems to indicate that these two modalities are equally effective for incidental vocabulary learning. This is why Brown et al. (2008) argue that, when it comes to these two modes, "the selection of preferred input mode should rest with the learner" (p. 148). When it comes to multimodal input, however, it is also true that it can contribute to the retention of information, as language "is processed in both visual and auditory working memory" (Kormos, 2020, p. 211). This creates stronger connections due to a phenomenon that, in educational psychology, is called verbal redundancy (Moreno & Mayer, 2002, p. 156). Therefore, multimodal input may be a useful teaching resource if we want to enhance our students' lexical retention.

2.2.6 Laufer and Hulstijn's (2001) construct of involvement

Another relevant concept within vocabulary research is the construct of involvement, which was developed by Laufer and Hulstijn (2001). This construct can be a useful way for teachers to conceptualise vocabulary learning. In fact, it was created in order to explain how foreign language learners can retain new lexical items more effectively. According to Laufer and Hulstijn (2001), the construct of involvement "can explain and predict learners' success in the retention of hitherto unfamiliar words" (p. 14). In other words, the more involved learners are during incidental tasks, the better the results of the learning process. Involvement can be explained as the result of three factors, need, search and evaluation, which will be described in the next paragraph.

Need is connected to motivation, and it happens when a learner wants to find out the meaning of a certain item. For instance, when "the learner is reading a text and an unknown word is absolutely necessary for comprehension, s/he will experience the need to understand it" (p. 14). Thus, it is the first step for vocabulary learning. As for search, it is the cognitive process of trying to find what a given form means or the form necessary to express a particular meaning. Therefore, search implies an action, such as looking up a word in a dictionary, or asking the teacher. The last factor (evaluation) takes place when a learner needs to choose between different words depending on what the context requires. This happens, for example, if a learner is doing a writing activity and looks up an L1 word in a bilingual dictionary, and they find three possible translations; the learner will then have to compare the three forms and guess which one is more appropriate in the context. This process of comparison is what Laufer and Hulstijn (2001) call evaluation, and it always "implies some kind of selective decision based on a criterion of semantic and formal appropriateness (fit) of the word and its context" (p. 15).

The construct of involvement has contributed to the investigation of how vocabulary is acquired. As Nation (2013) writes, "the research on the Involvement Load Hypothesis has been very productive and has shown the value of such an analytical approach to evaluating teaching activities" (p. 101). However, it has also been questioned because of its simplicity. The problem is that, "with only three categories, it does not include a lot of the factors that we know are important for vocabulary learning" (Nation, 2013, p. 100). Other factors that may also predict vocabulary acquisition are "awareness, negotiation, repetition, spacing of retrievals, creative use, interference and imaging" (p. 100). In spite of its simplicity, the concept of involvement can be a convenient framework for teachers, especially for assessing vocabulary activities. Overall, instructors should try to increase their students' need to find the meaning of new words; to provide tools to find these meanings; and to help them evaluate the different lexical options that are available.

2.2.7 Multiword items: collocations, idioms and phrasal verbs

Having dealt with different aspects of vocabulary studies within SLA research, I will now move on to one of the most essential types of lexical items: multiword items (MWIs). This category is often defined in opposition to single-word items, which are more commonly referred to simply as words. When language learners think about vocabulary acquisition, they usually think about looking up single-word items in the dictionary. However, students will have to face MWIs if they want to develop their lexical knowledge. Even though individual words are certainly relevant, "research has also shown that language also consists of large amounts of multi-word phrasal

vocabulary", and that "between one third and a half of all English discourse consists of formulaic language" (Schmitt, 2017, p. 6). Because of this, and because MWIs are particularly difficult to acquire, teachers and learners will have to deal with them at some point, whether they like it or not.

According to Wolter (2020), the "lack of consistency in terminology describing MWIs" (p. 493) is one of the reasons why they are often omitted in language courses. Even though classifications are not necessarily useful for learners, in this case they may help, as each type of MWI behaves differently depending on its context. That is to say, "different MWIs will display different qualities that often lead to a unique set of challenges for learners" (Wolter, 2020, p. 495). In general, there are "three types of MWIs that seem essential for L2 learners to achieve even modest levels of mastery in English" (Wolter, 2020, p. 495), which are collocations, idioms and phrasal verbs. I will now explain each of them and specify the difficulties they pose for learners.

The first type of MWIs, i.e. collocations, is not easy to define. A possible criterion for doing so is frequency of co-occurrence with other words, which requires having large corpora of lexical items available. Following this criterion, we consider a combination of words a collocation when "a particular word co-occurs with another word (or set of words) much more frequently than we would expect given that word's overall frequency" (Wolter, 2020, p. 498). Another definition is simply to say that, even though words can be combined in numerous ways to express the same ideas, native speakers will not judge all the combinations as correct; in fact, "only a small subset of these will be seen as acceptable" (Wolter, 2020, p. 498), while the rest, even if it is possible, will sound artificial.

When it comes to acquiring collocational knowledge, Wolter (2020) mentions different learning strategies that can help overcome the problem of their "pervasiveness and variability across registers" (p. 500). One of these strategies is choosing a limited number of collocations and exposing students to them repeatedly in one text. This leads to incidental learning, though intentional learning, in the case of collocations, is also possible, as "most studies have found that learners do show gains in collocational knowledge from explicit instruction" (Wolter, 2020, p. 501). Still, it must be acknowledged that this knowledge cannot come exclusively from deliberate teaching, especially because of the vast number of items to be acquired. In Wolter's (2020) view, the most effective approach would be using "carefully structured incidental reading activities, explicit instruction, or perhaps a combination of the two" (p. 503).

The second type of MWIs are idioms, and their main characteristic is that they vary in two areas: idiomaticity and fixedness. The former refers to the fact that, when learners encounter new idioms, their meaning is sometimes difficult to deduce by analysing their parts one by one, as it can be seen with examples like *between the hammer and the anvil*. It is worth observing that, as seen above, Grant and Nation (2006, p. 7) considered core idioms to be simply a specific kind of collocation, while Wolter (2020) sees collocations and idioms as two distinct types of MWIs. This shows how labels like MWIs, collocations and idioms often overlap, and to what extent this can end up confusing learners.

Fixedness refers to the accepted variability in the wording of a certain idiom, as some idioms are completely fixed, while others allow for some modifications. Wolter (2020, p. 504) provides the example *between a rock and a hard place*, which is totally fixed; if someone said *between a stone and a hard place*, the expression would sound unnatural to most native speakers. Other idioms are less fixed, and speakers can choose between certain possible words, as it happens with *a pain in the neck/rear* (but not *in the nose*). When teachers introduce new idioms, it may be a good idea to provide information about their level of fixedness.

Apart from idiomaticity and fixedness, which make idioms particularly difficult to learn, they present two additional problems: their number and their cultural specificity. According to Jackendoff (1997, as cited in Wolter, 2020, p. 504), in the English language there are at least 25,000 idioms. In a regular EFL course, teaching all of them would be impossible, which means that instructors and materials designers should focus on "uncovering which idioms are essential for learners and which can be left to later" (Wolter, 2020, p. 504). In addition, almost all idioms are connected in some way to the culture where they appeared, and they are often rarely used in other

parts of the English-speaking world. Though this is a difficulty, it can also make the learning process more interesting, as the teacher can take advantage of this specificity to talk about cultural differences.

As for the teaching of idioms, there is some debate about whether they should be taught integrated in meaningful contexts or in isolation. In general, "it seems that (. . .) approaches which present idioms in highly contextualized materials are likely to be more effective than those that rely on isolated presentation" (Wolter, 2020, p. 505). Still, in some cases they are classified according to their topic and presented in isolated lists, which may have the advantage of aiding the retention of idioms connected to the same semantic field.

The third type of MWIs are phrasal verbs (PVs), which are one of the most challenging vocabulary items for EFL learners, and which "many nonnative English speakers actually avoid using (. . .) altogether" (Gardner & Davies, 2007, p. 340). Wolter (2020) cites different reasons why PVs pose so many problems for students, the most important ones being that they are pervasive, extremely polysemous, and syntactically complex (pp. 495-496).

Regarding their pervasiveness, it is interesting to look at a study by Gardner and Davies (2007) in which the British National Corpus (BNC) was analysed to gain insight into the number of PVs in English and how they behave. By looking at the number of adverbial particles in the corpus, they concluded that "learners will encounter, on average, one [PV] in every 150 words of English they are exposed to, or roughly 2 per average page of written text" (Gardner & Davies, 2007, p. 347). Taking into account these numbers, we can say that phrasal verb constructions "constitute a major grammatical class" (2007, p. 347).

Another learning difficulty of PVs is their polysemous nature, as most PVs have more than one meaning, and some of them can be quite opaque in relation to their form. Gardner and Davies' (2007) found that, in the corpus they had analysed, each PV had on average 5.6 different meanings (p. 353). This explains why learning that the PV *take up* means "to pick up" may not always help learners, as it can have other, more opaque meanings—such as in the sentence *He took up the accordion at 60*. In consequence, "finding ways to reduce the learning burden can help a good deal with phrasal verbs" (Wolter, 2020, p. 497). As in the case of idioms, this would require finding the most frequent PVs; however, because of their polysemy, it would also be necessary to determine the most frequent meanings for each of these PVs. This is where frequency lists can be of great help (more on this in the next subsection).

The last difficulty of PVs is their syntax, especially regarding the movement of the particle, which may be optional, necessary, or—in the case of prepositional verbs—incorrect. In the case of transitive PVs, the particle can be moved after the object, a decision that is usually optional (both possibilities are grammatical); nevertheless, if the object is a personal pronoun, this movement becomes grammatically necessary. In the case of prepositional verbs, however, moving the preposition results in ungrammatical constructions and, in consequence, is not permitted. This confirms the idea that learning the form-meaning links is not enough if students want to use PVs productively.

As for the teaching techniques that work better with PVs, research is not clear about which ones should be used. Wolter (2020) claims that, in the two studies he reviewed, "it is difficult to conclude that either approach was particularly effective at teaching phrasal verbs" (p. 498). It may simply be that spending time focusing on PVs increases students' knowledge of these items, and not that any given teaching strategy is preferable. Still, bearing in mind their relevance and complexity, Wolter (2020) concludes that, in spite of the lack of conclusive research, "it would seem prudent to include phrasal verbs" (p. 498) in language courses.

2.2.8 Frequency lists: what lexical items should we teach?

As stated above, frequency lists are a useful tool for both instructors and learners, especially when it comes to PVs, which are ubiquitous and quite polysemous. The mentioned study by Gardner and Davies (2007) was one of the first deliberate attempts at creating an inventory of the most pedagogically relevant PVs. This list was improved by Liu (2011), who performed a multi-corpus analysis of the frequency of PVs in American and British English. Although in both analyses "PV frequency and ranking order was provided", a major problem of the list

was that they lacked "semantic information, especially in the case of polysemous items" (Garnier & Schmitt, 2015a, p. 4). Therefore, teachers and learners who use these frequency lists have no way of knowing which meanings of each PV are more useful, and they have to either learn all the meanings or rely on their intuition to choose which ones should be learned, which is not always a good criterion. This may lead them to the feeling that teaching/learning PVs is such a formidable task that it is not worth the effort, as there are too many meanings to be learned. To solve this problem, Garnier and Schmitt (2015a) developed the Phrasal Verb Pedagogical (PHaVE) List. Their main aim was "to narrow the scope of meaning senses of the most frequent PVs to be acquired" (p. 4) taking into account how frequently they appear in the Corpus of Contemporary American English. From the beginning, their corpus analysis was aimed at improving EFL instruction, material design and testing, and this is what makes the list such a fundamental tool for teachers.

The latter list consists of 150 PVs, which are the same that Liu (2011) included in his earlier list. While it is true that the PHaVE list covers a small number of items, especially taking into account how pervasive PVs are, Garnier and Schmitt (2015a) argue that learning these 150 items is already very positive for learners. The PVs they include "already cover 62.95% of all the total 512,305 PV occurrences in the BNC" (p. 18). In addition, as it is, the PHaVE list is already 38 pages long so, if it were longer, it might not be realistic to expect teachers/learners to use it. Therefore, a limited list is actually more efficient than a more extensive one.

As for the organisation of the list, the 150 entries are ordered from the most to the least frequent PVs in the analysed corpus. In each case, the user finds from one to four meaning senses, also ordered by frequency. In addition, each sense comes with a percentage "representing the relative frequency of that meaning sense in the total number of occurrences of the word" (Garnier & Schmitt, 2015a, p. 22), and also with an example sentence. The 30th entry in the list, for instance, is *stand up*. Its first meaning is "rise to a standing position after sitting or lying down" (2015a, p. 34), and its frequency percentage is 67.5%. However, its second meaning ("make public knowledge a privately held position" (2015a, p. 34) has a coverage of only 11%. The implication for teachers, therefore, is that the former sense will have to be taught earlier, because it will be more useful for both receptive and productive use. The authors decided to omit any senses with an occurrence percentage under 10% because they considered they "are not worth prioritizing for explicit attention" (Garnier & Schmitt, 2015a, p. 22). Overall, the PHaVE list is a very convenient resource for teachers, as it can help them decide what PVs should be prioritised in direct instruction, and which ones are dispensable. Learners, too, may benefit from using the list, especially if they are interested in deliberately developing their PV knowledge.

2.2.9 Remembering vocabulary: noticing, retrieval and creative use

Now that MWIs and frequency lists have been presented, I will turn to the topics of direct instruction and how vocabulary can be remembered most effectively. As it has been shown above, Nation (2021) contends that teaching vocabulary, while not entirely useless, is not always a justified investment of class time. If we go back to the four strands, as he argues, "direct vocabulary instruction is only one part of one of the four strands of a well-balanced course" (2013, p. 94), which is the language-focused learning strand.

One of the issues in teaching vocabulary is that teachers often perceive form-focused instruction as the most effective way of helping students develop their lexicon, since it "involves a deliberate focus on a particular learning goal and can draw on a teacher's specialist linguistic knowledge" (2013, p. 95). Therefore, it is not unusual for teachers to spend more time on the language-focused strand than on fluency development or on meaning-focused input/output, which are equally important. According to Nation (2013), though teaching vocabulary does matter, the instructor's goals should be limited to "giving attention to only high-frequency words, focusing on only the most important aspects of knowing a word, and not spending much time on each word" (p. 95).

A rather useful framework for thinking about how vocabulary should be taught are the three cognitive processes that, according to Nation, are involved in retaining new lexicon: noticing, retrieval and creative use (Nation, 2013, p. 102). Noticing consists in becoming aware of an item that, for some reason, stands out or seems useful to the

learner. It may happen while a student is reading or listening, but also when the teacher draws attention to a word and asks students about its possible meanings. Some processes that may help learners notice certain words are negotiation of meaning, providing short definitions while they are listening to or reading a story, and textual enhancement (2013, pp. 104-105); an example of the latter may be highlighting words in a text to make them more noticeable.

The second cognitive process is retrieval, also called recall. Once a lexical item has been noticed, if it is "subsequently retrieved during the task then the memory of that word will be strengthened" (Nation, 2013, p. 107), and it will probably be easier to recall in the future. This process is often opposed to recognition, which is precisely what multiple-choice vocabulary tests try to assess. These test types ask learners to select the meaning of a word from a limited number of possible options, something that is less cognitively demanding than retrieval. As for recall, it can be either receptive or productive. As it has been explained before, we say it is receptive when the learner encounters a form and has to find its meaning in their memory. Productive retrieval, on the other hand, takes place when the learner wants to express a meaning and has to remember the corresponding written/spoken form.

Another important aspect of retrieval is that it works better when it is spaced in time, rather than concentrated in one session. This is related to the concept of vocabulary recycling, which says that "it is essential that students are exposed to words often over time" (Schmitt, 2017, p. 6). Accordingly, learners should be asked to retrieve lexical items on different occasions after they have been encountered for the first time. After all, "a word is not fully learned through one meeting with it, even if this meeting involves substantial deliberate teaching" (Nation, 2013, p. 119).

Finally, the third cognitive process is creative use, which will further strengthen vocabulary retention. It takes place when, after encountering a lexical item, it is "subsequently met or used in ways that differ from the previous meeting with the word" (Nation, 2013, p. 110). This different use is what forces the student to consolidate the knowledge of the word. For example, a learner who has noticed the verb *cement,* may now see it used in a figurative sense in the sentence "we cemented our relationship with a drink" (Nation, 2013, p. 110). Creative use can also be productive. This might happen if a student takes a word such as *overwhelmed* and derives it following a known pattern, thus producing *underwhelmed*. This, of course, is more difficult for learners than the previous example.

A final issue that Nation deems important regarding direct vocabulary instruction is how detailed each explanation should be. If teachers want to know how thoroughly they should explain certain words, they should look at the reason for teaching them. When the purpose is to allow learners to continue with a communicative activity or a reading, "a short clear explanation is needed" such as "a translation, a (. . .) synonym, or a quickly drawn diagram" (Nation, 2013, p. 117). At other times, however, learners may need more detailed information about a lexical item, which is what Beck at al. (1987) called rich instruction. This type of teaching, nonetheless, is only necessary in the case of "high-frequency words and words for which the learner has special needs" (Nation, 2013, p. 117). In other cases, it would be unjustified to spend valuable class time discussing a lexical item that learners will seldom encounter. Some examples of rich instruction activities are the following: receiving input in which one word appears in many different contexts, activities in which students do research on a particularly difficult lexical item and then present it to their classmates, and practising the pronunciation of the different inflectional/derivational forms of one word.

Overall, the three cognitive processes of vocabulary learning—noticing, retrieval and creative use—can be a convenient way for instructors to go about lesson planning. Teachers who follow this framework should try to maximize the opportunities for noticing and for using (i.e. retrieving) new words. In addition, they should provide communicative contexts that require being creative and using the learned words in different ways. While many teachers already do this intuitively, thinking about the process more explicitly may make it less time-consuming.

2.3 Poetry as a tool for vocabulary acquisition

Up until this point, this section has reviewed some of the literature on poetry as a teaching material and some research on vocabulary studies. Although both fields have remained separate so far, they can be connected to SLA. As it has been explained above, the aim of this thesis is to determine whether poetry can contribute to EFL learners' vocabulary knowledge, which is an attempt to bring these two research areas together. Though this connection has rarely been established in past research, there are a few exceptions to this.

Khatib (2011) and Gómez Rodríguez (2018), for instance, have researched how poems can be used for form-focused practice in general. The former developed different language exercises from 42 English-language poems. Some texts he used to create re-ordering activities, cloze exercises and paraphrase selection tasks (2011, p. 166). In other cases, students were asked to focus on specific forms, such as passive-voice verbs or past tenses, and to transform them into active-voice structures or to put them in the present tense (p. 166). As for Gómez Rodríguez's (2018) study, though it included poetry in order to foster participants' communicative competence and aesthetic reading, the material was also used to develop their language knowledge. As stated earlier in the section, the researcher concluded that poetry had contributed to his student's lexicon (2018, p. 361), something that clearly supports my initial hypothesis. Apart from these two researchers, nevertheless, poetry has seldom been considered as a material for lexical learning.

As it will be detailed in the methodology section, for the present research, I designed an activity based on two poems that were expected to contribute to the participants' vocabulary. Certain lexical items from the poems were targeted, and students were asked to complete an activity using them. Some of these were PVs, while others belonged to different categories of lexical items. The fact that the targeted forms can be divided into two main groups will allow me to measure how each type of items is learned. Thus, I should be able to see if phrasal verbs are more easily acquired than other types of vocabulary. Apart from measuring vocabulary knowledge, the study will look into the learners' opinions of the poetry activity. This will be done by means of a perception questionnaire like the one used by Rodríguez and Puyal (2012). It should be acknowledged, however, that the scope of their research was much broader than mine, as their questionnaires focused on several areas of the participants' perceptions, such as linguistic skills, communication and intercultural competence (2012, p. 115). In spite of the more reduced size of my research, the methodology will be quite similar to theirs. The following section will provide more detailed information on the two selected poems, on the poetry-based activity and on the research instruments.

3. METHODOLOGY

3.1 Selection of teaching materials

As previously indicated, the hypothesis of this paper is that there will be an improvement in the students' vocabulary after using poetry in the classroom. To find out whether this is true or not, the internship period in the EOI of Terrassa was partially used to gather evidence connected to this hypothesis. During this period, I taught a total of 4 different sessions, which, in an EOI, amounts to 9 hours. However, only one of these sessions contained a poetry-based activity, because timing and curriculum constraints did not allow me to use other classes for the purpose of research.

In this type of end-of-master's thesis, it is more frequent to design an entire learning scenario and to implement it in the practicum centre, which is often a high school. In this case, nonetheless, my internship period took place in an EOI, which means there were certain requirements specific to the context. Ideally, my practicum intervention should have allowed me to implement a teaching unit using several poems within different activities. In this way, students would have experienced what poetry can offer as an EFL material. Nevertheless, in an EOI, teachers have to make sure that, by May, students will be ready to sit the certificate exams. Certain contents must be covered before the course ends, and learners are often concerned about not having practised enough during the previous months. In consequence, during my lessons I had to cover some aspects from the textbook that students needed to learn, although I was given some time to implement the mentioned poetry-based activity, which lasted about 35 minutes. Although this is no doubt a very short time, it was enough to expose students to two relevant poems and to conduct the necessary research.

The two mentioned poems were "Stop all the clocks" (Auden, 1936) and "January" (Updike, 1965), which may be found in **Annex 1**. The first one was written by W. H. Auden (1907-1973), a revered British poet well-known all over the English-speaking world for his technical mastery, while the author of the second one, John Updike (1932-2009), was an American novelist and poet often remembered for his "Rabbit" series of novels; both are modern classics in their respective literary traditions. Though these facts may seem superfluous, they are part of the English-speaking cultural world, which is why I considered it relevant to include a brief explanation about the authors within the poetry implementation.

These two texts were selected for several reasons, though the main one was their lexical richness. Taking into account that the chosen group was a B2.2 class, and that reading verse is almost invariably more complex than reading prose, the formal complexity of the poems was considered quite appropriate for the learners' level. In addition, the lexicon had some particularly challenging words, which learners were expected not to know, and this is what made these texts such an interesting teaching material.

In the case of "Stop all the clocks", its most clear interest lies in all the PVs that appear in the final stanza, such as *cut off*, *pack up* and *put out.* As I needed some lexical items that could be targeted for vocabulary acquisition, these phrasal verbs seemed good candidates for the study. In addition, there are also a few single-word lexical items that most learners were bound not to know (e.g. *muffled, mourner* and *moan*), and which complemented the above-mentioned PVs. As for "January", it was chosen because, although it had fewer challenging lexical items, they were also considered to be adequate. Words such as *burst* and *lace* were assumed to be new for many B2.2 learners, so I considered it would be a good opportunity for vocabulary acquisition. In addition, it was also selected because of its short extension, as it is only 16 lines long, and each of them contains between two and four words.

Apart from their lexicon, there are other reasons that make these two poems appropriate. One of them is content, which is certainly an essential part of the experience of reading poetry. Although form is extremely important in all poetry, if readers do not understand at least a part of the meaning, they will probably feel discouraged and abandon the effort. This is connected both to their lexical complexity and to whether the themes are more abstract or more concrete. Regarding the lexical complexity of the poems, even though both may be

difficult to understand, students will be reading them intensively, and they will have access to dictionaries and to their phones. Therefore, this should not be an impediment to understand the content.

When it comes to the themes of the poems, the significance of the topics was expected to motivate learners. In English literature, there are certainly instances of abstract or conceptual poetic traditions (e.g. Metaphysical poetry), which learners would probably find abstruse and boring. In this case, however, though the poems were written by two classic authors, their themes should be quite relatable for most people. "Stop all the clocks" deals primarily with the topic of loss, as its narrator expresses desolation and rage after the death of a very close person. Although it is certainly a sad topic, its universality will help readers understand what is happening almost from the beginning. "January", on the other hand, is much more mundane in topic. It describes a winter scene: the weather, the snow, the cold, and also the warmth and comfort of being indoors with a radiator. After the tragedy of the first poem, it transmits a pleasant feeling that students should be able to grasp with a couple of readings.

Another reason for selecting these two poems is their use of figurative language and rhetorical devices. Poets use these resources to make descriptions more vivid and visual, and they should make vocabulary retention easier for learners. In the first poem, the final stanza talks about putting out all stars and dismantling the Sun. Similarly, the second poem contains metaphorical language (e.g. "The sun a spark" and "The trees of lace"), which makes it more challenging, but also attractive. Apart from contributing to the learners' vocabulary retention, I considered that these rhetorical devices would force students to use their critical thinking skills in order to comprehend the poems.

A final argument for the choice of materials is that two YouTube videos were found in which the poems were performed. The first one (englishclasspoems, 2009) was a scene from the film *Four Weddings and a Funeral* in which one of the characters reads the poem in front of an audience. Since this is a rather well-known film, it was expected that some students might know it, which was supposed to make the poem seem less distant. Furthermore, the actor who performs it (John Hanna) has a strong Scottish accent, which I thought could add to the students' intercultural competence, especially if they were asked to pay attention to the pronunciation of certain words. For the second text, no film was found in which it was performed, but, in the chosen video (Emily, 2021), a voice actress read it very slowly with a clear General American accent. Therefore, I assumed learners would understand it much more quickly than they did in the case of the preceding poem.

3.2 Learning vocabulary with poems: the poetry-based activity

When it comes to the poetry-based activity itself, it consisted of the following exercises. First, the students were played the two YouTube videos before they had any access to the printed texts. This was done because I wanted learners to make a comprehension effort in order to develop their listening competence (*competència específica 2* (CE2)). I was aware that understanding the first poem might be too difficult for students, though this comprehension struggle was already considered a part of the first contact with the poems. Having to listen to them with no written text was meant to have learners notice the language and develop their aesthetic appreciation. If students had had the texts from the beginning, they would have focused much more on the words than on the listening experience.

After this initial listening, they were given a worksheet (see **Annex 1**) with three parts: the two poems, a vocabulary exercise and some discussion questions. The videos were then played again, which created a multimodal experience. After this, students had time to read the texts at their own pace, and they were asked to underline any unknown words or expressions. They were allowed to use either their mobile phones or paper dictionaries to look up the meaning of these words/expressions. This part of the activity was meant to make learners pay closer attention to the vocabulary of the texts and to help them understand the content. In other words, it was supposed to develop their reading comprehension competence (CE4).

Now that students had interacted with the written texts, some contextual details were provided regarding the authors' lives and their importance within the English-speaking world. Then, there was a brief whole-class exchange in which I asked some general questions, such as what each of the poems talked about, if there was anything they had not understood, what the differences were between the two (in language and content) and their personal opinions. The next exercise of the implementation was perhaps the most essential for lexical practice: students were asked to do a vocabulary exercise that targeted some specific lexical items. It was a gap-filling activity with 12 unrelated sentences, which students had to complete using the lexical items they had been working with. However, they were not given any possible options, because the idea was for them to go back to the poems in order to spot the necessary items. I encouraged students to work in pairs or in groups so they could help each other out. Afterwards, we corrected in a whole-class setting.

The third part of the activity was mostly unrelated to lexical acquisition, though it required learners to consider the two poems they had just read. It was a group discussion in which they had a number of questions about poetry-related topics. The exercise had been designed to develop their oral production competence (CE3), and the questions centred mostly on the themes of the first poem and on the students' views about reading poetry more generally. The main idea was to get them to talk about how they felt about poetry for a while. Finally, they were asked to share their views in a whole-class setting, and I asked them some of the worksheet questions aloud.

3.3 Research methodology and justification

Now that the materials and the poetry-based activity have been described, I will present the methodology followed to collect the necessary data, which consisted of two instruments. The first one was a vocabulary test that was administered twice: two days before the poetry-based activity and a few days afterwards. The second instrument was a student perception questionnaire, which was administered only afterwards. The purpose of the former was to measure to what extent the learners' vocabulary had improved following the poetry-based activity, while the purpose of the latter was to collect information about their views on the activity and on poetry as a material. These two instruments will now be described and their design will be justified.

The vocabulary test (see **Annex 2**) provided quantitative data about the acquisition of the 12 lexical items that had been targeted in the poetry activity. The decision to have a pre-test was due to the fact that, in order to determine the improvements in the students' lexicon, I needed information about their prior knowledge of the targeted forms. Otherwise, it would have been impossible to tell if they had acquired the vocabulary during the poetry-based activity, or if they already knew it from the beginning, which would have rendered the research quite useless. Thanks to the two tests, the results have offered reliable information about the extent to which the learners' lexicon has improved between two points in time.

Regarding the targeted lexical items themselves, the main criterion for selection was their difficulty. In other words, they were supposed to be challenging for B2.2 students. This is why the following 12 items were chosen: *sweep up*, *put out*, *cut off*, *pour away*, *pack up*, *dismantle*, *spark*, *lace*, *moan*, *prevent someone from doing something*, *purr*, and *burst*. Of course, the selection process was limited by the lexical richness of the two texts, which, as it has been indicated, were rather short. Still, they were complex enough for me to find appropriate words and expressions. In addition, the targeted items included 5 phrasal verbs that were also adequate for my research purposes. The latter were selected because an additional research interest of the study was to compare the acquisition of PVs with that of other lexical items; in the present paper, these "other lexical items" will be referred to as non-phrasal verbs (non-PVs).

It is worth pointing out that only 2 of the 5 targeted PVs appear in Garnier and Schmitt's (2015b) PHaVE list (see **Subsection 2.2.8**). This, nevertheless, is not necessarily a problem, as it implies they are not among the 100 most frequent PVs in British and American English. In consequence, it is highly probable that learners will not know their meanings. As for the 2 items that do appear in the list, they are *put out* and *cut off*. In the case of the former,

the meaning with which it is used in the poems—"stop STH from burning or shining" (2015b, p. 13)—appears with a usage frequency of 14%. The latter, as it is used in the poems, would correspond to the meaning "interrupt SB as they are speaking" (2015b, p. 15), which has a frequency of 24.5%. These meanings appear in the positions 51 and 61 in the PHaVE list respectively. This implies they are not very frequent, as the list follows a decreasing order of frequency. Therefore, they should be above the lexical repertoire of B2.2 learners. As for the MWI *prevent someone from doing something*, it cannot be considered a PV and, thus, it does not appear in the PHaVE list. In consequence, it has been considered more appropriate to categorise it as a non-PV.

In relation to the questions of the test, students were allowed to answer in one of three different ways: with a synonym, with a definition, or with the Catalan or Spanish counterpart. I gave them all these options because I wanted evidence about their knowledge, regardless of the language in which they expressed it. In the questions, the lexical items appeared in isolation, rather than in meaningful contexts, because the linguistic context would have allowed learners to guess the correct answer. It was also decided not to provide different meanings that students could choose from, as I wanted to make sure there was meaning recall—that is, having to retrieve the meaning from one's memory—, as opposed to meaning recognition, which is much less demanding. Students were not allowed to use their phones or any dictionaries to look up the items, so they had to rely only on their knowledge at the moment of doing the test. In addition, they were not told this was related to a poetry activity they were to do later, since I did not want them to anticipate any further activities using the lexical items. As for the post-test, it consisted of the same 12 lexical items that appeared in the pre-test. I considered that the vocabulary had to be kept constant as, otherwise, the pre- and post-implementation results would not have been comparable.

The second instrument was the student perception questionnaire (see **Annex 3**), which the learners answered, together with the post-test, in the last of my internship sessions. The questionnaire had 5 questions, which were aimed at providing both quantitative and qualitative data about the learners' subjective experience of working with poetry in class. In the first three questions, students had to indicate to what extent they agreed or disagreed with certain statements using a five-point Likert scale. The first two statements ("I think the poem-based activities have been enjoyable" and "I think the poem-based activities have been effective to improve my English") were connected to how learners felt about the specific activity they had done. That is why the words *enjoyable* and *effective* were included. The third statement ("I would have preferred to work with other materials, rather than with poems") was related to their general view of poetry as a material.

As it has been mentioned above, the perception questionnaire has been partially modelled after Rodríguez and Puyal's (2012) case study. As a part of their research, they also asked students to express their agreement or disagreement with a number of Likert-scale statements. Their implementation was a teaching unit based on a poem and passages from a novel, yet their questionnaire was considerably longer than mine, as it considered other aspects apart from the learners' views on literature. In my case, however, it was considered that three questions were enough for my purposes, especially considering the time constraints derived from the context of my internship.

The last two questions of the questionnaire were open-ended; one of them asked students to suggest any other materials they would have preferred to use rather than poetry, while the other one allowed them to add further comments. These were included in order to provide a qualitative counterpoint to the Likert-scale statements, and to give learners the chance to express their opinions in more detail. Neither the vocabulary test nor the three Likert-scale questions would have provided such an opportunity, and this is why the open-ended questions were deemed necessary. Indeed, the most central part of the research would have been possible without them, yet their addition has made the study more complete.

It was also decided that the perception questionnaire would be administered only after the poetry activity. Even though I was aware that Rodríguez and Puyal (2012) had asked their students to complete the questionnaires both before and after their teaching unit, this was because they were trying to measure *the evolution* of their student's intercultural competence, linguistic skills, etc. In the present study, however, administering the

questionnaire only afterwards was considered to be more appropriate. This was because I came to the conclusion that, since my students had never done any similar activities before, asking them what they thought about using poetry as a material would not have provided any significant information. What I was interested in was how they perceived my specific poetry activity after its implementation. In addition, I thought adding a perception questionnaire to the pre-test would have told students they would be working with poetry in the future. Therefore, when they had been given the activity worksheet, they might have recognised the 12 lexical items from the pre-test and, in consequence, they would have focused especially on the targeted vocabulary, and not on the poems in general. Though this would not have been a serious problem, it might have led to unnaturally high scores in the post-test, which would have decreased the reliability of the data.

Finally, another relevant choice that was made regarding the two instruments is that both would be completely anonymous. In fact, students were explicitly told not to write their names anywhere. Therefore, when looking at the results of the two tests, I will have no way of knowing whose answers I am reading, which will prevent any undesired biases. In addition, this decision was expected to allow students to express any criticisms more sincerely.

After administering the tests and the questionnaires, the final part of the methodology will consist in presenting the results and discussing their implications. For the description, the tests will be marked as if they were regular exams: every correct answer will count 1 point, so the maximum score will be 12. In addition, I will also provide 0.5 points when a student gives a definition/synonym that does not correspond to the meaning of the word in the poems but that is correct nonetheless. This is the case, for instance, of the noun *lace*, which may refer either to a cloth material or to the string used to fasten one's shoes. If a student provides the latter meaning, although it does not match how it is used in the second poem, I will consider they had some knowledge about the word, so it will be counted as partially right.

I will then distribute the scores in a number of tables (see next section) to facilitate the comparison between the learners' pre- and post-implementation performances. After this, I will look at the data in the tables and try to find relevant patterns. If I can find any, these will allow me to draw conclusions about the effectiveness of my poetry-based intervention, and I may be able to either confirm or refute my hypothesis. The conclusions, in turn, may be generalised, at least to some degree, beyond the specific circumstances of the study.

Regarding the data from the perception questionnaires, I will also use a table to present the results from the Likert-scale statements. Finally, I will collect the students' answers to the open-ended questions, which have been added at the end of this paper (see **Annex 4**). Though the answers to the Likert-scale statements will be analysed quantitatively, the answers to the open-ended questions will have to be approached qualitatively. This means that the students' most relevant comments will be considered and discussed one by one.

4. RESULTS AND DISCUSSION

4.1 Personal account of the poetry implementation

Implementing the poetry-based activity with the B2.2 group allowed me to observe how it worked out in the classroom, rather than just analysing the tables with the results. Thus, before moving on to the description of these results, I will briefly provide a subjective account of my experience and of how students responded to the activity.

As it has been stated, the activity began with a first listen of the poems without the texts. Students were played the two YouTube videos, the first of which elicited a more enthusiastic response from most learners. This was because, as it had been predicted, some of the learners had seen *Four Weddings and a Funeral* and recognised the scene, so we talked about the film for a while. They were also surprised to find out that "Stop all the clocks" was a well-known poem in English literature and were curious to learn more about it. Students found the pronunciation of the Scottish actor extremely difficult to understand, and, until I gave them the text, they did not grasp the contents of the poem. Then, they realised what it talked about, and, after being played the video for the second time, some students even seemed moved by the sadness of the topic. As for the second poem, learners did not struggle as much with it, and, taking into account their comments, I believe they found it more enjoyable. Though these are extremely subjective impressions, the students' spontaneous reactions are relevant if one wants to know whether they liked or not a given activity.

Regarding the implementation of the vocabulary exercise, seeing that students asked about the meaning of quite a few words—though they also looked them up in the dictionaries—, I assumed the vocabulary was challenging enough for their level. This was the idea behind the choice of materials, as I wanted them to feel the need to discover the meaning of these unknown lexical items. When it comes to the third exercise (the final group discussion), learners simply followed the questions in the worksheet and exchanged their views on the poems they had just read. Although in the beginning they did show interest in the topic, they seemed to run out of ideas after some minutes, something that was not unusual in discussion activities in general. My perception of the whole activity, however, was rather positive. The students' comments showed me that most of them had enjoyed listening to both poems. Still, regarding the first text, it is difficult to know whether they liked it due to its aesthetic qualities or because of its connection to a famous film.

4.2 Results

As it has been detailed in the Methodology section, the pre- and post-tests that students were asked to complete were the main source of information for the research. Approximately one week elapsed between these two tests and, fortunately, the same 13 students attended class throughout the whole of my implementation, so there were no changes in the participants. This will allow me to compare the pre- and post-test scores in a more reliable way.

Before discussing the data from the research, the disposition of the results will be explained. As it may be seen in **Table 2** and **Table 3** below, the scores of the different items appear under the headings "test 1", "test 2" and so on; it is worth noting, however, that there is no connection between the numbers in the pre- and in the post-test results. This means that the "test 1" in the pre-implementation does not necessarily correspond to the same student as the "test 1" in the post-implementation. This is because the tests were shuffled to guarantee anonymity; therefore, there is no correspondence between them.

After marking the pre- and post-tests, I entered the points for each lexical item in two tables (**Table 2** and **Table 3**) using Excel. I then added them up in order to find the total score for each test. I also calculated the average score for each of the 12 items individually, which has provided information about which lexical items most students already knew from the beginning (such as *dismantle* and *pack up*), and which ones were mostly

unknown (such as *put sth. out* and *cut off*). In the same two tables, I have also included the average of the total scores, which is particularly useful because it provides a general idea of how the whole class performed before and after the implementation.

In addition, **Table 4** has also been created to summarise the information in the previous two tables. More specifically, it includes the average of all the students' scores for each lexical item. Because the pre- and post-test scores have been put side by side, comparing them will be considerably easier. Finally, I created **Table 5**, which groups the 12 lexical items into two categories: phrasal verbs and all the other types of lexical items (non-PVs). To fill it in, I calculated the average score of each grouping for the 13 students and I also found the total averages, which are in bold. The division into these two groupings will provide information on how learners acquired PVs as opposed to non-PVs. Most relevantly, the final column includes two percentages that quantify the extent of their improvement in each of the two categories.

As for the results from the perception questionnaires, this instrument was administered to the same group of students who had answered the post-tests, which means there should have been 13 participants. Two learners, nevertheless, left their questionnaires blank. This was possibly due to a personal choice, as they had been explicitly told that, on the other side of the vocabulary test, they would find some subjective questions. In consequence, I have decided to discard their questionnaires, which means that the total number of participants is 11.

As stated in the Methodology section, the three Likert-scale questions asked students to reflect on the enjoyableness and effectiveness of the implementation, as well as on their views on poetry as a material. The results gathered from this research instrument have been expressed in **Table 6** in the form of percentages, which should make it easier to see the distribution of their answers. In the following paragraphs, I will first deal with the vocabulary-test results (**Tables 2, 3, 4 and 5**), then move on to the Likert-scale statements (**Table 6**), and finally discuss the answers to the open-ended questions. Only the most relevant among these will be presented, though the students' complete comments are available in **Annex 4**.

4.2.1 Vocabulary test results

	test 1	test 2	test 3	test 4	test 5	test 6	test 7	test 8	test 9	test 10	test 11	test 12	test 13	Average score
1. sweep up	1	1	1	0	0	0	0	0	0	0	0	0	1	0,31
2. dismantle	1	1	1	1	0	1	0	0	1	1	0	0	1	0,62
3. spark	0	0	1	1	0	1	0	0	1	1	0	0	1	0,46
4. lace	0	0	0	0	0	0	0	0	0	0	0	0	1	0,08
5. put sth. out	0	0,5	0	0	0	0	0	0	0	0	0	0	1	0,12
6. moan	0	0	0	0	0	0	0	0	1	0	0	0	1	0,15
7. cut off	0	0,5	0	0	0	0	0	0	0	0	0	0	0	0,04
8. prevent sb. from doing sth.	0	0	0,5	0	0	0	0	1	0	0	1	1	1	0,35
9. pour sth. away	1	0	0,5	0	0	1	0	1	1	0	0	0	0	0,35
10. purr	0	0	0	0	0	1	0	0	1	1	0	0	1	0,31
11. pack up	1	1	1	1	1	1	1	1	1	0	0	1	1	0,85
12. burst	0	0	0	0	0	0	0	0	1	0	0	0	1	0,15
Total score	**4**	**4**	**5**	**3**	**1**	**5**	**1**	**3**	**7**	**3**	**1**	**2**	**10**	**3,77**

Table 2. Pre-test scores

	test 1	test 2	test 3	test 4	test 5	test 6	test 7	test 8	test 9	test 10	test 11	test 12	test 13	Average score
1. sweep up	1	1	1	1	1	0	1	0,5	1	1	1	0	1	0,81
2. dismantle	0	1	1	1	1	0	0	1	1	1	1	1	1	0,77
3. spark	1	1	1	1	1	0	1	0	0	1	1	1	1	0,77
4. lace	0	1	0	1	1	0	1	0	0	1	1	0	1	0,54
5. put sth. out	0	1	0	0	0	0	0	0	0	1	0	0	0	0,15
6. moan	0	1	0	0,5	1	1	0	0	1	1	0	0	1	0,50
7. cut off	0	0	0	0,5	0	0	0	0	0	0	1	0	0	0,12
8. prevent sb. from doing sth.	0,5	0	0	0	0	1	0	0	0	0	1	0	0	0,19
9. pour sth. away	0	0	0	0	1	0	0	0	0	1	1	1	0	0,31
10. purr	0	1	0	0	0	1	0	1	0	1	1	1	1	0,54
11. pack up	1	1	1	1	1	1	1	1	1	1	1	1	1	1,00
12. burst	0	1	0	0	0	0	1	0,5	1	1	1	0	1	0,50
Total score	**3,5**	**9**	**4**	**6**	**7**	**4**	**5**	**4**	**5**	**10**	**10**	**5**	**8**	**6,19**

Table 3. Post-test scores

As it may be seen in **Table 2** and **Table 3**, the results indicate a general improvement in the learners' knowledge of the targeted forms. In the pre-test, the average total score was 3.77, while in the post-test it was 6.19. This means there was an increase of 64% in the learners' overall knowledge, which is a considerable improvement. It is important to bear in mind, however, that the highest possible score was 12, which means that the marks are quite low in absolute terms. In the pre-test, this is quite unsurprising, as I had deliberately chosen lexical items that learners were supposed to struggle with. However, in the post-test, the average is lower than it might have been expected. This becomes more evident if we look at the scores test by test. In the pre-test, there are only two total scores over 6 (a 7 and a 10), and the lowest mark is 1. Therefore, if the pass mark is set at 6, most people could be said to have failed. As for the post-test, the scores tend to be considerably higher, as 5 students obtained 7 points or more (though no one reached the maximum score). Nevertheless, more than half of the class still had a mark under 7 and, this time, the lowest mark was a 3.5.

Item	Average pre test score (max.=1)	Average post test score (max.=1)
1. sweep up	0,31	0,81
2. dismantle	0,62	0,77
3. spark	0,46	0,77
4. lace	0,08	0,54
5. put sth. out	0,12	0,15
6. moan	0,15	0,50
7. cut off	0,04	0,12
8. prevent sb. from doing sth.	0,35	0,19
9. pour sth. away	0,35	0,31
10. purr	0,31	0,54
11. pack up	0,85	1,00
12. burst	0,15	0,50
Total score (max. = 12)	**3,77**	**6,19**

Table 4. Pre- and post-test scores compared

Pre test															
	test 1	test 2	test 3	test 4	test 5	test 6	test 7	test 8	test 9	test 10	test 11	test 12	test 13	Total average	Increase (%)
PVs average	0,60	0,60	0,50	0,20	0,20	0,40	0,20	0,40	0,40	0,00	0,00	0,20	0,60	**0,33**	
non-PVs average	0,14	0,14	0,36	0,29	0,00	0,43	0,00	0,14	0,71	0,43	0,14	0,14	1,00	**0,30**	
Post test															
	test 1	test 2	test 3	test 4	test 5	test 6	test 7	test 8	test 9	test 10	test 11	test 12	test 13	Total average	
PVs average	0,40	0,60	0,40	0,50	0,60	0,20	0,40	0,30	0,40	0,80	0,80	0,40	0,40	**0,48**	45
non-PVs average	0,21	0,86	0,29	0,50	0,57	0,43	0,43	0,36	0,43	0,86	0,86	0,43	0,86	**0,54**	80

Table 5. Phrasal verb and non-phrasal verb average scores

In **Table 5**, which groups the data into PVs and non-PVs, the results also show certain trends. In this case, it should be remembered that the maximum score in all cases is 1, and not 12. This is because I calculated the average from the grades of the different lexical items. The problem was that, among the targeted items, the number of PVs and non-PVs was not equal (i.e. they were 5 and 7, respectively). This means that, if I had not used the averages, the PV scores would have been unnaturally lower than those of non-PV, and that the two categories would not have been comparable.

If we focus on the column with the total averages, it becomes evident that there was an improvement in both categories, which was to be expected. The scores for the PV questions increased from 0.33 to 0.48, which represents an improvement of 45%. The non-PV questions, on the other hand, went from 0.3 to 0.54; that is to say, students improved by 80%, an increase that is considerably higher. As for the data of individual students taken separately, it confirms the same tendency: the averages for PV items are slightly better in the post-test than in the pre-test. The highest number before the activity was 0.6, while afterwards it climbed to 0.8. Regarding non-PVs, the improvement stands out more clearly: in the pre-test, most averages were under 0.4, while, afterwards, there were six people over 0.4, and some even reached a 0.86. Thus, it cannot be denied that, in general, the

acquisition of non-PVs was more successful than that of PVs. This is something that will be considered in the discussion section of the paper.

4.2.2 Perception questionnaire results

Statement	Strongly agree		Agree		Neutral		Disagree		Strongly disagree	
	students	%	students	%	students	%	students	%	students	%
1. Enjoyableness	2	18,18	7	63,64	2	18,18	0	0	0	0
2. Effectiveness	3	27,27	6	54,55	2	18,18	0	0	0	0
3. Other materials	0	0	1	9,09	6	54,55	3	27,27	1	9,09

Table 6. Distribution of the answers to the Likert-scale statements

Regarding the distribution of the students' personal views, in general, their answers point to a positive attitude towards poetry. When it comes to the first statement ("I think the poem-based activities have been enjoyable"), most learners (64%) said they agreed with it. There was also an 18% who strongly agreed, and another 18% who felt neutral towards it. The second statement ("I think the poem-based activities have been effective to improve my English") yielded similar results: roughly 55% of the participants expressed agreement. In this case, the "strongly agree" option was marked by 27%, and the "neutral" option was chosen by 18%. Quite tellingly, no one expressed any disagreement or strong disagreement with statements 1 or 2.

The final statement ("I would have preferred to work with other materials, rather than with poems") was expressed in a negative way, which means that, if students had really enjoyed the poems, the expected distribution of answers would have been mostly "disagree" and "strongly disagree". The results, however, point in another direction, as most people (55%) felt neutral towards the statement. Some participants (27%) disagreed with it, and there was one student (9%) who strongly disagreed with it. Interestingly, another learner (9%) agreed with the statement. The latter is the only slightly negative opinion towards poetry in the whole set of questionnaires, and we will soon return to it. In general, however, most answers evidence a rather positive attitude towards the poetry-based activity and towards the materials used.

After the three Likert-scale statements, there were the two open-ended questions: "If you agreed with the last sentence, what other materials would you have preferred to use?" and "Do you have any other comments or suggestions?". These were phrased in a way that suggested they were optional, which might explain why there were only two answers to the first question and six to the second question. I will now present some of the most interesting comments, which have been transcribed verbatim.

There was one student who answered "I appreciate all materials in class" to the first question, and "I love poetry" to the second one, which shows this person was probably already predisposed towards poetry before the activity. In another answer, a student valued the originality of the activity: "It was nice to work with poems, it's an alternative or an original way of learning because we are not used to". Yet another positive comment to the second question was "No, the poem activity was perfect to *learn some more vocabulary*. The poems were nice" (emphasis added). This is interesting because, although the questionnaire had not mentioned vocabulary anywhere, the answer evidences this person was aware of having learned vocabulary. In addition, it seems he/she valued the aesthetic qualities of the materials ("the poems were nice"), rather than seeing them just as a means to an end.

Finally, another learner wrote they would have preferred using "videos" rather than poems. Relevantly, this is the same student who marked "agree" in the third Likert-scale statement, two answers that are quite valuable. As stated before, the open-ended questions were supposed to provide a space for students to express their opinions freely, whether they were positive or negative. In fact, I had encouraged them to express any dissatisfaction with the activity or the materials, and, because I wanted them to be sincere, I had stressed that the tests were anonymous. Still, the latter answer was the only relatively negative one in relation to poetry as a material, and this is why I consider it so relevant.

4.3 Discussion and limitations

4.3.1 Discussion of the results

Now that the results have been presented, I will consider their implications for the present Master's Thesis. As far as the vocabulary tests are concerned, the average scores from the pre- and post-tests point to an improvement in the students' knowledge of the targeted forms. As it may be seen in **Table 2** and **Table 3**, while the average score in the pre-tests was 3.77, in the post-tests it was 6.19. In relative terms, this implies there was an increase of 64% percent in the students' average knowledge of the assessed items. This is a considerable improvement, although, in absolute terms, the students' performance was much less remarkable. If we consider the pre-test scores, only two learners passed the test, which means there was a lot of room for improvement. In the post-test, 6 people passed, which is certainly an increase; however, taking into account that there were 13 participants, it implies more than half of the participants did not pass the post-test. For the present thesis, nonetheless, the most relevant indicator is the change in the students' vocabulary between their pre- and post-implementation states of knowledge. Even though some students had quite low scores in the post-test (e.g. a 3.5), in the light of the results, it can be concluded that the initial hypothesis has been confirmed: the class improved their vocabulary after the poetry-based activity.

As for the distinction between the acquisition of PVs and that of non-PVs, the activity has also contributed to the learning process. However, the data reveals that the students were clearly more successful in learning non-PVs than PVs. More specifically, in the pre-test, the total average for the non-PV questions was 0.3, while the same parameter in the post-test was 0.54; in other words, the marks improved by 80%. When it comes to PVs, the pre- and post-test averages were 0.33 and 0.48 respectively, which represents a 45% increase. The difference may seem considerable enough to conclude that poetry is less effective for learning PVs than for other types of vocabulary. This, nevertheless, would be an overgeneralisation, as it will be seen in the limitations subsection.

Regarding the perception questionnaires, the results of both the Likert-scale statements and the open-ended questions show that most students had a positive view of the implemented activity and the poems. This is especially evident in the case of the first two statements. The vast majority of the learners (80%) expressed either agreement or strong agreement with them, which indicates they considered the activity to be both enjoyable and effective. The third statement, which referred to the possibility of using other materials apart from poems, yielded less clear results: most students (55%) were neutral towards it, and one person expressed a preference for other materials. Nevertheless, around 40% of the participants disagreed or strongly disagreed with the statement, which is also encouraging in relation to poetry.

The answers to the two open-ended questions, in spite of their paucity, were positive as well. In particular, there are two comments that stand out because they refer to the originality of the poems and to their usefulness for learning new vocabulary. Another student also used this part of the questionnaire to express their preference for including videos, rather than poems, in the lessons. Though this is not a negative comment per se (it is merely an opinion), it is a reminder that not all learners were satisfied with the activity. This is relevant because poetry, as it has been stated in the theoretical framework, is not always perceived as an appropriate instructional material. In fact, as McKay (1982) explains, poetry is particularly complex in terms of language, structure and cultural specificity (p. 529). Although these issues can also be framed as advantages (see **Subsection 2.1.4**), it is equally true that they may lead some students to feel reluctant to read poetry. This should be respected, and teachers should try not to overuse poetry—or any other material—as an EFL resource. In other words, though the focus of the present thesis has been poetry, it is quite evident that variety is essential when it comes to authentic materials. Even though many teachers are bound to have a preference for certain cultural products, this is not a valid argument for excluding other materials from their lessons. Still, we should not lose sight of the fact that the students' views have been rather positive towards the use of poetry, and that more research is needed to draw definitive conclusions either in favour or against poetry within EFL. The next subsections will focus on the main

limitations of the research instruments, and they will provide some possible solutions for future studies in this area.

4.3.2 Limitations of the vocabulary tests

The most evident limitation of the vocabulary test is that the sample of students was too reduced. Because of the circumstances of my internship, I was only able to implement the poetry-based activity with one group in the school. In addition, only 13 students attended class during the weeks when I taught my sessions. Consequently, even though there was a clear increase in their vocabulary knowledge, it is not possible to establish any irrefutable relationship of causality between the poetry activity and the students' learning process. A possible solution to this would be to replicate the study with more groups of the same level; this would certainly be possible, as in the EOI of Terrassa there are several B2.2 English groups. Of course, more time and resources would be needed, but the data would increase in reliability.

Yet another problem was that the two poems limited the number of lexical items that could be targeted for acquisition. This is an unavoidable issue with authentic materials, yet in the case of poetry, the texts tend to be extremely short, which means there are few words one can choose from. This is why I had tried to find appropriate poems in terms of lexicon. The first text ("Stop all the clocks") was probably a good choice, as it contained numerous phrasal verbs that the students were unlikely to know, but the second one ("January") was slightly more problematic. This was because it contained only a few items that were challenging enough for a B2.2 level, so I had no choice but to target these words. In consequence, the items from the second poem were very different from the rest, and the 12 targeted items ended up being a mix of PVs and other, unrelated words (such as *purr* and *lace*). In hindsight, it seems it would have been better to work with items more consistent in terms of semantic field and grammatical category.

If I were to design the study again, I would substitute "January" for another poem which included 5 adequate single-word items. This would allow me to solve another design issue, which is that, in the present study, there were 5 PVs and 7 non-PVs. Therefore, when I separated the results into two different sets, I had to use averages, which made managing the data more difficult. If there had been 5 PVs and 5 single-word verbs, I could have compared the students' performance in these two categories more easily. As it has been discussed, the results indicate that students learned non-PVs more effectively than PVs. This, nevertheless, does not necessarily mean that poetry should never be used for the acquisition of phrasal verbs. In order to determine how PVs are acquired in comparison to other word types, it would be necessary to carry out research with more participants and more diverse implementations.

This, in fact, is related to another shortcoming of the research: even if the mentioned study with 5 PVs and 5 single-word verbs were implemented, it would still be rather difficult to establish if the learners' lexical improvements had actually been due to the poems. If there was an improvement in their vocabulary, it would be undeniable that, at least, poetry is not detrimental to their development. If, on the other hand, the learners' performance in the post-test was worse than in the pre-test, it would be evident that the activity was not well designed for vocabulary acquisition. In the present case, it seems the implementation has had a positive impact on the students, although we cannot ascertain if this is because of the chosen poems or due to a combination of several factors.

An improvement that could solve this shortcoming would be to include a control group that completed a similar activity but without poetry. Thus, the control group would work with the same 12 lexical items as in the poetry-based activity, though they would encounter them in a more conventional EFL context. A possibility would be to have students in the control group listen to and read an article that contained the targeted lexical items. They would then repeat roughly the same process as in the poetry implementation. That is to say, they would look up any unknown items in the text and complete a gap-filling activity (which would be exactly the same as the one for the treatment group). Finally, there would be a group discussion connected to the topic of the text. Thus, by means of a control group, the study could isolate the variable of the discursive genre used and keep all the other

circumstances constant. After the two groups had answered the pre- and post-tests, the researcher could compare the two data sets, and this would lead to more reliable conclusions regarding the usefulness of poetry in EFL.

Finally, the last shortcoming of the tests is connected to the type of assessment used to measure vocabulary knowledge. As explained in the previous section, the test questions had no context (see **Annex 2**), as I did not want learners to have any clues regarding the meaning of the items. However, I did not realise that some of them, such as *lace*, were clearly polysemous, which means that the test did not make answering with the appropriate meanings essential. For example, a student may have known that *lace* means *encaje* in Spanish, but might have written the translation *cordó de les sabates*, which is probably a more prototypical meaning of the word. In this case, there would have been no way of determining if the student only knew the latter meaning of the item, or if they knew both. When something like this happened, I chose to give the student 0.5 points, as the answer could not be considered wrong. An alternative would have been to design a test that made answering with the required meanings essential and that ruled out any additional meanings. In other words, it might have been better to use multiple-choice questions, which are more effective when dealing with polysemous items, because the researcher can control the available answers. Even though guessing is always an issue in the case of multiple-choice tests, in the present case, reliability and essentialness were more important priorities.

4.3.3 Limitations of the perception questionnaires

Regarding the perception questionnaires, though they have fewer limitations than the tests, the paper would not be complete if they were not addressed. The first shortcoming is that the information collected by means of the questionnaires was very subjective. Of course, this instrument is inherently subjective, as it tries to measure how a group of people feel or think about a certain experience. In this case, the data obtained cannot provide any conclusions about the actual effectiveness of poetry activities in general. All it can do is reflect what students felt after working with poems in class. In this sense, the questionnaire did what it was supposed to do, though it could have been improved. In hindsight, I believe I could have included more Likert-scale statements that students could agree/disagree with, following Rodríguez and Puyal's (2012) methodology. As it is, there are two statements that ask students what they thought about the activity itself, and only one about their views on poetry as a material. This means there is an information gap, which could have been filled in by including more statements about poetry (e.g. "I would like to keep using poetry in class" and "I found the poems interesting"). This would have made my research more thorough, though it would have taken students longer to complete the questionnaires. Because the implementation time was limited, I considered three statements were enough, yet further research would benefit from more complete perception questionnaires.

Another issue is that the learners' answers to the questionnaire may not be entirely reliable. As I was both the researcher and the trainee teacher who implemented the activity, some biases were unavoidable, especially because I had a good relationship with the students. In addition, most of them were aware the classes were a part of my internship, and they wanted me to feel as comfortable as possible, so they were enthusiastic about all the activities I proposed. This is why I believe their perceptions may have been influenced positively by the context. Of course, this is not to say that the questionnaire results are utterly unreliable or that the participants lied. Instead, it means that some biases are unavoidable when the teacher and the researcher are the same person, and that this should be taken into consideration when looking at the students' opinions.

Finally, a related shortcoming is that the students' views were based only on the two poems they had read and listened to. Furthermore, most learners had never worked with poetry in English before. This means the material was a novelty for them, which may have contributed to their enjoyment. If the implementation had been longer and they had been exposed to different poems for several weeks, their views might have been more representative of how students respond to poetry as an EFL material. This is precisely another reason why further research is necessary regarding students' perceptions.

In sum, in spite of the limitations of the research design, the results collected using both instruments have been relevant for my research purposes. The test scores have confirmed that students have certainly improved their vocabulary after the poetry-based implementation, and the questionnaires have shown that the participants have generally perceived the activity as enjoyable and effective. Therefore, though the study certainly has quite a few shortcomings, it is also essential to value what has been achieved with such a limited implementation and in such a short time.

5. CONCLUSIONS

This Master's Thesis has focused on the topic of using poetry in EFL classes. Its aim has been to determine if poems can be a helpful material for the acquisition of vocabulary in the case of adult learners. Thus, I have tried to explore how students can increase their vocabulary knowledge by reading and listening to poems as well as by doing activities connected to this discursive genre. The hypothesis of the study has been that the results of the research would show an improvement in the learners' vocabulary after doing a poetry-based activity.

The research has been conducted in the EOI of Terrassa, and the participants have been a group of 13 students of EFL. The students completed a poetry-based activity in which they had to read and listen to two poems, do a vocabulary exercise, and discuss the materials in groups. In addition to the implementation itself, I designed a vocabulary test aimed at assessing the learners' knowledge of the lexical items they had worked on. The test was administered both before and after the activity itself, and it was expected to provide evidence for the corroboration or refutation of the hypothesis. Apart from the test, learners had to answer a questionnaire about their perceptions regarding the poetry-based activity.

The results have shown that, after the implemented activity, learners had clearly improved their knowledge of the targeted lexical items. This seems to imply that the poetry-based activity has contributed to their vocabulary. As for the questionnaire, the data indicated most students deemed the activity enjoyable and effective for their learning process. These results, even though they are subjective, are quite encouraging for the future of poetry as an EFL material.

After dealing with the results, the paper has discussed the shortcomings of the research. One of its limitations is that the methodology did not consider the need for a control group. Therefore, it is virtually impossible to tell if the lexical improvement of the learners has derived from the inclusion of the poems, or if any other material would have led to similar results. Another relevant limitation lies in the two poems chosen. If the study were to be repeated, the texts should be selected more carefully so that they included more appropriate lexical items, especially in relation to the categories of phrasal verbs and non-phrasal verbs. Moreover, students may have been influenced by the novelty of poetry, as they had never used this material in EFL classes before. This shortcoming should be taken into consideration, and further research should expose learners to more poems and to more diverse vocabulary-acquisition exercises.

All in all, the present research can be said to support my initial hypothesis: the poems have helped most students improve their vocabulary knowledge. Nevertheless, given all the shortcomings in the design, it is evident that more research is needed in this area. Therefore, if similar studies are carried out in the future, it will be important to employ more reliable methodologies. So far, the field of poetry as a teaching material has been rather unexplored. However, poems have a great deal to offer to teachers and students alike, so let us hope that researchers will continue to investigate the uses of poetry in foreign language acquisition.

REFERENCES

Auden, W. H. (1936). *Stop all the clocks*. https://web.cs.dal.ca/~johnston/poetry/stopclocks.html

Beck, I. L., McKeown, M. G., & Omanson, R. C. (1987). The effects and uses of diverse vocabulary instructional techniques. In M. G. McKeown & M. E. Curtis (Eds.), The nature of vocabulary acquisition (pp. 147–163). Lawrence Erlbaum Associates, Inc.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, *20*(2), 136–163.

Coleman, A. (1929). *The teaching of modern foreign languages in the United States: A report prepared for the Modern foreign language study*. The Macmillan Company.

Emily. (2021, February 20). *January by John Updike* [Video]. YouTube. https://www.youtube.com/watch?v=vPsZj9GVOnM

englishclasspoems. (2009, September 22). *Funeral Blues—Four Weddings and a Funeral* [Video]. YouTube. https://www.youtube.com/watch?v=DDXWclpGhcg

Front page. (n.d.). Hello Poetry. https://hellopoetry.com/

Gardner, D., & Davies, M. (2007). Pointing Out Frequent Phrasal Verbs: A Corpus-Based Analysis. *TESOL Quarterly*, *41*(2), 339–359. https://doi.org/10.1002/j.1545-7249.2007.tb00062.x

Garnier, M., & Schmitt, N. (2015a). The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research : LTR*, *19*(6), 645–666. https://doi.org/10.1177/1362168814559798

Garnier, M., & Schmitt, N. (2015b). *Vocabulary Resources: PHaVE List*. Norbert Schmitt. https://www.norbertschmitt.co.uk/vocabulary-resources

Gómez Rodríguez, L. F. (2018). English learners' voices on integrating poetry through a transactional approach in an EFL classroom. *Literatura y Lingüística*, *37*, 355–380. https://doi.org/10.29344/0717621X.37.1387

Grant, L. E., & Nation, I. S. P. (2006). How many idioms are there in English? *ITL - International Journal of Applied Linguistics*, *151*(1), 1–14. https://doi.org/10.2143/ITL.151.0.2015219

Khatib, M. (2011). A new approach to teaching English poetry to EFL students. *Journal of Language Teaching and Research*, *2*(1), 164-. https://doi.org/10.4304/jltr.2.1.164-169

Khatib, M., Rezaei, S., & Derakhshan, A. (2011). Literature in EFL/ESL Classroom. *English Language Teaching*, *4*(1), Article 1. https://doi.org/10.5539/elt.v4n1p201

Kormos, J. (2020). How Does Vocabulary Fit Into Theories of Second Language Learning? In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 207–222). Routledge. https://doi.org/10.4324/9780429291586-14

Kramsch, C., & Kramsch, O. (2000). The Avatars of Literature in Language Study. *The Modern Language Journal*, *84*(4), 553–573.

Langer, J. A. (1997). Literacy Issues in Focus: Literacy Acquisition through Literature. *Journal of Adolescent & Adult Literacy*, *40*(8), 606–614.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1–26. https://doi.org/10.1093/applin/22.1.1

Liu, D. (2011). The Most Frequently Used English Phrasal Verbs in American and British English: A Multicorpus Examination. *TESOL Quarterly*, *45*(4), 661–688. https://doi.org/10.5054/tq.2011.247707

McKay, S. (1982). Literature in the ESL classroom. *TESOL Quarterly*, *16*(4), 529–536. https://doi.org/10.2307/3586470

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology*, *94*(1), 156–163. https://doi.org/10.1037/0022-0663.94.1.156

Nation, I. S. P. (2007). The Four Strands. *Innovation in Language Learning and Teaching*, *1*(1), 2–13. https://doi.org/10.2167/illt039.0

Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139858656

Nation, I. S. P. (2020). The Different Aspects of Vocabulary Knowledge. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies*. Routledge. https://doi.org/10.4324/9780429291586-14

Nation, I. S. P. (2021). Is it worth teaching vocabulary? *TESOL Journal*, *12*(4), e564. https://doi.org/10.1002/tesj.564

Oxford English Dictionary. (2024). *Vocabulary, n., sense 4*. In *Oxford English Dictionary*. https://doi.org/10.1093/OED/8754098917

Poetry Foundation. (2025, April 18). The Poetry Foundation. https://www.poetryfoundation.org/

Qian, D. D. (2002). Investigating the Relationship Between Vocabulary Knowledge and Academic Reading Performance: An Assessment Perspective. *Language Learning*, *52*(3), 513–536. https://doi.org/10.1111/1467-9922.00193

Rodríguez, L. M. G., & Puyal, M. B. (2012). Promoting Intercultural Competence through Literature in clil Contexts. *Atlantis (Salamanca, Spain)*, *34*(2), 105–124. https://www.jstor.org/stable/26453027

Schmitt, N. (2017). Vocabulary: Principles and practice. *English Teaching Professional*, *109*, 4–6. https://www.norbertschmitt.co.uk/articles

Updike, J. (1965). *January*. Poem Hunter. https://www.poemhunter.com/poem/january-30/

Webb, S. (2020). Introduction. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 1–12). Routledge. https://doi.org/10.4324/9780429291586-1

Wolter, B. (2020). Key Issues in Teaching Multiword Items. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 493–510). Routledge. https://doi.org/10.4324/9780429291586-31

ANNEXES

ANNEX 1. Worksheet for the poetry-based activity

Poetry in English

"Stop all the clocks" by W. H. Auden

Stop all the clocks, cut off the telephone,

Prevent the dog from barking with a juicy bone,

Silence the pianos and with muffled drum

Bring out the coffin, let the mourners come.

Let aeroplanes circle moaning overhead

Scribbling on the sky the message He Is Dead,

Put crepe bows round the white necks of the public doves,

Let the traffic policemen wear black cotton gloves.

He was my North, my South, my East and West,

My working week and my Sunday rest,

My noon, my midnight, my talk, my song;

I thought that love would last for ever: I was wrong.

The stars are not wanted now: put out every one;

Pack up the moon and dismantle the sun;

Pour away the ocean and sweep up the wood;

For nothing now can ever come to any good.

"January" by John Updike

The days are short,

The sun a spark

Hung thin between

The dark and dark.

Fat snowy footsteps

Track the floor.

Milk bottles burst

Outside the door.

The river is

A frozen place

Held still beneath

The trees of lace.

The sky is low.

The wind is gray.

The radiator

Purrs all day.

Read the poems again and underline the words you do not understand. Try to guess their meanings using their context. Then, use a dictionary to check what they mean.

Vocabulary

1 - Use words and phrasal verbs from the poems to fill the gaps in the following sentences.

1. The fire began because of a _____ from the fireplace.
2. Could you please _____ the broken glass from the floor before someone gets hurt?
3. After the picnic, we had to _____ all our things and head back home.
4. We had to _____ the candles before leaving the room.
5. She's always _____ about her workload, but she never asks for help.
6. The hot-air balloon _____, so they could not begin the expedition.
7. The Internet router broke down, so we were suddenly _____ in the middle of the video call.
8. If you don't lock the door properly, you won't _____ anyone _____ getting inside.
9. The kitten began to _____ softly when I stroked its back.
10. I accidentally added too much salt to the soup, so I had to _____ it _____ and start again.
11. The police _____ the car and inspected its parts separately.
12. The _____ curtains in the living room were beautiful.

Discussion

Listen to "Stop all the Clocks" and "January" again and discuss in pairs:

- What do you think about the funeral scene? Did you find the poem touching or too melodramatic for the situation?
- Do you think using poetry in funerals is appropriate? And what about using sad songs? Is there any other type of music you would prefer?
- Which of the two poems did you like best? Do you ever read poetry in your own language? Do you think song lyrics should be considered poetry?

ANNEX 2. Vocabulary pre- and post-tests

Vocabulary test (1)

Write a definition, a synonym or a Spanish/Catalan translation for each of these words:

1. sweep up (phrasal verb):

   _____

2. dismantle (verb):

   _____

3. spark (noun):

   _____

4. lace (noun):

   _____

5. put something out (phrasal verb):

   _____

6. moan (verb):

   _____

7. cut off (phrasal verb):

   _____

8. prevent somebody from doing something (expression):

   _____

9. pour something away (phrasal verb):

   _____

10. purr (verb):

    _____

11. pack up (phrasal verb):

    _____

12. burst (verb):

    _____

Vocabulary test (2)

Write a definition, a synonym or a Spanish/Catalan translation for each of these words:

1. sweep up (phrasal verb):

   _____

2. dismantle (verb):

   _____

3. spark (noun):

   _____

4. lace (noun):

   _____

5. put something out (phrasal verb):

   _____

6. moan (verb):

   _____

7. cut off (phrasal verb):

   _____

8. prevent somebody from doing something (expression):

   _____

9. pour something away (phrasal verb):

   _____

10. purr (verb):

    _____

11. pack up (phrasal verb):

    _____

12. burst (verb):

    _____

ANNEX 3. Perception questionnaire

2 - Consider the two poem-based activities you have done in class and, for each statement, put an X in one of the boxes.

I think the poem-based activities have been enjoyable.

Strongly agree	Agree	Neutral	Disagree	Strongly disagree
☐	☐	☐	☐	☐

I think the poem-based activities have been effective to improve my English.

Strongly agree	Agree	Neutral	Disagree	Strongly disagree
☐	☐	☐	☐	☐

I would have preferred to work with other materials, rather than with poems.

Strongly agree	Agree	Neutral	Disagree	Strongly disagree
☐	☐	☐	☐	☐

If you agreed with the last sentence, what other materials would you have preferred to use?

Do you have any other comments or suggestions?

ANNEX 4. Answers to the open-ended questions (perception questionnaire)

[The following answers have been reproduced verbatim]

First open-ended question:

If you agreed with the last sentence, what other materials would you have preferred to use?

"I appreciate all materials in class."

"Videos"

Second open-ended question

Do you have any other comments or suggestions?

"It was nice to work with poems, it's an alternative or an original way of learning because we are not used to. So thankyou!!"

"I love poetry."

"Maybe you can improve speaking with more conviction and volume."

"As a suggestion, I would like to say that I would preffered to have more homework. Nevertheless, the classes were so dinamic and fun. Thank you for your job with us."

"No, the poem activity was perfect to learn some more vocabulary. The poems were nice"

"It was an enjoyable activitie"